



Improving CO_2 assimilation efficiency in *Cupriavidus necator* H16

Giorgia Tibaldero MSc (Hons)

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

Supervised by: Katalin Kovács
Naglis Malys
Nigel Minton

Declaration

Unless otherwise acknowledged, the work presented in this thesis is my own. No part has been submitted for another degree at the University of Nottingham or any other institute of learning.

Abstract

In this fast-developing world, there is an increasing demand for biodegradable and renewable materials derived from cheap and wildly available resources. Waste gases from industry (steel manufacturing, oil refining, coal and natural/shale gas) can be used as alternatives to petroleum-based solutions. In this context, the development of technologies allowing us to meet new trends and demands is becoming increasingly essential.

Several microorganisms are known to have the potential to address this challenge. In particular, the model organism *Cupriavidus necator* H16 is able to produce naturally biodegradable thermoplastics that can be used as renewable source for the plastic industry growing the bacteria on CO_2 and H_2, as a sole carbon and energy source.

In this study a dual strategy is used in an attempt to improve CO_2 assimilation efficiency in *C. necator* H16. The first strategy involved the heterologous expression of genes for self-assembled proteinaceous organelles (α-carboxysomes) in *C. necator,* to observe if their carbon concentrating mechanism (CCM) boosts the carbon fixation efficiency of the organism. I have shown that it is possible to heterologously express the *cso* operon in *C. necator*, however the functionality of the α-carboxysomes could not be confirmed. Further analysis on the mechanism driving RubisCO encapsulation in the α-carboxysomes was carried out using molecular modelling techniques. Where, the surface electrostatic potential showed to play a role during the carboxysome self-assembly process. The second strategy aimed to test LysR independent and LysR dependent constitutive transcription of the *cbb* operons. The singular mutation of CbbR (A167V) proved to be advantageous for enhancing CO_2 fixation possibly due to the constitutive expression of both *cbb* operons and absence of catabolite feedback.

Acknowledgments

First, I would like to thank my supervisors Katalin Kovács, Naglis Malys and Nigel Minton, for all of the help, support and advice they have given me. Special thanks to Katalin, I learnt a lot from you, you helped me to grow and develop as a scientist and I really appreciate the time you devoted for scientific discussions during the past four years.

I would like to extend my gratitude to Christof Jager for his support and direction in the molecular dynamics project and for hosting me in his group. Thank you, Marko Hanzevacki for the time you devoted to teaching me and your presence as source of intel in the world of molecular modelling, despite the distance.

I would also like to express my gratitude to everyone in B67 for their friendship and support, with a special thanks to Marco Garavaglia for his kindness, his willingness to always help and for his scientific insights. I couldn't have made it through my PhD without sharing my day with some great people, who lifted my spirits in a period full of great challenges. Thank you.

Finally, I would like to thank my family. Especially, my mum for all the sacrifices she made to put me through university and the love she pours in our calls. Even when separated, we are close. Thank you, Cameron for your never ending will to help, for your care in times of distress, for sharing this experience with me and for your love I can always rely on.

Table of contents

Declaration *3*

Abstract *4*

Acknowledgments *5*

List of figures *10*

List of tables *13*

Abbreviations *14*

Chapter 1 – Theoretical background *16*

1.1 Carbon: building block of life and role in global warming **16**

1.2 *Cupriavidus necator* H16 ..... **17**

1.3 CO_2 assimilation pathways **18**

1.4 Genomic organisation of *cbb* operons in *C. necator* H16 ..... **25**

1.5 Regulation of Calvin-Benson-Bassham pathway in *C. necator* H16 ..... **27**

1.6 RubisCO **29**

1.6.1 Multiple forms of RubisCO 30

1.6.2 RubisCO active site and catalysis 31

1.6.3 RubisCO assembly and metabolic repair 34

1.7 Carbonic anhydrases and hydrogenases in *C. necator* H16 ..... **37**

1.7.1 Carbonic anhydrases 37

1.7.2 Hydrogenases 38

1.8 Carboxysomes: metabolic modules for carbon fixation **40**

1.8.1 Carboxysomes types with different components and evolution 43

1.8.1.2 Structure and assembly of β-carboxysome 44

1.8.1.1 Structure and assembly of α-carboxysome 46

1.8.2 Caboxysomal carbonic anhydrase comparative analysis 49

1.8.3 Carboxysomes bioengineering 51

1.9 Aims of the project **52**

Chapter 2 - Materials and Methods *53*

2.1 Bacterial strains, media, and growth conditions **53**

2.1.1 Bacterial strains 53

2.1.2 General chemicals 54

2.1.3 Media 54

2.1.3.1 Lysogeny broth medium 54

2.1.3.2 Low salt lysogeny broth (LSLB) 54

2.1.3.3 Hannahan's Broth (SOB Medium) 54

2.1.3.4 Schlegel minimum medium (MM) 54

2.1.3.5 *Li et al.* 2012 minimal medium (L-MM) 55

2.1.4 Culture growth 56

2.1.5 Bacterial storage 56

2.2 Antibiotics **57**

2.3 DNA manipulation **57**

2.3.1 Plasmids....57
2.3.2 Oligonucleotide primers....62
2.3.3 Isolation of plasmid DNA....72
2.3.4 Isolation of chromosomal DNA....72
2.3.5 PCR analysis of DNA....72
2.3.6 Restriction analysis of plasmid DNA....72
2.3.7 Agarose gel electrophoresis....73
2.3.8 Purification of PCR products *via* gel extraction....73
2.3.9 DNA quantification....73
2.3.10 HiFi DNA Assembly....74
2.3.11 Sequencing of DNA constructs....74

2.4 Transformation of bacterial strains....74
2.4.1 Preparation of *E. coli* electrocompetent cells....74
2.4.2 Electroporation of *E. coli* electrocompetent cells....75
2.4.3 *E. coli* transformation using $CaCl_2$....75
2.4.4 Preparation of *C. necator* H16 electrocompetent cells (Ehsaan M, unpublished)....76
2.4.5 Electroporation of *C. necator* H16 electrocompetent cells (Ehsaan M, unpublished)....76
2.4.6 Mobilisation of plasmids from *E.coli* to *C. necator H16* by conjugation....77
2.4.7 Whole genome sequencing....78

2.5 Plasmids and strain construction....79
2.5.1 Construction of deletion and insertion plasmids....79
2.5.2 Construction of *C. necator* H16 allelic mutants....79
2.5.3 Construction of expression plasmids....80
2.5.3.1 Construction of expression plasmids for the evaluation of an inducible orthogonal expression system in *E. coli* and C. *necato*r H16....80
2.5.3.2 Construction of plasmids for the heterologous expression of *cso* operon from *H. naepolitanus* in *E. coli* and *C. necator* H16....81
2.5.3.3 Construction of expression plasmids with chimeric *cso* operon replacing *cbbLS* genes from *H. neapolitanus* with *C. necator* wild type and mutant form....83

2.6 Protein expression and analysis....85
2.6.1 Growth conditions for protein analysis....85
2.6.2 Denaturing polyacrylamide SDS-PAGE gel electrophoresis analysis....85
2.6.3 Coomassie Blue staining....86
2.6.4 MALDI-TOF mass spectrometry analysis....86

2.7 Transmission electron microscopy (TEM) analysis....87
2.7.1 Sample preparation using iPGel Kit (GenoStaff®, Japan) and TEM imaging....87
2.7.2 Sample preparation and Cryo-TEM tomography analysis....88

2.8 RFP Fluorescence assays....88

2.9 Growth assay in organo-autotrophic conditions in microbioreactor (BioLector Pro®)....89

2.10 Growth assay in autotrophic conditions in serum bottles....90

2.11 Fermentation in autotrophic and mixotrophic conditions in DASGIP® Parallel Bioreactor System....90

2.12 HPLC sample preparation and analysis....92

2.13 PHA quantification using GC-MS 92

2.14 Bioinformatics 94
2.14.1 Plasmid map analysis 94
2.14.2 Sequence databases 94
2.14.3 BLAST analysis 94
2.14.4 Next generation Sequencing (NGS) analysis 94

2.15 Molecular modelling computational methods 94
2.15.1 Model systems 94
2.15.2 Electrostatic potential 95
2.15.3 Force filed parameters 95
2.15.4 Molecular dynamic (MD) simulations 96
2.15.5 Analysis 96

Chapter 3 - Characterisation of the Halothiobacillus neapolitanus cso operon in Cupriavidus necator H16 98

3.1 Introduction 99

3.2 Results 103
3.2.1 Heterologous expression and formation of carboxysomes in *E. coli* and *C. necator* H16 103
3.2.1.1 Expression of *cso* operon in *E. coli* 103
3.2.1.2 Expression of *cso* operon in *C. necator* H16 106
3.2.1.3 Imaging of *E. coli* strains expressing genes for carboxysomes using transmission electron microscopy (TEM) 109
3.2.1.4 Imaging of *C. necator* strains expressing genes for carboxysomes using transmission electron microscopy (TEM) 111
3.2.2 Functional characterisation of carboxysomes in *C. necator* strains 115
3.2.3 Heterologous expression and formation of carboxysomes with RubisCO activases in *E. coli* and *C. necator* 122
3.2.3.1 Expression of *cso* operon with RubisCO activases in *E. coli* 123
3.2.3.2 Expression of *cso* operon with RubisCO activases in *C. necator* 123
3.2.3.3 Imaging of *C. necator* strains co-expressing genes for carboxysomes and RubisCO activases, using transmission electron microscopy (TEM) 129
3.2.3.3.1 TEM 129
3.2.3.3.2 Cryo-TEM tomography 133
3.2.4 Function of carboxysomes with activases in a heterologous host 134

3.3 Discussion 140

3.4 Future work 145

Chapter 4 - Biological relevance of RubisCO surface charge for recruitment and incorporation in α-carboxysomes: Insights from computational modelling 146

4.1 Introduction 147

4.2 Results 149
4.2.1 Electrostatic potential of *H. naepolitanus* and *C. necator* H16 RubisCO in gas phase 149
4.2.2 Molecular modelling of *H. naepolitanus* and *C. necator* H16 wild type and mutant RubisCO 152
4.2.3 Functional complementation of *C. necator* ΔΔ*cbbLS* mutant by *CnMut*RubisCO in autotrophic conditions 171

4.3 Discussion 174

4.4 Future work 180

Chapter 5 – Deregulation of the cbb operon in C. necator H16 for improved carbon fixation rates 181

5.1 Introduction 182

5.2 Results 185

5.2.1. Construction of a reporter system for evaluating an inducible, orthogonal expression cassette in *C. necator* H16 185

5.2.1.1 Investigating expression levels of the inducible, orthogonal system in *E. coli* and *C. necator* H16 186

5.2.1.1.1 Evaluation of the orthogonal expression system in *E. coli*: 186

5.2.1.1.2 Evaluation of the orthogonal expression system strength in *C. necator* H16 190

5.2.2 Biological effect of constitutive or inducible expression of the *cbb* operons in *C. necator* mutants 199

5.2.2.1 Evaluation of mutant strains with constitutive expression of the *cbb* operon under organo-autotrophic conditions in microbioreactor (BioLector Pro®) 201

5.2.2.2 Evaluation of mutant strains with inducible expression of the *cbb* operon under organo-autotrophic conditions in microbioreactor (BioLector Pro®) 206

5.2.2.3 Fermentation of the *C. necator* CbbR A167V mutant strain under autotrophic conditions (DASGIP® Parallel Bioreactor System) 208

5.2.2.4 Fermentation of the *C. necator* CbbR A167V mutant strain under mixotrophic conditions (DASGIP® Parallel Bioreactor System) 211

5.3 Discussion 216

5.4 Future work 219

Chapter 6 – General discussion 221

6.1 General discussion and future directions 221

6.2 Concluding remarks 228

References 230

Appendix 261

List of figures

Figure 1.1: Schematic outline of the Calvin-Benson-Bassham cycle..... 19

Figure 1.2: Schematic outline of two pathways: the reductive TCA cycle and the dicarboxylate/4-hydroxybutyrate cycle.....20
Figure 1.3: Schematic outline of the reductive acetyl-CoA pathway 21
Figure 1.4: Schematic outline of two pathways: the 3-hydroxypropionate pathway and the 3-hydroxypropionate/4-hydroxybutyrate cycle.....22
Figure 1.5: Schematic outline of the reductive glycine pathway 24
Figure 1.6: Schematic view of *cbb* operons of *C. necator* H16.....26
Figure 1.7: Schematic representation of the intercistronic region between *cbbR* gene and *cbb* operon and role of DNA bending and LTTR binding at the ABS and RBS sites in LTTR-dependent transcriptional activation 28
Figure 1.8: Hexadecameric structure of RubisCO with details about active site and CsoS2 peptide interaction site.....33
Figure 1.9: Schematic view of the *cso* operon from *H. naepolitanus* and of the link between carboxysome CO_2 fixation and the central carbon metabolism. 42
Figure 3.1: Expression of *cso* operon under inducible promoter in *E. coli*.....104
Figure 3.2: Expression of *cso* operon under constitutive promoter in *E. coli* 105
Figure 3.3: Expression of *cso* operon under inducible promoter in *C. necator*.....107
Figure 3.4: Expression of *cso* operon under constitutive promoter in *C. necator*.....108
Figure 3.5: TEM imaging of *E. coli* strains expressing *cso* operon under inducible and constitutive promoter 110
Figure 3.6: TEM imaging of *C. necator* strains expressing *cso* operon under inducible promoter 113
Figure 3.7: TEM imaging of *C. necator* strains expressing *cso* operon under constitutive promoter. 114
Figure 3.8: Functional complementation of *Hn* and *Cn*RubisCO in *C. necator* Δ*cbbLS* deletion strain 116
Figure 3.9: Functional complementation of *Hn* and *Cn*RubisCO in *C. necator* Δ*phaCAB* Δ*cbbLS* deletion strain 118
Figure 3.10: Functional complementation of *C. necator* Δ*cbbLS* Δ*panC* and *C. necator* Δ*phaCAB* Δ*cbbLS* Δ*panC* with *cso* operon in organo-autotrophic conditions.....120
Figure 3.11: Functional complementation of *C. necator* Δ*phaCAB* Δ*cbbLS* Δ*panC* with *cso* operon under constitutive and inducible promoter in autotrophic conditions.....120

Figure 3.12: Expression of *cso* operon with or without RubisCO activases in *C. necator* ΔΔ*cbbLS* Δ*panC* in LB media 124
Figure 3.13: Expression of *cso* operon with or without RubisCO activases in *C. necator* ΔΔ*cbbLS* Δ*panC* grown in 0.4% SG MM 126
Figure 3.14: Expression of *cso* operon with or without RubisCO activases in *C. necator* ΔΔ*cbbLS* Δ*panC* grown in 80mM Na-formate MM 128
Figure 3.15: TEM imaging of *C. necator* Δ*panC* strains expressing *cso* operon with and without RubisCO activases..... 131
Figure 3.16: TEM imaging of *C. necator* Δ*phaCAB ΔpanC* strains expressing *cso* operon with and without RubisCO activases..... 132
Figure 3.17: Growth curve of *C. necator* ΔΔ*cbbLS* Δ*panC* expressing the *cso* operon with and without RubisCO activases in organo-autotrophic conditions..... 135
Figure 3.18: Functional complementation of adapted *C. necator* ΔΔ*cbbLS* Δ*panC* expressing the *cso* operon with and without RubisCO activases in organo-autotrophic conditions 136
Figure 3. 19: Growth curve of *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* expressing the *cso* operon with and without RubisCO activases in organo-autotrophic conditions 138
Figure 3.20: Functional complementation of adapted *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* expressing the *cso* operon with and without RubisCO activases in organo-autotrophic conditions..... 139
Figure 4.1: Representation of the electrostatic surface potential representation of *Hn*RubisCO and *Cn*RubisCO large and small subunit calculated in gas phase 149
Figure 4. 2: *Cn*RubisCO dimeric protein structure with highlighted mutation sites..... 150
Figure 4.3: Comparison between electrostatic potential of *Hn*RubisCO and *CnMut*RubisCO..... 151
Figure 4.4: RMSD analysis of *Hn*, *Cn* and *CnMut*RubisCO at three ionic strengths (20mM, 150mM and 1000mM) 154
Figure 4.5: RMSF analysis of *Hn*, *Cn* and *CnMut*RubisCO at three ionic strengths (20mM, 150mM and 1000mM) 156
Figure 4.6: Protein structure of *Hn*, *Cn* and *CnMut*RubisCO showing loop flexibility 158
Figure 4.7: Hexadecameric structure of *Hn*RubisCO showing CbbL flexible loop location..... 159
Figure 4.8: RDF analysis between Cl^- and positively charged amino acids of *Hn*, *Cn* and *CnMut*RubisCO in three ionic strengths (20mM, 150mM and 1000mM) 161
Figure 4.9: RDF analysis between Na^+ and negatively charged amino acids of *Hn*, *Cn* and *CnMut*RubisCO in three ionic strengths (20mM, 150mM and 1000mM) 163
Figure 4.10: Ion Density Analysis of anions and cations during MD simulation of *Hn*, *Cn* and *CnMut*RubisCO at 20 mM 166
Figure 4.11: Ion Density Analysis of anions and cations during MD simulation of *Hn*, *Cn* and *CnMut*RubisCO at 150 mM 168

Figure 4.12: Solvent accessible and interaction surface of *Hn*RubisCO in the hexadecameric structure.....170
Figure 4.13: Adaptation experiment of *C. necator* ΔΔ*cbbLS* mutant strain carrying *CnMut*RubisCO in autotrophic conditions.....172
Figure 4.14: Complementation experiment of *CnMut*RubisCO in *C. necator* ΔΔ*cbbLS* mutant strain 173
Figure 5.1: Schematic representation of the *cbb* operon and molecular engineering strategy.....183
Figure 5.2: Growth and fluorescence measurements of *E. coli* carrying either the arabinose inducible promoter or the orthogonal expression system in LB media.....188
Figure 5.3: Growth and fluorescence measurements of *E. coli* carrying control plasmids in LB media.....189
Figure 5.4: Growth and fluorescence measurements of *C. necator* carrying either the arabinose inducible promoter or the orthogonal expression system in LB media.....192
Figure 5.5: Growth and fluorescence measurements of *C. necator* carrying control plasmids in LB media.....193
Figure 5.6: Growth and fluorescence measurements of *C. necator* carrying either the arabinose inducible promoter or the orthogonal expression system in LB 1% Glucose media.....195
Figure 5.7: Growth and fluorescence measurements of *C. necator* carrying control plasmids in LB 1% Glucose media.....196
Figure 5.8: Growth and fluorescence measurements for the rescue plasmid assay of the orthogonal expression system in *E. coli*.....198
Figure 5.9: Growth of *C. necator* P_{trp} CBB allelic mutant under organo-autotrophic conditions after overnight incubation in 0.4% SG MM.....202
Figure 5.10: Growth of *C. necator* CbbRA167V allelic mutant under organo-autotrophic conditions after overnight incubation in 0.4% SG MM.....203
Figure 5.11: Growth of *C. necator* P_{trp} CBB allelic mutant under organo-autotrophic conditions after overnight incubation in the same conditions.....204
Figure 5.12: Growth of *C. necator* CbbRA167V allelic mutant under organo-autotrophic conditions after overnight incubation in the same conditions.....205
Figure 5.13: Growth of *C. necator* P_{BAD} CBB allelic mutant under organo-autotrophic conditions after overnight incubation in 0.4% SG MM.....207
Figure 5.14: Growth and gas consumption profile of *C. necator* CbbR A167V under autotrophic conditions 210
Figure 5.15: Growth, VFAs consumption and PHAs composition of *C. necator* CbbR A167V under aerobic mixotrophic conditions.....212
Figure 5.16: Growth and VFAs consumption of *C. necator* H16 and *C. necator* CbbR A167V during two-phase fermentation.....214
Figure 5.17: Gas consumption profile of *C. necator* H16 and *C. necator* CbbR A167V during two-phase fermentation. 215

Appendix 1: Expression of *cso* operon under inducible promoter in *C. necator* Δ*phaCAB* 261
Appendix 2: Expression of *cso* operon under constitutive promoter in *C. necator* Δ*phaCAB* 262
Appendix 3: Expression of *cso* operon with or without RubisCO activases in *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* in LB media 263
Appendix 4: Expression of *cso* operon with or without RubisCO activases in *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* grown in 0.4% SG MM 264
Appendix 5: Expression of *cso* operon with or without RubisCO activases in *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* grown in 80mM Na-formate MM 265
Appendix 6: Ion Density Analysis of anions and cations during MD simulation of Hn, Cn and CnMutRubisCO at 1000 mM..... 266

List of tables

Table 2.1: List of strains used in this study. 53
Table 2.2: List of antibiotics used in this study 57
Table 2.3: List of plasmids used in this study. 57
Table 2.4: List of oligonucleotides used in this study 62

Abbreviations

µL	Microliter
µm	Micrometer
AAA	ATPase Associated with diverse cellular Activities
AD	Anaerobic digestion
ACK	Acetate kinase
ACS	Acetyl-CoA synthase
ATP	Adenosine triphosphate
BMCs	Bacterial microcompartments
CA	Carbonic anhydrases
CBB	Calvin-Benson-Bassham
CBs	Carboxysomes
CCMs	CO_2-concentrating mechanisms
CODH/ACS	Carbon monoxide dehydrogenase or acetyl-Co-A synthase
FDH	Formate dehydrogenase
fs	Femtosecond
GC-MS	Gas chromatography-mass spectrometry
GCS	Glycine cleavage/synthase
GR	Glycine reductase
G3P	Glyceraldehyde-3-P
HPLC	High performance liquid chromatography
IDA	Ion density analysis
kb	Kilobases
kDa	Kilodalton
L	Litre
LB	Lysogeny broth
LSLB	Low salt lysogeny broth
LTTR	LysR-type transcriptional regulator
LOC	Limiting oxygen conditions
M	Molar
MBH	Membrane bound hydrogenase
MD	Molecular dynamics
MEP	Molecular Electrostatic Potential
min	Minute
mL	Millilitre
mm	Millimetre
mM	Millimolar
MTP	Microtiter plate
NAD+/H	Nicotinamide adenine dinucleotide
NADP/H	Nicotinamide adenine phosphate dinucleotide
ns	Nanosecond

nmRC	Nanoscale and Microscale Research Centre, University of Nottingham
O/N	Grown overnight
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
PFOR	Pyruvate ferredoxin oxidoreductase
PHAs	Polyhydroxyalkanoates
PPI	Ribose-5-phosphate isomerase
PRK	Phosphoribulokinase
ps	Picosecond
RDF	Radial distribution function
RFU	Relative Fluorescence Units
rGly	Reductive glycine pathway
RMSD	Root mean squared deviation
RMSF	Root mean square fluctuation
RuBP	Ribulose-1,5-biphosphate
RubisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
s	Second
SG MM	Sodium gluconate minimal media
SH	Soluble hydrogenase
TCA	Tricarboxylic acid
TEM	Transmission electron microscopy
TPI	Triosephosphate isomerases
VFAs	Volatile fatty acids
VWA domain	Von Willebrand factor A domain
XuBP	Xylulose-1,5-bisphosphate
2-PGA	2-phospho-glycolate
3-PGA	3-phospho-glycerate
3-HP	3-hydroxypropionate
3-HP pathway	3-hydroxypropionate pathway
3-HP/4-HB	3-hydroxypropionate/4-hydroxybutyrate

Chapter 1 – Theoretical background

1.1 Carbon: building block of life and role in global warming

Carbon is the primary building block of life on earth. In nature, carbon occurs in a variety of organic and inorganic forms and its cycles are an important driving force for circulating C between the atmosphere, hydrosphere, lithosphere and biosphere. One of the most important forms in which carbon exists is carbon dioxide (CO_2), that serves as the only essential carbon source for plants and autotrophic bacteria. During this process, carbon is converted into its organic form and incorporated into biomass. Interest in carbon dioxide fixing bacteria has been steadily increasing since 1940. Early work was largely confined to the study of CO_2 consumption in different conditions rather than on unravelling metabolic pathways involved into CO_2 fixation. The latter was reported for the first time in 1940 by Wood and Werkman during *Propionibacterium pentosaceum* glycerol fermentation (Wood and Werkman, 1940). Later, CO_2 assimilation was proven to be a specific pathway by growing numerous heterotrophic bacteria on heavy carbon dioxide (Slade *et al.*, 1942). Since then, several investigations have helped elucidate possible mechanism and pathways involved in the carbon incorporation process as well as, understanding diversity and taxonomy of CO_2 fixing autotrophic microorganisms (Saini *et al.*, 2011).

In recent decades, it became clear that the carbon cycle on our planet was unbalanced, as a consequence of the increased CO_2 emissions into the atmosphere caused by fossil fuel combustion and changes in land use. Fossil fuels account for 85% of total emissions and the land usage change for about 15% (Richardson *et al.*, 2011). Carbon dioxide among other gasses is the most abundant component of greenhouse gas (GHG), and its involvement in the raising temperature on the planet has been known for a long time (Arrhenius, 1896; Qiancheng Ma,

1998). Recent reports on climate change emphasise the need prompt interventions for its mitigation through limiting or preventing GHG emissions, as well as activities that reduce their concentrations in the atmosphere (Edernhofer *et al.*, 2014; IPCC, 2018).

The urgent need to both to reduce CO_2 emissions and its atmospheric concentrations underpin the need to exploit tools from the natural world that enable carbon fixation and its use for the production of bioenergy, microbial biodiesel, biohydrogen and biodegradable plastics (Rittmann, 2008; Wang *et al.*, 2008).

In this scenario, the selection of an appropriate microbe able to satisfy different demands as a model organism and as possible industrial chassis, is paramount.

1.2 *Cupriavidus necator* H16

In the past decade, several studies highlighted the great potential of strain *Cupriavidus necator* strain H16 (*Ralstonia eutropha*; formerly *Alcaligenes eutrophus*) as a model organism, and as a chassis for biotechnological production, especially bioplastics (Abascal *et al.*, 2005; Brigham *et al.*, 2012; Grousseau, Lu, *et al.*, 2014; Chen *et al.*, 2015). A Gram-negative facultative chemolithoautotrophic, β-proteobacterium, *C. necator* is characterised by a genome exhibiting a high degree of flexibility in its carbon and energy metabolism (Pohlmann *et al.*, 2006; Cramm, 2009; Little *et al.*, 2019). It is perhaps best known for its ability to grow in aerobic conditions using CO_2 and H_2 as sole sources of carbon and energy, but can also grow heterotrophically on many different organic compounds such as sugars, organic acids, fatty acids, or plant oils (Kusian and Bowien, 1997; Shimizu *et al.*, 2013). The organism is able to alternate between heterotrophic or autotrophic growth modes and store energy in form of polyhydroxyalkanoates (PHAs) (Boyle and Morgan, 2011; Shimizu *et al.*, 2013, 2015). Industrial scalability of growth conditions and genetic amenability make *C. necator* an attractive host for the

production of bioplastic and for biofuel industries (Shimizu *et al.*, 2015; Torella *et al.*, 2015).

1.3 CO_2 assimilation pathways

A wide variety of prokaryotes, dispersed among both the archaeal and bacterial domains, contribute to the formation of organic matter through the biosynthetic process of CO_2 fixation (Kusian and Bowien, 1997; Saini *et al.*, 2011). To date, seven main pathways are known to be utilised by microbes for carbon dioxide assimilation.

The first to be discovered about 66 years ago (Thauer, 2007) was the Calvin cycle. (Bassham *et al.*, 1954). In honour of these pioneer researchers, the Calvin cycle became known as the Calvin-Benson-Bassham (CBB) pathway. Besides, its discovery in plants it is reported to be widespread among eukaryotes such as algae, and to occur within prokaryotes in green-sulphur bacteria (Ivanovsky *et al.*, 1999); α-proteobacteria (Meijer, 1994; Fuhrmann *et al.*, 2003); β-proteobacteria, such as *C. necator* (Saini *et al.*, 2011) and cyanobacteria (Otsuki, 2001). The CBB pathway is also referred to as the reductive pentose pathway as a consequence of the enzyme pool being shared with the oxidative pentose pathway. As mentioned above, it is the main route used for assimilation and fixation of CO_2 in proteobacteria such as *C. necator* and anaerobic phototrophs (Gibson and Tabita, 1996; Kusian and Bowien, 1997; Jessup M. Shively *et al.*, 1998). The first step of CO_2 fixation is carried out by RubisCO (Ribulose-1,5-bisphosphate carboxylase/oxygenase), which carboxylates the substrate Ribulose-1,5-biphosphate (RuBP). The unstable end-product of the reaction hydrolyses to form two molecules of 3-P-glycerate (3-PGA). Subsequently, it is reduced into glyceraldehyde-3-P (G3P) by the consecutive action of phosphoribulokinase (PGK) and glyceraldehyde 3-P dehydrogenase (GAP; Figure 1.1) (Kusian and Bowien, 1997). For every three CO_2 molecules entering the cycle, six triose phosphates are produced. Of

these, one is directed toward the tricarboxylic acid (TCA) cycle and five molecules will be used to regenerate the substrate ribulose-5-P. The former is phosphorylated in a final step to ribulose-1,5-bis-P (RuBP) by phosphoribulokinase (PRK) (Figure 1.1) (Bowien and Kusian, 2002; Schwartz *et al.*, 2009).

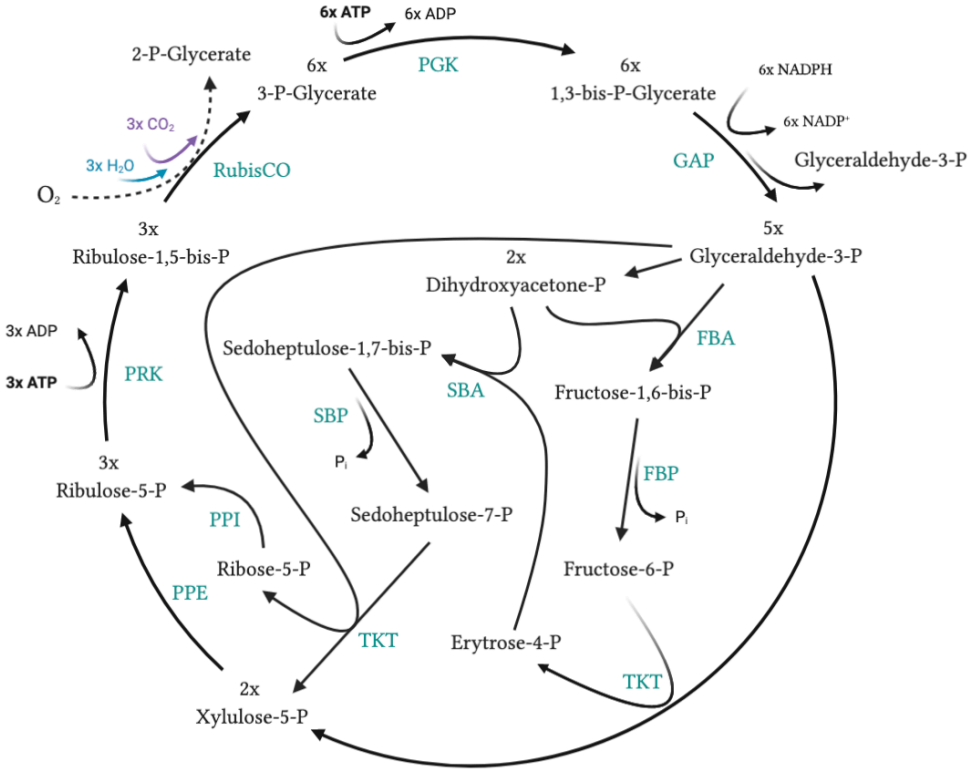


Figure 1.1: Schematic outline of the Calvin-Benson-Bassham cycle. Image describing the details of reactions and enzymes involved in the CBB pathway. The following abbreviations are used for enzymes in uppercase (green): Fructose-1,6-bisphosphate aldolase (FBA); Fructose-1,6-bisphosphatase (FBP); Glyceraldehyde-3-phosphate dehydrogenase (GAP); 3-phosphoglycerate kinase (PGK); Phosphoglycolate phosphatase (PGP); Pentose-3-phosphate 3-epimerase (PPE); Pentose-3-phosphate isomerase (PPI); Phosphoribulo-kinase (PRK); Ribulose-1,5-bisphosphate carboxylase/oxigenase (RubisCO); Sedoheptulose-1,7-bisphosphate aldolase (SBA); Sedoheptulose-1,7-bisphosphatase (SBP); Transketolase (TKT); Trisephosphate isomerase (TPI). Image created in BioRender.com.

In 1966, another pathway named the reductive tricarboxylic acid (TCA) cycle (Figure 1.2, in black) was found to be operating in *Chlorobium limicola* (formerly, *Chlorobium thiosulfatophilum*) (Evans *et al.*, 1966). Most enzymes involved in this cycle overlap with the oxidative TCA cycle, with the exception of ATP-citrate synthase, 2-oxoglutarate synthase and fumarate reductase that catalyse the reactions in the reductive direction. This reductive pathway has been

mainly adopted by anaerobic and microaerobic microorganisms (Hügler *et al.*, 2005; Saini *et al.*, 2011).

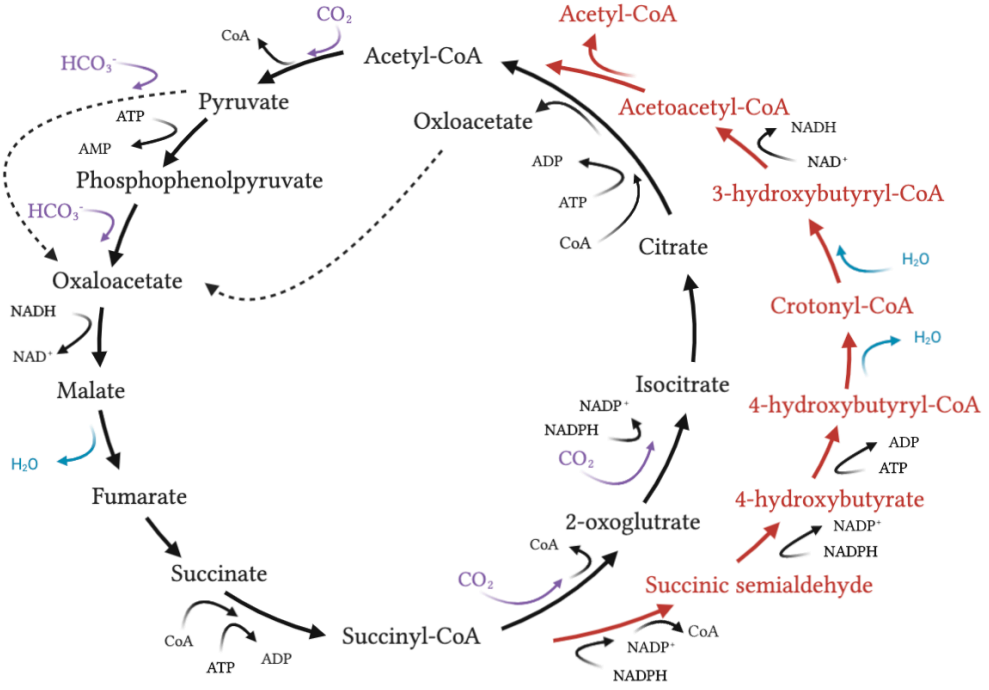


Figure 1.2: Schematic outline of two pathways: the reductive TCA cycle and the dicarboxylate/4-hydroxybutyrate cycle. Image describing a diagram of two pathways: the reductive TCA cycle (in black) and the dicarboxylate/4-hydroxybutyrate cycle (in red). Image created in BioRender.com.

The reductive acetyl Co-A pathway (Figure 1.3) is a third pathway that can mediate CO_2 fixation. It is also known as Wood-Ljungdahl pathway and through it autotrophic acetogens and methanogens synthesize acetic acid and methane from inorganic carbon sources. Unlike the other fixation pathways, there is no regeneration of the primary CO_2 acceptor and carbon dioxide enters the pathway through two different paths called the western and eastern branches (Ragsdale and Pierce, 2008; Saini *et al.*, 2011). The western branch involves a single step process operated by the key anoxic enzyme CODH/ACS (carbon monoxide dehydrogenase or acetyl-Co-A synthase), where a singular molecule of carbon dioxide is incorporated into the carboxylic group of an acetate molecule (Ragsdale *et al.*, 1983). While, the eastern branch (also known as methyl branch) reduces carbon dioxide through many steps into the methyl group of

an acetate molecule (Figure 1.3). For reviews see (Ragsdale, 2008; Ragsdale and Pierce, 2008).

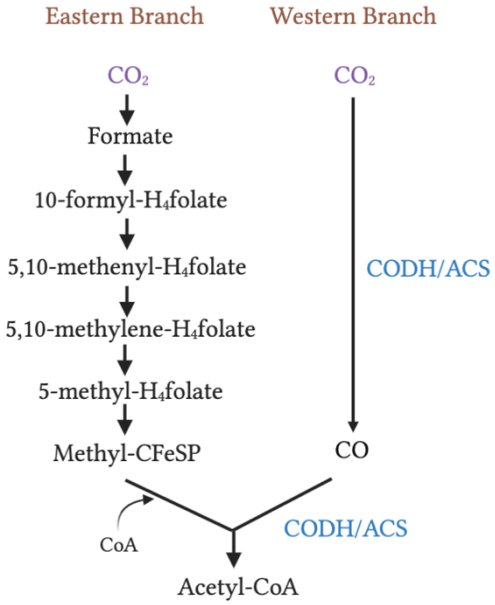


Figure 1.3: Schematic outline of the reductive acetyl-CoA pathway. Image describing the two branches of the reductive acetyl-CoA pathway. The following abbreviations are used for enzymes in uppercase (blue): Carbon monoxide dehydrogenase or acetyl-Co-A synthase (CODH/ACS). Image created in BioRender.com.

The fourth pathway known for CO_2 incorporation is named 3-hydroxypropionate pathway (3-HP pathway; Figure 1.4 in black) (Herter *et al.*, 2002). It is a cyclic pathway involving the regeneration of the starting molecule, acetyl-CoA. At each turn of the cycle two molecules of CO_2 are fixed into one molecule of glyoxylate. The glyoxylate is then converted into pyruvate by a dedicated pathway (the glyoxylate assimilation pathway). The pathway was discovered in the green non-sulfur phototrophic bacterium *Chloroflexus aurantiacus* (Holo, 1989) and seems to be present only in organisms that belong to the Chloroflexaceae phylum (Klatt *et al.*, 2007), one of the five major lineages of phototrophic bacteria. A related pathway operates in some acidophilic archaebacteria of the phylum *Crenarchaeota,* such as A*cidianus brierleyi, Sulfolobus metallicus* and *Metallosphaera sedula* (see 3-hydroxypropanoate/4-hydroxybutanate cycle).

The 3-Hydroxypropionate/4-hydroxybutyrate (3-HP/4-HB) cycle (Figure 1.4), the fifth pathway for CO_2 assimilation was first reported in aerobic autotrophic organisms (Berg *et al.*, 2007). This cyclic pathway has been split into two main parts. The first one operates the transformation of acetyl-CoA and two molecules of bicarbonate into succinyl-CoA (Figure 1.4, in black), while the other part forms two molecules of acetyl-CoA from succinyl-CoA (Figure 1.4, in green). Despite the similarity with 3-HP pathway, the enzymes involved in the catalytic reactions are not the same. In this pathway, for each turn of the cycle two molecules of bicarbonate are fixed, but the carbons are not exported as glyoxylate. Rather, they are converted directly to acetyl-CoA by a route that is similar to that used by *Clostridium* for fermentation of succinate (Saini *et al.*, 2011).

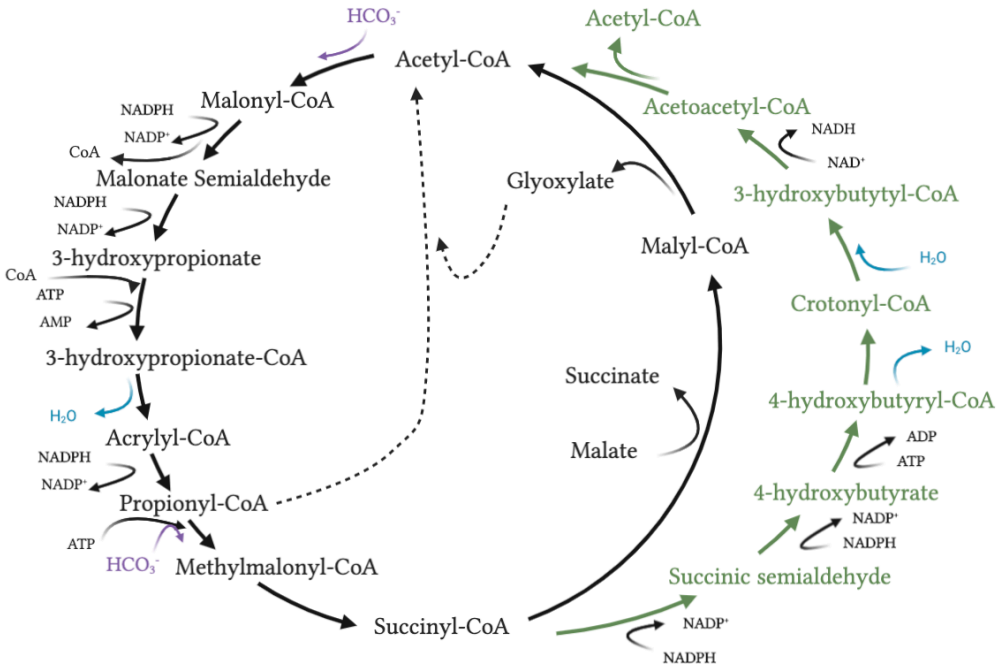


Figure 1.4: Schematic outline of two pathways: the 3-hydroxypropionate pathway and the 3-hydroxypropionate/4-hydroxybutyrate cycle. Image describing a diagram of two pathways: the 3-hydroxypropionate pathway (in black) and the 3-hydroxypropionate/4-hydroxybutyrate cycle (in green). Image created in BioRender.com.

The dicarboxylate/4-hydroxybutyrate cycle (Figure 1.2; in red) was recently associated with the autotrophic growth of *Ignicoccus hospitalis* (*Desulfurococcales*) (Huber, Gallenberger, *et al.*, 2008). Its name is due to the similarity with reductive TCA cycle and 3-HP/4-

HB pathway. The main difference between this and the 3-hydroxypropionate/4-hydroxybutyrate cycle is in the formation of succinyl-CoA. During the first reaction of the dicarboxylate/4-hydroxybutyrate cycle the transformation of acetyl-CoA and two inorganic carbon to succinyl-CoA is achieve by pyruvate synthase and pyruvate carboxylase (Figure 1.2, black dotted lines). For reviews see (Huber, Gallenbergger, *et al.*, 2008; Saini *et al.*, 2011).

The most recently discovered autotrophic pathway was first proposed as a synthetic pathway and named the reductive glycine pathway (rGly) (Cotton *et al.*, 2018). Only in 2020, it was re-discovered this times as a natural and efficient CO_2 assimilation pathway used by the sulphate-reducing bacterium *Desulfovibrio desulfuricans* G11 (Sánchez-Andrea *et al.*, 2020). In this pathway one molecule of CO_2 is first reduced to formate, which is then reduced further and condensed with a second molecule of CO_2 ultimately forming glycine. During the autotrophic growth of *D. desulfuricans* CO_2 assimilation proceeds primarily through the glycine reduction route (Figure 1.5; in black) rather than the serine route (Figure 1.5; in blue). Furthermore, under the same conditions *via* the glycine route, the acetyl-phosphate is converted into acetyl-CoA, using the combined activity of acetate kinase (ACK) and acetyl-CoA synthase (ACS). A third molecule of CO_2 is incorporated with acetyl-CoA to form pyruvate. The rGly is one of the most efficient pathways for CO_2 fixation, due to its low ATP demand (1 to 2 ATP per pyruvate molecule produced), and the efficiency, as well as, the thermodynamic feasibility of all of the steps involved. For this reason, the synthetic pathway *via* serine route has already been successfully engineered in several hosts strains: *Escherichia coli*, *C. necator* and yeast (Yishai *et al.*, 2017; Gonzalez De La Cruz *et al.*, 2019; Claassens *et al.*, 2020).

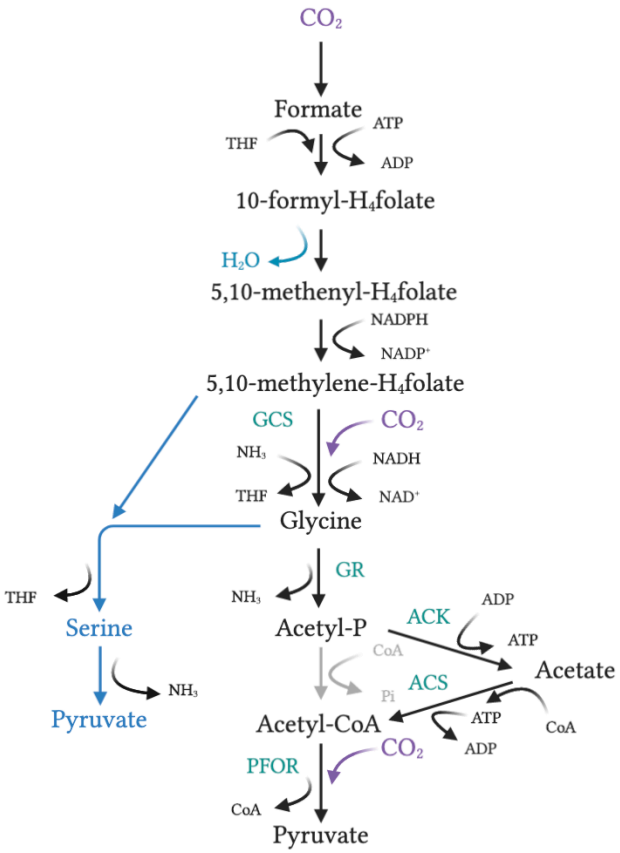


Figure 1.5: Schematic outline of the reductive glycine pathway. Image describing the two branches of the reductive glycine pathway: glycine reduction route (in black) and serine route (in blue). The following abbreviations are used for enzymes in uppercase (green): Acetate kinase (ACK); Acetyl-Co-A synthase (ACS); Glycine cleavage/synthase (GCS); Glycine reductase (GR); Pyruvate ferredoxin oxidoreductase. Image created in BioRender.com

Proteobacterial aerobic chemoautotrophs such as *C. necator* and anaerobic phototrophs (purple bacteria) use the CBB cycle as the principal pathway of carbon assimilation (Gibson and Tabita, 1996; Kusian and Bowien, 1997; Jessup M. Shively *et al.*, 1998). The organisation of the genes differs in distinct bacteria, but the overall presence of one or several *cbb* operons is a common feature (Bowien and Kusian, 2002).

1.4 Genomic organisation of *cbb* operons in *C. necator* H16

The multi replicon genome of *C. necator* is composed of two chromosomes and a megaplasmid (pHG1) (Pohlmann *et al.*, 2006). While, the information regarding housekeeping genes, central carbon and energy metabolism are enclosed in the chromosomes, the megaplasmid codes for a number of additional metabolic pathways (Schwartz *et al.*, 2003; Pohlmann *et al.*, 2006). The genes encoding the enzymes necessary for lithoautotrophic growth of *C. necator* are duplicated in the genome. One copy is situated on chromosome II and an almost identical copy resides on the megaplasmid pHG1 (Figure 1.6) (Kusian and Bowien, 1997). The chromosomal operon (cbb_c) is composed of 13 genes, encoding eight of the ten enzymes required for the CBB cycle, and is 15.2 kilobases (kb) in length. The megaplasmid operon (cbb_p) comprises 12 genes and has a total length of 12.8 kb (Bowien and Kusian, 2002). The enzymes involved into the CBB cycle (genes displayed in Figure 1.6 and pathway in Figure 1.1) reside in the duplicated operons, apart from triosephosphate isomerases (TPI) and ribose-5-P isomerase (PPI) (Gibson and Tabita, 1996; Kusian and Bowien, 1997; Jessup M. Shively *et al.*, 1998). If the structures of the two *cbb* clusters are compared, it is possible to see that the pHG1 operon lacks the *cbbB* gene coding for the formate dehydrogenase-like protein (FDH) (Figure 1.6) (Bowien and Kusian, 2002; Pohlmann *et al.*, 2006). Further details regarding the function of *cbbX* and *ccbY* genes will be discussed in the next section.

The transcription of each operon is driven by a promoter upstream of the *cbbL* gene, which is regulated by the product of the divergently orientated *cbbR* gene at the 5' end of the operon (Figure 1.6). The function of the *cbbR* copy on the megaplasmid is impaired by a 28 bp frameshift deletion near the 5'-end of its coding sequence (Windhövel and Bowien, 1991). Therefore, the regulation of both *cbb* operons is under the influence of *cbbR* chromosomal copy (Kusian *et al.*, 1995).

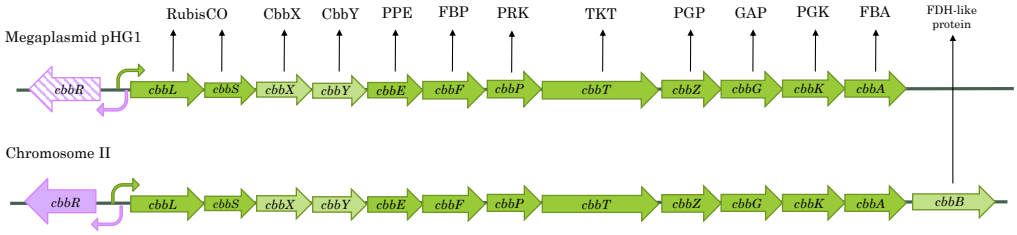


Figure 1.6: Schematic view of *cbb* operons of *C. necator* H16. Image describing the structural *cbb* genes (in green) and the transcriptional regulator (in lilac). The transcriptional regulator on the megaplsmid pHG1 carries a 28 bp deletion, which disable the function (in lilac stripes). The following abbreviations are used for the enzymes in the uppercase: Fructose-1,6-bisphosphate aldolase (FBA); Fructose-1,6-bisphosphatase (FBP); Glyceraldehyde-3-phosphate dehydrogenase (GAP); 3-phosphoglycerate kinase (PGK); phosphoglycolate phosphatase (PGP); Pentose-3-phosphate 3-epimerase (PPE); phosphoribulo-kinase (PRK); Ribulose-1,5-bisphosphate carboxylase/oxigenase (RubisCO); Transketolase (TKT).

1.5 Regulation of Calvin-Benson-Bassham pathway in *C. necator* H16

As a facultative autotroph *C. necator* H16 needs to control expression of its two *cbb* operons to avoid utilizing the CBB cycle when it is not required. The simultaneous transcription of entire operons is driven by a σ^{70} promoter ($P_{cbbL;}$ Figure 1.7 A). Since both operons are simultaneously transcribed, and no alternative internal transcription starts have been identified (Schäferjohann *et al.*, 1996; Jeffke *et al.*, 1999; Bowien and Kusian, 2002), they form a regulon under the control of the chromosomal transcriptional activator CbbR (Windhövel and Bowien, 1991; Kusian *et al.*, 1995).

CbbR is a member of the LysR-type transcriptional regulators (LTTRs) and is the product of the *cbbR* gene which is separated by 167 bp from the chromosomal *cbbL* in *C. necator* (Figure 1.7) (Jeffke *et al.*, 1999; Bowien and Kusian, 2002). Further studies highlighted the differential expression between the promoter-proximal *cbbLS* genes and distal *cbb* genes (reviewed in Bowien and Kusian, 2002; Dangel and Tabita, 2015). Firstly, the molar ratio between RubisCO and PRK was reported to be at least 5:1 in *C. necator* (Leadbeater *et al.*, 1982) and secondly the transcript of RubisCO genes was shown to be more abundant than the distal mRNA of the operon (Schäferjohann *et al.*, 1996). The relative abundance of transcripts is affected by the formation of a mRNA based stem-loop transcription terminator in the intergenic region between *cbbS* and *cbbX* (Schäferjohann *et al.*, 1996). This results in differential gene expression between the polycistronic *cbb* operon (Figure 1.6).

Like the other LTTRs, CbbR is composed of three functional domains (Maddocks *et al.*, 2008). The helix-turn-helix DNA binding domain (DBD) at the N-terminal, mediates the interaction of the regulator with the *cbb* promoter region. The rigid helix linker domain serves to prevent the interaction between the other two domains that are

involved in the oligomerisation process. The C-terminal domain, also known as the recognition domain (RD) or effector domain, is involved in the interaction with the effectors molecules (Kusian and Bowien, 1997; Bowien and Kusian, 2002; Maddocks *et al.*, 2008).

Usually, LTTR binding sites are located in close proximity from the promoter target region. The LTTR typically recognizes a partially symmetric sequence with a T-N_{11}-A motif. In *C. necator,* the intergenic segment of 167 bp encodes for the entire control region of the *cbb* operon. Within this region there are an activation binding site (ABS or A-site) and a distal regulator binding site (RBS or R-site), with which the CbbR interacts (Kusian and Bowien, 1997; Bowien and Kusian, 2002). The A-site overlaps with the -35 box of the *cbb* promoter P_{cbbL}, while both A-site and R-site cover almost the entirety of the *cbbR* promoter P_{cbbR} region (Figure 1.7 A). This is indicative of negative autoregulation (Kusian and Bowien, 1997; Bowien and Kusian, 2002). A-site and R-site are regulated by binding of a CbbR tetramer each, which provokes sharp DNA bending. After the interaction with positive effectors (Figure 1.7 B), the DNA tension is released allowing CbbR tetramer, in ABS position, to contact an RNA polymerase and start CbbL transcription (Maddocks *et al.*, 2008).

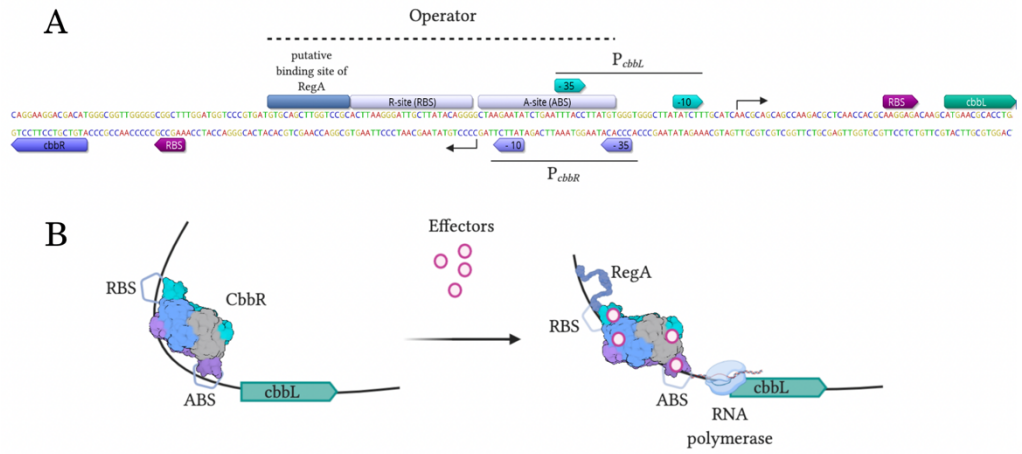


Figure 1.7: Schematic representation of the intercistronic region between *cbbR* gene and *cbb* operon and role of DNA bending and LTTR binding at the ABS and RBS sites in LTTR-dependent transcriptional activation. (A) Details of the operator-promoter region of *C. necator* H16 between the ORF of *cbbL* and *cbbR*. Image created with Geneious R11.1. (B) Details about DNA bend relaxation after the effectors binding to the LTTR tetramer. Image created with BioRender.com.

Phosphoenolpyruvate (PEP) was identified as transcriptional repressor of both operons in *C. necator* (Grzeszik *et al.*, 2000), but a complex regulation mechanism accounting for cellular energy state, was expected to be in place. Recently, homologues of RegA/B system, including functionally relevant motives, were found in *C. necator* (Gruber *et al.*, 2017). It was shown that RegA interaction with CbbR *via* their DBD domain increasing the stability of the DNA/CbbR complex in *R. sphaeroides*, possibly enhancing the transcription rate (Dubbs *et al.*, 2000; Dangel and Tabita, 2015b). Experimental analysis has since revealed the central role of RegA in the transcriptional control of *cbb* operons in *R. sphaeroides* and *R. palustris*, depending on energy levels and redox state of the cell (Dangel and Tabita, 2009, 2015b; Gruber *et al.*, 2017).

1.6 RubisCO

RubisCO is addressed as one of, if not, the key enzyme of the CBB cycle. Despite its relevance, it is not the ideal enzymatic catalyst. The roots of these imperfections reside in the major forces that have acted on RubisCO during evolution. Over the past 3.5 billion years, earth's atmosphere and biosphere evolved and went through dramatic changes in terms of the concentration of its components. The advent of photosynthesis caused a sharp rise in oxygen concentration, starting from the anaerobic environment where RubisCO firstly evolved, to reach 35% oxygen in the late Paleozoic (Berner *et al.*, 2003) and to the present day levels of 21%. Concurrently, there was an important decline in CO_2 concentration to 200 ppm V about 280 mya (Mora *et al.*, 1996), increasing again to the present day 409.8 ppm V on the rise (Berner and Kothavala, 2001; Lindsey, 2020).

This huge change in the environmental conditions might be the cause of RubisCO renowned inefficiency, with slow catalytic turnover rate (k_{cat} = 1 to 12 s^{-1}), low affinity for CO_2 and a wasteful oxygenase side reaction. In fact, RubisCO catalyses the incorporation of oxygen on

the RuBP substrate, producing only one molecule of 3-P-glycerate (3-PGA) and one molecule of 2-P-glycolate (2-PGA) (Bhat *et al.*, 2017). This toxic secondary product is recycled into 3-P-glycerate *via* the glyoxylate pathway, usually involving ATP-dependent reactions and loss of CO_2 (Figure 1.1, dotted line). The detoxification of 2-P-glycolate starts with its conversion to glyoxylate, which can only be redirected to the central carbon metabolism by wasting fixed carbon (one CO_2 per two glyoxylate molecules) (Berg, 2011). Although the precise loss of carbon caused by the glyoxylate assimilation in bacteria is not known, it was estimated to be 21% of the net CO_2 fixation for C3 plants (Cegelski and Schaefer, 2006). Recent analysis highlighted that RubisCO specificity is not the main factor responsible for its efficiency. Rather, it is the trade-off between the carboxylation velocity and CO_2 affinity. There was no observed correlation between O_2 catalysis rate and the other measured kinetic parameters of RubisCO (Savir *et al.*, 2010). As a result, RubisCO optimal performance is predominately determined by the local CO_2 concentration and only affected moderately by the presence of oxygen (Savir *et al.*, 2010).

From an evolutionary point of view, other mechanisms have evolved to tackle the above issues. Several organisms show active CO_2-concentrating mechanisms (CCMs) which increase the substrate at the active site of RubisCO, lessening its imperfections (Badger and Price, 1992). Further details regarding CCMs can be found in the carboxysome section 1.8 below.

1.6.1 Multiple forms of RubisCO

Extensive research in the field over the last 50 years have revealed that there are four forms of RubisCO. Form I, II and III were discovered to be linked to RuBP-dependent CO_2 fixation activity (for review see (Ashida *et al.*, 2005)). Form IV enzymes, found in the methionine salvage pathway (Ashida *et al.*, 2003), were named

'Rubisco-like' proteins. The CO_2–fixing role of form III is still in doubt due to the lack of an identifiable PGK in archaea (Finn and Tabita, 2004).

RubisCO I and II forms were recognized as directly involved in the CBB cycle and correlated with autotrophic growth. These enzymes can be found in a vast variety of organisms such as: chemo-, organo-, and phototrophs; bacteria algae and plants. As a consequence of their heterogeneity, Form I enzymes have been divided in four clades. Forms IA and B are members of the 'Green'-RubisCO group, and forms IC and D belong to the 'Red'-RubisCO group (form IC in *C. necator*) (Badger and Bek, 2008a).

There has been previous speculation on the reason for the considerable RubisCO diversification in proteobacteria, where RubisCO forms IAc, IAq, IC and II are all present. It is likely connected with the metabolic flexibility required by autotrophic growth, where CO_2 levels are not constant and can abruptly fluctuate. It is intriguing that they have adopted a survival strategy where different forms of RubisCO can be used in different environmental niches. The absence of RubisCO form IB has raised equal interest and is linked to questions about the evolution of two types of carboxysomes in proteobacteria and cyanobacteria (Badger *et al.*, 2002) (for review see (Rae *et al.*, 2012)).

1.6.2 RubisCO active site and catalysis

The most common form of RubisCO, form I, is a 550 kDa complex formed by eight large (CbbL in *C. necator*; 55 kDa) and eight small subunits (CbbS in *C. necator;* 15 kDa). The large subunits assemble in a toroid of antiparallel dimers, where four small subunits at each end cap the structure forming a hexadecameric structure (Figure 1.8) (Bhat *et al.*, 2017).

Despite the differences in amino acid sequence between different RubisCO forms, the secondary structure of the large catalytic subunit

is extremely conserved throughout the clades (Andersson, 2008). The RubisCO active site is located at the interface of two dimers, between the C-terminal of one large subunit and the N-terminal of the second large subunit (Figure 1.8; in orange). The arrangement of the dimers in a hexadecameric structure allows all eight active sites to face the outside solvent. The main reaction catalysed by RubisCO entails multiple steps and association with intermediates of variable stability (Bhat *et al.*, 2017). To be functional, RubisCO needs prior interaction with Mg^{2+} cofactor, which ligates to two carboxylated ligands provided by Asp203 and Glu204 (using *Nicotiana tabacum* nomenclature throughout the section; Figure 1.8 C) and three water molecules. Subsequently, the enzyme is carbamylated on the Lys201 residue by a CO_2 molecule, that is not involved in the catalysis. The carbamylation is necessary for correct RuBP binding in the active site. The carbamylated Lys201 is stabilised by the interaction with the Mg^{2+} ion (Andersson, 2008).

After RubisCO activation, the RuBP binding replaces two of the water molecules coordinating with the Mg^{2+} ion, with the remaining water molecule replaced by CO_2 in the carboxylation step (Andersson, 2008). The gaseous substrate CO_2 interacts with conserved residue Lys334 on loop 6 of the large subunit and the Glu60 and Thr65 in the N-terminal region of the adjacent large subunit during catalysis (Figure 1.8 C). Upon substrate binding, the conformation of the RubisCO large subunit transitions from the open state to the closed state *via* two conformational changes (Bhat *et al.*, 2017). The closing mechanism requires movements of the loop 6 (residues 331-338), the C-terminal strand (residues 463 to the C-terminal) of one large subunit and the N-terminal loop (residue 63-68) of the adjacent large subunit (Figure 1.8). The loop 6 at the C-terminal of the structure transitions over the bound RuBP, trapping it in the catalytic pouch. Then, the C-terminal tail latches across the loop 6 pinning it down, closing the active site (Bracher *et al.*, 2017). This conformational

change generates the physical environment required for the electrophilic attack of RuBP by CO_2 or O_2.

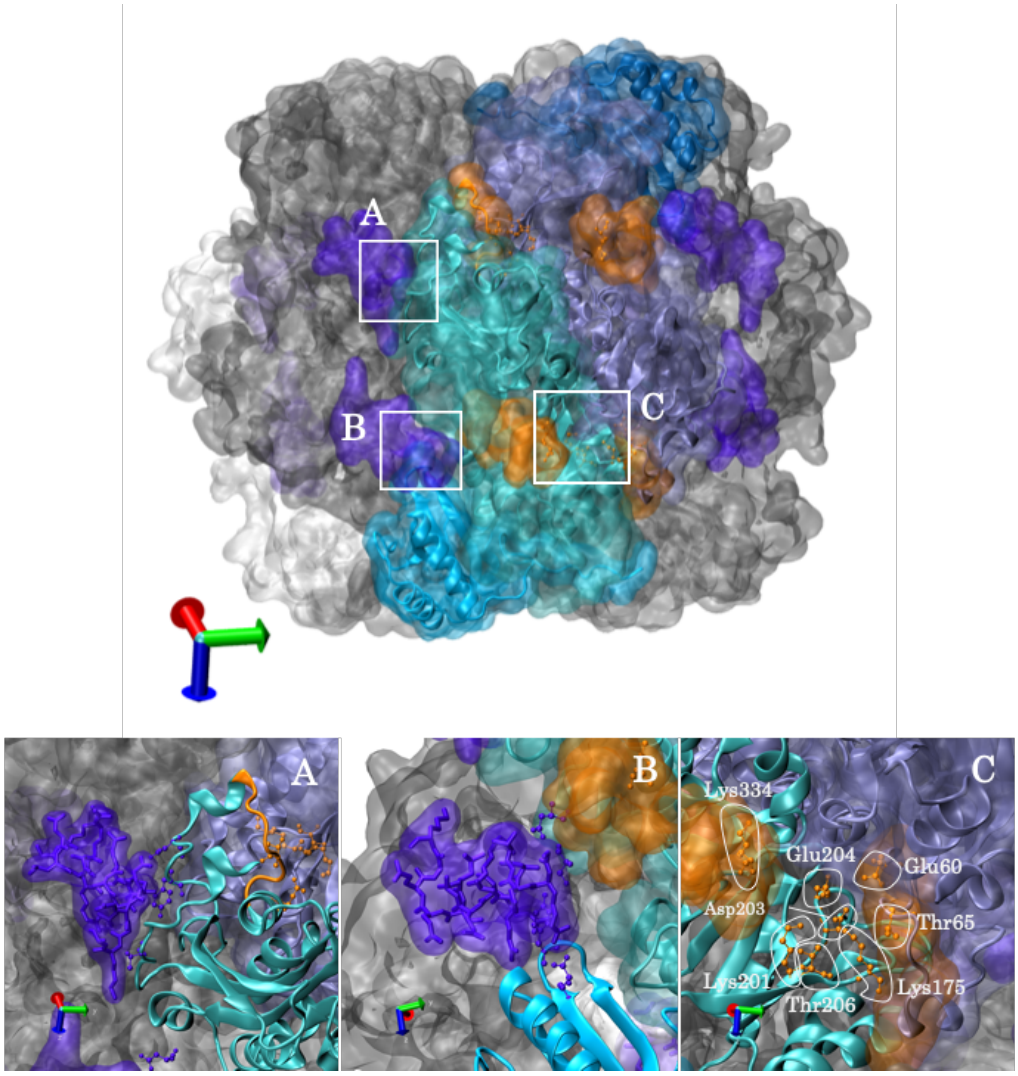


Figure 1.8: Hexadecameric structure of RubisCO with details about active site and CsoS2 peptide interaction site. The hexadecameric structure of the RubisCO (from *H. naepolitanus* RubisCO; PDB: 6uew) is displayed in grey with highlighted two subunits involved in the catalysis. L1S1 dimer depicted in turquoise for the large subunit and light blue for the small subunit. L2S2 dimer depicted in lilac for the large subunit and blue for the small subunit. Regions of L1 and L2 subunits involved in catalytic conformational rearrangement in orange. Amino acid residues involved in the catalytic activity of the enzyme in orange CKP with white tags. CsoS2 peptides shown in purple. (A) and (B) Interaction site between *H. naepolitanus* RubisCO (PDB: 6uew) and CsoS2 recruiting peptides. (C) Active site of *Hn*RubisCO with highlighted amino acid residues involved in the catalytic reaction. Image created in VMD.

The carboxylation of the substrate RuBP requires five steps and three transition states: enolisation of the RuBP at C3; carboxylation of the 2,3-enediolate at C2, hydration of the ketone at C3, cleavage of the C2-C3 bond and the stereospecific protonation of one 3-PGA product. The first step involves the conversion of the ketone form of the substrate enediol in order to direct the attack of CO_2 to the C2 of

RuBP (Bracher *et al.*, 2017). This task is accomplished by the deprotonation of the C3 by Lys201 and the protonation of the oxygen in C2 by Lys175 (Figure 1.8 C). The electrophilic attack of CO_2 on the C2, leads to the formation of a branched six-carbon intermediate, which is hydrolytically halved and stereo-specifically protonated into two molecules of 3-PGA. The overall reaction results in a net intake of inorganic carbon. Unfortunately, the incorporation of an oxygen molecule on the enediolate leads to the formation of 2-PGA and 3-PGA with no net gain of carbon (Andersson, 2008; Bracher *et al.*, 2017). Additionally, the erroneous re-protonation of the enediolate intermediate cause the creation of the 'misfire' sugar-phosphate D-xylulose-1,5-bisphosphate (XuBP), which inactivates the active site (Bhat *et al.*, 2017).

1.6.3 RubisCO assembly and metabolic repair

The synthesis and assembly of the RubisCO require ancillary proteins, named molecular chaperonins. Chaperonins are ATP driven double-ring complexes able to provide a compartment in which molecules can fold in isolation without aggregation. The chaperonins involved in the folding of RubisCO are either the chloroplast chaperonin Cnp60 or the bacterial homolog GroEL/ES (Hayer-Hartl, 2017). Recent developments in cryo-electron crystallography allowed the study of the transient stage of interaction between GroEL/ES and RubisCO, giving in-depth details on the interaction between the proteins and the encapsulation process (Chen *et al.*, 2013). For further details on GroEL/ES subunit composition and activity refer to review (Bracher *et al.*, 2017).

Once assembled, the multistep catalytic activity of the error-prone RubisCO generates 'misfire' sugar phosphates, which inhibit the activity of the enzyme. As mentioned before, an inactive enzyme can also be the result of RuBP substrate binding prior to RubisCO activation *via* carbamylation. In both cases, the release of the

inhibitor is made possible by the interaction and remodelling of the RubisCO by RubisCO activases (Mueller-Cajar, 2017). These enzymes belong to the conserved AAA+ (ATPases Associated with diverse cellular Activities) ATPase family, which in its hexameric conformation release RubisCO from its inhibitory molecules (Jordan and Chollet, 1983; Portis, 2003). Since their discovery in the 1980s' (Portis and Salvucci, 2002), three distinct classes of Rubisco activases have now been identified: green-type associated with form IA and II RubisCO; green-type associated with form IB RubisCO, and; red-type (Mueller-Cajar *et al.*, 2011; Tsai *et al.*, 2015; Bhat *et al.*, 2017).

The RubisCO activases of green-type associated with RubisCO form IB were the first to be discovered (Portis and Salvucci, 2002) possibly due to their presence in all higher plants and therefore greatly studied (Portis, 2003). They are thought to be evolutionarily derived from the cyanobacteria, where their homologs are still linked with the carboxysomal RubisCO form IB (Mueller-Cajar, 2017).

The RubisCO activases of red-type are associated with the RubisCO form IC and ID. Until recently, their presence was assumed to be restricted to plants. Nevertheless, they were first discovered in the proteobacterium *Rhodobacter sphaeroides* (Mueller-Cajar *et al.*, 2011) and later on its homolog was identified in *C. necator* H16 (Gruber *et al.*, 2017). The enzyme is encoded by the *cbbX* gene, in the proximity of the *cbbL* and *cbbS* genes in the *cbb* operon (Figure 1.6) (Gibson and Tabita, 1997). Its crucial role was demonstrated by the deletion of the encoding gene ablating the autotrophic growth of *C. necator* H16 and *R. sphaeroides* (Gibson and Tabita, 1997; Bowien and Kusian, 2002).

The most recent addition to the RubisCO activases family is the green-type associated with RubisCO form IA and II. They are encoded by the *cbbQ* and *cbbO* genes of chemoautotrophic bacteria, such as *Halothiobacillus naepolitanus* and *Acidithiobacillus ferrooxidans* (Sutter *et al.*, 2015; Tsai *et al.*, 2015). The genes are usually in the

genomic surroundings of the RubisCO genes. The *cbbQ* gene encodes for hexameric AAA+ subunits (about 30 kDa each) and the *cbbO* for a RubisCO adaptor protein (about 80 kDa). Further biochemical characterisation has shown that these proteins work as a hetero-oligomeric complex in a Q_6O_1 stoichiometry. The CbbQ functions as an activase, while interacting with the von Willebrand factor A (VWA) domain of the CbbO (Sutter *et al.*, 2015; Tsai *et al.*, 2015). There is a gene pair *cbbQ/cbbO* that is associated with a carboxysomal gene cluster encoding a RubisCO form IAc in *H. naepolitanus* (Sutter *et al.*, 2015). The CbbQ complex was shown to interact with CbbO and to be part of the carboxysome shell (Sutter *et al.*, 2015).

Another enzyme plays an important role in preserving RubisCO activity: CbbY, a XuBP phosphatase (Bracher *et al.*, 2015). RubisCO side activity creates small amounts of XuBP, which is a strong inhibitor for the enzyme itself. CbbY converts XuBP in RuBP, allowing substrate recycling. However, XuBP inactivation of RubisCO appears to be less detrimental than the loss of CbbX activity (Pearce, 2006; Parry *et al.*, 2008; Bracher *et al.*, 2015). In fact, *cbbY* knock-out strains of *C. necator* H16 *and R. sphaeroides* displayed normal autotrophic growth, with only reduced RubisCO activity (Gibson and Tabita, 1997; Bowien and Kusian, 2002).

1.7 Carbonic anhydrases and hydrogenases in *C. necator* H16

Autotrophic CO_2 assimilation is a very energy-intense process and to control it sensibly the facultative autotrophs required an accurate system to couple fixation rate to their needs (Codd and Kuenen, 1987). Therefore, organisms rely on an efficient system of controlling pH and CO_2 concentration to perform carbon assimilation (Gai *et al.*, 2014). During autotrophic growth in *C. necator*, carbon fixation proceeds through the CBB pathway, which requires a consistent supply of CO_2, as the main building block, and energy to carry out the required reactions.

1.7.1 Carbonic anhydrases

Carbonic anhydrases (CA) are very important to feed a high concentration of CO_2 to the RubisCO, during autotrophic metabolism. In *C. necator* H16 four genes were identified as coding for CAs in the genome. Two CA genes are localised on chromosome I and two on chromosome II. The *can* and *can2* genes code for β-CA (locus tag H16_A0169 and H16_B2270, respectively). While, the *caa* gene encodes a putative periplasmatic α-CA (locus tag H16_B2403) and the *cag* gene a γ-like-CA/acetyltransferase (locus tag H16_A1192) (Pohlmann *et al.*, 2006). The presence of multiple classes of enzymes in *C. necator* H16 suggests the unique roles they play in fine-tuning CO_2 transport and metabolism. Further study tested each putative role and intracellular localisation of the CA enzymes in *C. necator* H16 (Gai *et al.*, 2014). The α-CA enzyme was proposed to convert CO_2 to bicarbonate in the periplasmatic space. This hypothesis is supported by the natural balance of the Caa-catalysed reaction, which lies towards the hydration of CO_2. Furthermore, deletion, imbalanced expression and mislocalisation of Caa all negatively impacted the fitness of *C. necator* mutant strains, underlying its involvement in CO_2 transport and bicarbonate supply (Gai *et al.*, 2014). Additionally, Can was suggested to be responsible for the supplementation of CO_2

to the central metabolism in the cytosol, given the substrate specificity of the enzyme and the dependency of the *can* deletion strains from external CO_2 feed for survival (Kusian *et al.*, 2002; Gai *et al.*, 2014). Cag could potentially perform the same reaction, but its role wasn't wholly elucidated in the study (Gai *et al.*, 2014) and thus requires further investigation. Lastly, the activity of Can2 was highlighted as potentially involved in the pH homeostasis of the cell.

1.7.2 Hydrogenases

In nature, microorganisms harvest the reducing power of H_2 as an energy source. Knallas bacteria, such as *C. necator* H16, are well known for their ability to grow in aerobic autotrophic conditions and to use H_2 for reducing power, while using O_2 as terminal electron acceptors (Lauterbach and Lenz, 2019). In *C. necator* H16 the energy conservation is mediated by two [Ni-Fe]-hydrogenases: a membrane-bound hydrogenase (MBH) and a soluble hydrogenase (SH). The hydrogenases are encoded by two gene clusters with an overall length of 80 kb localised on the megaplsmid pHG1 (Lenz *et al.*, 2002). The two transcriptional units, the MBH operon and the SH operon, are coordinatively regulated by *hoxA, hoxB, hoxC* and *hoxJ* gene products, forming a regulon (Lenz *et al.*, 1994, 2002). The two upstream strong promoters are recognised by an alternative sigma factor σ^{54} (RpoN) of the RNA polymerase. However, additional transcriptional starting sites with conventional σ^{70} were also discovered, suggesting an even more complex level of regulation (Lenz *et al.*, 2002). In general, two physiological conditions are met to allow gene expression: the presence of H_2 in the media and the absence of organic acids, where the degree of repression correlates directly with the quality of the substrate (Lenz *et al.*, 2002). The MBH is anchored *via* the C-terminal of the small subunit to the cytoplasmatic membrane facing the periplasm and performs H_2 oxidation splitting the H_2 into two protons and two electrons. Where, the latter are directly moved to the cytochrome-b as the primary electron acceptor of the respiratory

chain, resulting in proton motor force used for ATP synthesis (Lenz *et al.*, 2002; Lauterbach and Lenz, 2019). The cytoplasmatic SH provides reducing equivalents to the cell, coupling the H_2 oxidation with the reduction of the nicotinamide cofactor NAD^+ to NADH (Lauterbach and Lenz, 2019).

1.8 Carboxysomes: metabolic modules for carbon fixation

Membrane-bound organelles for compartmentalization of specific functions are the key feature of all eukaryotic cells. Bacteria also possess organelles, however, they are not present throughout the domain. They are, instead, irregularly distributed and frequently provide key functions for the microorganisms' success in their specialised niche (Gonzalez-Esquer *et al.*, 2016). More than 60 years ago, polyhedral bodies composed only of proteins were discovered in electron micrographs within members of the cyanobacteria (Drew and Niklowitz, 1956) and later on in chemoautotrophs (Shively *et al.*, 1970). They were subsequently identified as proteinaceous shells encapsulating enzymes required for carbon fixation, and they were named accordingly, carboxysomes (CBs) (Shively *et al.*, 1973). Through sequence comparison of encapsulating shell proteins in microbial genomes, it was shown that carboxysomes are part of a family of related organelles, known as bacterial microcompartments (BMCs). X-ray studies on the crystallographic structure of carboxysome subunits (Kerfeld *et al.*, 2005; Tanaka *et al.*, 2008; Klein *et al.*, 2009) led to the discovery of three main types of shell proteins and to an icosahedral shell model, where the facets are composed by (pseudo)hexameric building blocks capped by pentameric vertices. BMC-H, a hexamer of proteins containing a single PF00936 domain (Kerfeld *et al.*, 2005), BMC-T, a trimer of proteins containing two PF00936 domains (Klein *et al.*, 2009) and BMC-P, a pentamer formed by proteins containing one PF03319 domain (Tanaka *et al.*, 2008) are the building blocks of the BMC architecture.

The presence of BMCs was predicted in 23 bacterial phyla (Axen *et al.*, 2014), where they assume a key role in the success of the organism, mainly dictated by the function of the encapsulated enzymes. Despite their functional diversity, only two BMCs namely:

α and β carboxysomes are known to take part into the autotrophic metabolism of the CBB cycle (Kerfeld and Melnicki, 2016).

The carboxysome is the core component of what is known as the CO_2–concentrating mechanism (CCM) (Badger and Price, 1992; Kaplan and Reinhold, 1999; Price *et al.*, 2008), which works together with CO_2 diffusion and HCO_3^- uptake transporters to accumulate inorganic carbon, primarily as HCO_3^-, into the cytoplasm of the cell (Figure 1.9 B) (Price *et al.*, 1998; Dobrinski *et al.*, 2005). HCO_3^- then diffuses through the proteinaceous shell layer and accumulates in the lumen where CA enzymes converts it into CO_2 (Figure 1.9 B). Subsequently, it is fixed by RubisCO into a three-carbon sugar (3-phosphoglycerate; 3-PGA) that can freely diffuse through the proteinaceous barrier and take part in the CBB cycle.

The putative mechanism for efficient carboxysomal C_i fixation involves RubisCO and carbonic anhydrase enzymes sequestration within a selectively permeable protein shell layer (Figure 1.9 B, in blue). The molecular mechanism by which CBs enhance the catalytic performance of RubisCO is partially due to the co-localisation with CA activity. The CA localisation model has shown that CA *nuclei* are present throughout the shell structure (Cot *et al.*, 2008) and are surrounded by RubisCO zones (Rae *et al.*, 2013). Moreover, it has been experimentally proven that α and β-carboxysome shells provide CO_2 diffusion resistance (Reinhold *et al.*, 1991; Cai *et al.*, 2009) preventing leakage of this C_i species and oxygen influx, yet permitting the transit of HCO_3^-, Mg^{2+}, RuBP and 3-PGA. Sequestration of this rather inefficient enzyme into a microcompartment allows carboxysomes to function as a CO_2 trap that provides RubisCO with a substrate enriched environment, promoting enhanced fixation rates.

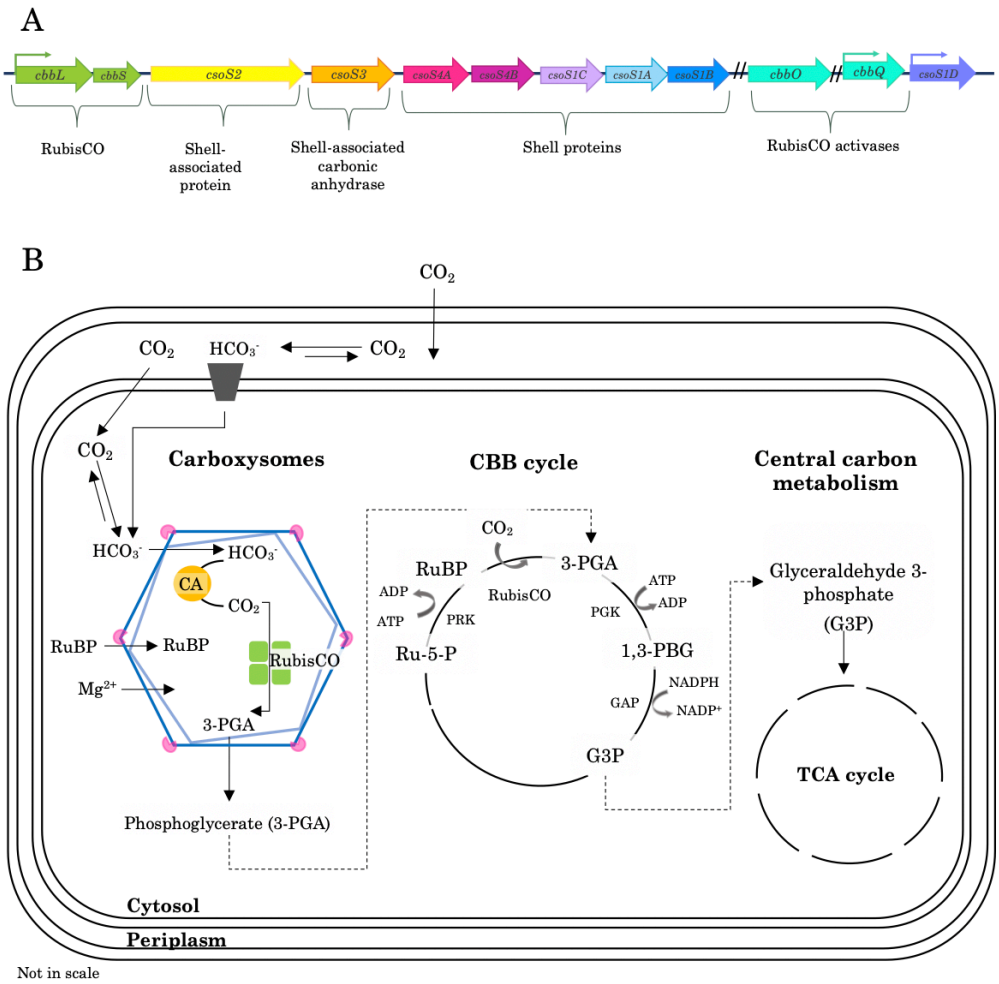


Figure 1.9: Schematic view of the *cso* operon from *H. naepolitanus* and of the link between carboxysome CO_2 fixation and the central carbon metabolism. (A) Diagram of the *cso* operon from *H. naepolitanus* with adjacent genes encoding for RubisCO activases: CbbQ and CbbO. (B) Schematic view of the link between CO_2 fixation in the carboxysomes and the metabolic pathways in lithoautotrophic bacteria.

1.8.1 Carboxysomes types with different components and evolution

Alpha and beta-carboxysomes share similar general architecture and functional principles, but they differ in several details as phylogenetic, sequestered RubisCO type, gene organisation, including protein lumen arrangement and therefore ultimately in the assembly method.

Previous research (Tabita, 1999), clustered RubisCO form I into two distinct evolutionary groups. Phylogenetic affinity was found, between the RubisCO amino acid sequence of oceanic cyanobacteria and proteobacteria species with a Rubisco form called IA. Other cyanobacteria phyla instead showed sequence similarity with those of higher plants containing RubisCO form IB. Cyanobacteria expressing RubisCO IA came to be known as α-cyanobacteria and those using Rubisco IB form as β-cyanobacteria. Interestingly, among autotrophic bacteria possessing RubisCO genes, it is possible to discern between those containing carboxysomes based on protein sequence (Badger and Bek, 2008a). For this reason, Rubisco form IA can be further divided in two subgroups: IAc associated with carboxysome structures and IAq without any obvious correlation with CBs operon. The classification was later expanded on by Badger *et al*. (Badger *et al.*, 2002), who discovered that carboxysome genes also typify specific phylogenetic clades. Hence, the carboxysomes of α-cyanobacteria carrying form IAc RubisCO (encoded by *cso* operon) came to be known as α-carboxysomes and the carboxysomes of β-cyanobacteria containing RubisCO form IB (encoded by *ccm* operon), as β-carboxysomes. Despite their fairly similar external appearance and the conserved function, α and β-carboxysomes revealed to have a manifest difference in terms of protein components and internal shell architecture.

1.8.1.2 Structure and assembly of β-carboxysome

Previous studies revealed that the differences between α-carboxysomes and β-carboxysomes is already apparent from the organisation of their respective encoding genes. β-carboxysome accessory protein and genes are dispersed throughout the cyanobacterial genome. The core elements are mostly located in the canonical *ccm* operon, or occasionally near RubisCO *rbcLS* genes, which are adjacent to the core operon (Rae *et al.*, 2012; Axen *et al.*, 2014).

Recent models of β-carboxysomes showed that at least 11 proteins are needed for the correct assembly in the model organism *Synechococcus elongatus* PCC7942 (Rae *et al.*, 2013). These peptides fall into two main classes: shell layer proteins and lumen proteins (CcmM-35, RbcL, and RbcS). The former class divides into inner shell bicarbonate dehydration/RubisCO-organizing layer (formed by CcaA, CcmM-58, CcmN) and outer shell BMC-like layer (formed by CcmK2-4, CcmL, and CcmO) (Rae *et al.*, 2013; Kerfeld and Melnicki, 2016). Genetic deletion analysis allowed to differentiate between essential and accessory genes for the formation of a functional carboxysome, identifying CcmK, CcmM, and CcmN as of central importance for the assembly (Rae *et al.*, 2013; Kerfeld *et al.*, 2018).

Model species for β-carboxysome exhibited two isoforms of the CcmM protein (CcmM-58 and -35) that are encoded by the *ccmM* gene and are the most important β-carboxysome scaffolding proteins (Long *et al.*, 2010; Rae *et al.*, 2013). CcmM-58 has the toilsome role of recruiting several proteins such as CA, and outer shell proteins (Long *et al.*, 2007, 2010; Cot *et al.*, 2008) while interlinking adjacent subunits of RubisCO. Whilst, CcmM-35 isoform is expected to be confined to the carboxysome lumen. Electron microscopic studies of β-carboxysomes (Kaneko *et al.*, 2006) displayed para-crystalline arrays due to the structural role of CcmM-35 protein, which is thought to

organise RubisCO enzymes into layers (Long *et al.*, 2011). As a part of the shell organising proteins, CcmN, contains hexapeptide-repeat domains (Pfam00132) on its N-terminus, which are needed for recruiting CcmM and an encapsulation peptide (EP). The latter on its C-terminus possesses an amphipathic helix, used to recruit the shell proteins (Kinney *et al.*, 2012).

In the proposed model, the outer shell is composed of protein complexes formed by CcmK2-4 (BMC-H, hexameer), CcmO (BMC-T, trimer) and CcmL proteins (BMC-P, pentamer) (Kerfeld *et al.*, 2005, 2010; Tanaka *et al.*, 2008; Kinney *et al.*, 2011). Crystal structures revealed the capacity to form sheets or linear strips of hexamers of CcmK1, CcmK2, and CcmK4 proteins (Kerfeld *et al.*, 2005; Tanaka *et al.*, 2009), where each hexameric unit has a positively charged pore which is thought to mediate charged-solute transit, via an unknown mechanism (Kerfeld *et al.*, 2005). A homologous relationship was established between CcmO and CcmK genes, with CcmO being composed of tandem repeats of a CcmK-like unit (J M Shively *et al.*, 1998; Rae *et al.*, 2013). Nevertheless, the expected three-dimensional conformation of the protein is a trimer, with a structure similar to those of CcmK hexamers. Given the typical icosahedral structure of the carboxysome model, the CcmL protein containing the PF03319 domain, forming pentagonal pentamers, fits the role of structural vertices (Tanaka *et al.*, 2008, 2009). Recently, a gene, *ccmP* was identified as an orthologue of the α-carboxysome csoS1D gene (Cai *et al.*, 2012). CcmP protein is expected to reassemble the crystal structure of CsoS1D, providing a potential pore for large metabolite transit.

The assembly of β-carboxysomes starts with nucleation of the enzymatic core by CcmM, followed by its encapsulation in the shell through interaction with the core protein CcmN (Cameron *et al.*, 2013). Initially, soluble RuBisCO is nucleated by domains of the

CcmM protein, where the N-terminal γ-carbonic anhydrase domain of CcmM interacts with both the shell (Cai *et al.*, 2016) and with the N-terminal domain of CcmN. Subsequently, the CcmN amphipathic alpha helix, which was shown to play a key role during carboxysome assembly, makes contact with the outer shell components (Kerfeld and Melnicki, 2016). The pentameric vertices close the shell and the newly formed CB redistributes along the cytoskeletal protein ParA (Kerfeld *et al.*, 2018).

1.8.1.1 Structure and assembly of α-carboxysome

Contrary to the dispersed nature of β-carboxysome gene organisation (Kinney *et al.*, 2011), α-carboxysomes are organised into a collected operon (*cso)* encoding between 8 to 11 proteins, depending on species (Cannon *et al.*, 2003; Heinhorst *et al.*, 2006; Roberts *et al.*, 2013). The canonical *cso* operon structure carries genes for RubisCO form IA large (*cbbL)* and small (*cbbS*) subunits, a protein with putative lumen organisation role (*cso*S2), β-carbonic anhydrase (*cso*SCA also called *cso*S3), followed by five to seven, small genes forming the shell proteins: *cso*S1A to D, *cso*S4A and B in *H. naepolitanus* (Figure 1.9 A) (Cannon *et al.*, 2003; Axen *et al.*, 2014). In the genome, the *cso* operon is surrounded by conserved gene clusters, called superloci, which typically encode for ancillary proteins and C_i uptake pumps involved in the carboxysome functionality (Turmo *et al.*, 2017).

The carboxysome shell proteins (ranging between 10 to 11 kDa) form the typical architecture of BMC domains with CsoS1A to C forming flattened, hexagonal hexamers and CsoS1D trimers (Tsai *et al.*, 2007; Klein *et al.*, 2009). It is important to note that the presence of a central pore in each hexamer, with a specific size and charge distribution may be responsible for substrate selective transit (HCO_3^-,RuBP and Mg^{2+}) or product (3-PGA) exit (Kinney *et al.*, 2011). Recent crystallographic studies of CsoS1D protein structure revealed the distinctive, dual-layer hexametric conformation of the protein able, when assembled into trimeric unit, to show an 'open' and 'closed' state. This suggests

that an interchange between the conformations could gate the transit of large RubisCO substrate RuBP (Klein *et al.*, 2009). Thus, it would explain the relevance of CsoS1D protein in the correct and functional assembly of CBs during *H. naepolitanus* operon heterologous expression (Menon *et al.*, 2008; Bonacci *et al.*, 2012).

The typical icosahedral geometry of BMC shells requires the presence of a pentagonal shaped protein to fill the vertices. This function is fulfilled by pyramidal pentamer proteins known as CsoS4A and B (Tanaka *et al.*, 2008), that conform to the idealized structure of α-carboxysomes.

Over the years several hypotheses on the function of the intrinsically disordered protein CsoS2 have emerged. Recent studies assigned it a crucial role during the assembly of subshell structure of α-carboxysomes (Rae *et al.*, 2013; Cai, Dou, *et al.*, 2015). CsoS2 in *H. neapolitanus* can be found in two isoforms: full length and in a truncated form, of 130 kDa and 85 kDa, respectively (Baker *et al.*, 1999). The long form CsoS2B is formed of three regions (N-, M- and C-) while the short form CsoS2A possesses only two regions (N- and M-). Each of those regions have a particular role during the structural assembly of the α-CBs. Work from the Cai group hypothesised that the CsoS2B N-region was involved in the recruitment of CsoS1 shell proteins, the M-region interacted and aggregated with RubisCO while the C-region was responsible for anchoring CsoS2 to the forming shell (Cai, Sutter, *et al.*, 2015).

For further information, the main carboxysome lumen proteins: RubisCO and β-carbonic anhydrase please refer to section 1.6 and 1.8.2, respectively.

As already mentioned (section 1.6.3), numerous ancillary proteins are involved in the folding and activation process of RubisCO. Bioinformatic studies, have shown, in the recent past, genomic clusters with a strong association with the α-carboxysome genes

(Jorda *et al.*, 2013; Axen *et al.*, 2014; Desmarais *et al.*, 2019). Further research work highlighted a PCD-like family of proteins associated with the α-carboxysome superloci, involved in the folding and assembly of the hexadecameric structure of the carboxysomal RubisCO. The protein family was renamed after the study α-carboxysome RuBisCO assembly factor (or acRAF) (Wheatley *et al.*, 2014). In many α-carboxysome superloci, RubisCO activase genes, such as *cbbX* and *cbb*OQ, were mined during bioinformatic studies (Axen *et al.*, 2014). Further experimental studies showed that the CbbQ was interacting directly with the carboxysomal shell proteins in *H. neapolitanus* (Sutter *et al.*, 2015) and that the CbbQ forms a complex with CbbO, where the VDA domain of the later interacts with the inactive RubisCO to release the inhibitory RuBP (Tsai *et al.*, 2015).

The assembly process of the α-carboxysome has always been more challenging to unravel than the β-carboxysome counterparts, given the absence of recognisable domains in the unique subshell assembly protein, CsoS2 (Cannon *et al.*, 2003; Cai, Dou, *et al.*, 2015). The importance of the CsoS2 protein in the process was highlighted by its interaction with both RubisCO and CBs shell proteins (Liu *et al.*, 2018) and the absence of carboxysomal structures in CsoS2 knockout strains (Cai, Dou, *et al.*, 2015). Using electron cryo-tomography techniques, carboxysomes were captured during assembly in *H. neapolitanus,* observing simultaneous formation of the shell proteins and packing of RubisCO along the inner surface of the nascent shells (Iancu *et al.*, 2010). Until very recently the details of the interaction between CsoS2 and RubisCO during the assembly were not established. A recent study, carried out by Oltrogge *et al.* 2020, deepened our understanding of the interactions occurring between the *H. naepolitanus* carboxysomal RubisCO and the CsoS2 protein, showing the crystallographic structure of the hexadecameric RubisCO associated with multiple peptides from the N-terminal of the

CsoS2 (Figure 1.8 A and B). The interaction site between the peptides and RubisCO was mediated by highly conserved residues with a predominantly electrostatic interface, with salt bridges and π-cation interactions. They were localised at the conjunction between RubisCO large and small subunits, ensuring the encapsulation of only the functional holoenzyme (Figure 1.8 A and B). Five highly conserved residues on the CsoS2 peptides were observed interacting with the RubisCO dimeric structure: residue Arg3 engaging in a salt bridge with Asp94 of the CbbS; residue Arg9 interacting with a salt bridge and π-cation with Asp360 and Phe346 of the CbbL, respectively; residue Arg10 engaging in two π-cation bonds with Tyr72 and Tyr96 of the CbbL and CbbS, respectively, as well as, in a salt bridge with Asp69 of the CbbL; while, both peptide residues Gly17 and Lys18 interact with Asp26 of the CbbL *via* two hydrogen bonds and a salt bridge, respectively. Furthermore, the multivalency of the interaction was shown to be a key factor during the recruitment and assembly process, given the weak interaction between an individual peptide and RubisCO residues, where eight peptides would be needed to engage for the recruitment of a functional holoenzyme in its hexadecameric conformation (Oltrogge *et al.*, 2020).

All the previous findings are consistent with a model suggesting the simultaneous assembly of the shell and core of carboxysomes, with the N-terminal of the CsoS2B recruits CsoS1 proteins, increasing the local concentration until the threshold required for self-assembly is reached. At the same time, the M-region recruits the RubisCO and the C-terminal anchors the resulting structures, to the shell. While, CsoS2A promotes the further packing of RubisCO (Iancu *et al.*, 2010; Cai, Sutter, *et al.*, 2015; Kerfeld *et al.*, 2018; Oltrogge *et al.*, 2020).

1.8.2 Caboxysomal carbonic anhydrase comparative analysis

Carbonic Anhydrase (CA) is the second essential enzyme of the carboxysome core. Carboxysomal CO_2 accumulation beyond natural

rates is achieved through the activity of carboxysome-specific CA enzymes. Boosting fixation yields, CA converts HCO_3^- to CO_2 in the proximity of RubisCO, saturating the active site of the enzyme.

Only two classes of CA are found in association with carboxysomes (see review (Cannon *et al.*, 2010)). The CA of α-carboxysomes CsoSCA (formerly known as CsoS3) was identified as β-type encoded by the *csoS3* gene (Figure 1.9 A) (So *et al.*, 2004; Sawaya *et al.*, 2006). β-carboxysomes instead, possess two genes encoding for CA activity: the conserved γ-CA domain in the structural lumen protein CcmM (Cot *et al.*, 2008; Peña *et al.*, 2010; Long *et al.*, 2011) and in some strains β-CA, encoded by the *ccaA* gene, is also encapsulated in the carboxysome (Long *et al.*, 2007; Peña *et al.*, 2010).

The α–carboxysomes CA enzyme is thought to interact directly with the shell, through a domain on its N-terminus (Sawaya *et al.*, 2006). Likewise, in β-carboxysomes, where γ-CA domain of CcmM interacts directly with outer shell proteins (Cai *et al.*, 2016). Whereas, the compartmentalization of CcaA is expected to be dependent on its interaction with CcmM protein (Long *et al.*, 2007).

Interestingly, it has been proven experimentally (Long *et al.*, 2010) that CcmM protein is a core component for carboxysome assembly and correct folding of β-carboxysomee, whilst deletion of CsoSCA in *H. neapolitanus* (Dou *et al.*, 2008) or CcaA in *Synechocystis s*p. PCC 6803 (So *et al.*, 2002) and *S.* PCC7942 (Long *et al.*, 2011) does not affect carboxysome formation.

1.8.3 Carboxysomes bioengineering

Carboxysomes contributed to the genomic versatility and success of several bacterial in specific environmental niches. The spatial segregation from the central metabolism, and the enhanced activity yields offered by bacterial microcompartments raised a strong interest, especially in the last decade, for bioengineering (Kerfeld *et al.*, 2018). For their use in synthetic biology, they must be genetically amenable to manipulation and design. They already display an incredible ability to incorporate chimeric and heterologous RubisCO species while retaining their functionality (Menon *et al.*, 2008; Cai *et al.*, 2016). Furthermore, the shell proteins can be altered to change the permeability of the shell itself (Cai, Sutter, *et al.*, 2015). While, the feasibility of expression of heterologous α- and β-carboxysomes in *E. coli* (Bonacci *et al.*, 2012; Cai *et al.*, 2016) has already been tested on a proof of concept basis. Moreover, a study succeeded in the first expression and analysis of *H. neapolitanus cso* operon into a Gram-positive model species, *Corynebacterium glutamicum* (Baumgart *et al.*, 2017). Furthermore, another recent study achieved a functional reconstruction of the CCM in *E. coli* (Flamholz *et al.*, 2020).

These findings support the overall modularity of α-carboxysome systems and the feasibility of an orthogonal expression, while taking in to account possible adaptation strategies to achieve optimal functionality in different host organisms.

1.9 Aims of the project

It is the aim of the work described in this thesis to improve CO_2 assimilation efficiency in *C. necator* H16 through two different strategies.

The first strategy, involving the heterologous expression of self-assembled proteinaceous organelles (α-carboxysomes) is implemented in order to observe if their carbon concentrating mechanism (CCM) boosts the carbon fixation efficiency of *C. necator* H16. Further efforts were devoted to the investigation of ancillary proteins: acRAF and RubisCO activases, for carboxysome assembly and functionality in this heterologous host.

At the same time, *in silico* analysis of the relevance of RubisCO surface electrostatic potential, as a mechanism driving RubisCO encapsulation in the α-carboxysomes shell, were carried out using molecular dynamics analysis. Additionally, the modelled RubisCOs were expressed *in vivo* to confirm their functionality.

In the second strategy, the natural regulation of the CBB system is bypassed, in order to investigate the benefit of placing the *cbb* operon under LysR independent and LysR dependent constitutive transcription of the *cbb* operons to assess if these modifications prove to be advantageous for enhancing the efficiency of CO_2 fixation of the CBB cycle.

Chapter 2 - Materials and Methods

2.1 Bacterial strains, media, and growth conditions

2.1.1 Bacterial strains

The bacterial strains used in this study are listed in Table 2.1 and include wild-type and genetically modified *E. coli* and *C. necator* H16 strains.

Strain	Genotype/Description	Reference/Source
E. coli		
DH5-α	F- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17(rk-, mk+) *pho*A *sup*E44 *thi*-1 *gyr*A96	CRG strain collection
S17-1 λpir	*recA, thi, pro*, *hsd*R17(rk-, mk+)RP4: 2-Tc:Mu:Km Tn*7* *λpir*, Tp^R Sm^R	CRG strain collection
C. necator		
H16	wild-type	DSM 428
H16 P_{trp} CBB	*cbbR* and P_{cbbL} defective strain and replacement with P_{trp}	This study
H16 P_{BAD} CBB	*cbbR* and P_{cbbL} defective strain and replacement with P_{BAD}	This study
H16 CbbR A167V	*cbbR* and P_{cbbL} defective strain and replacement with mutated *cbbR* A167V	This study
H16 Δ*panC*	*panC* defective strain	SBRC deletion mutant (Garavaglia M and McGregor C, unpublished)
H16 ΔΔ*cbbLS*	*cbbLS* from Chr II and megaplasmid defective strain	SBRC deletion mutant (Arenas L. C, unpublished)
H16 ΔΔ*cbbLS* Δ*panC*	*cbbLS* from Chr II and megaplasmid *panC* defective strain	This study
H16 Δ*phaCAB*	*phaC phaA phaB* defective strain	SBRC deletion mutant (Arenas L. C, unpublished)
H16 Δ*phaCAB* ΔΔ*cbbLS*	*phaC phaA phaB* and *cbbLS* from ChrI I and megaplasmid defective strain	SBRC deletion mutant (Arenas L. C, unpublished)

H16 Δ*phaCAB* ΔΔ*cbbLS* Δ*panC*	*phaC phaA phaB* and *cbbLS* from Chr II and megaplasmid *panC* defective strain	This study

Table 2.1: List of strains used in this study.

2.1.2 General chemicals

Unless stated otherwise, all the chemicals were purchased from Sigma-Aldrich.

2.1.3 Media

2.1.3.1 Lysogeny broth medium

Standard lysogeny broth (LB) was used for general procedures in *E. coli* and *C. necator* H16. It consisted of 10 g/L tryptone (Oxoid), 5 g/L yeast extract (Oxoid), 10 g/L NaCl and NaOH to pH 7.0. LB agar was prepared by addition of 15 g/L Technical agar No. 3 (Oxoid) to the broth. LB broth and agar were sterilised by autoclaving.

2.1.3.2 Low salt lysogeny broth (LSLB)

LSLB was used for culture growth of *C. necator* H16 in conjugative plasmid procedures (section 2.4.6). It consisted of standard LB containing 2.5 g/L NaCl, pH 7.0. It was supplemented with 15 % (w/v) sucrose in conjugative plasmid procedures for gene replacement (section 2.4.6). LSLB-15 % sucrose agar was prepared by addition of 15 g/L Technical agar No. 1 (Oxoid) to the broth. Sterilisation was achieved by autoclaving.

2.1.3.3 Hannahan's Broth (SOB Medium)

SOB medium was used for the preparation of *C. necator* H16 competent cells. It was reconstituted following manufacturer's instructions. The medium consists of 20 g/L tryptone, 5 g/L yeast extract, 2.4 g/L $MgSO_4$, 0.5 g/L NaCl and 0.186 g/L KCl and sterilised by autoclaving.

2.1.3.4 Schlegel minimum medium (MM)

Chemically defined medium was used for growth assays and contained Na_2HPO_4 x 12 H_2O 9 g/L, KH_2PO_4 1.5 g/L, NH_4Cl 1 g/L, $MgSO_4$ x 7 H_2O 0.2 g/L, $CaCl_2$ x 2 H_2O 20 mg/L, Fe(III)NH_4-Citrate 1.2 mg/L, trace element solution (SL7) 1 mL/L and NaOH to pH 6.9

(Biebl and Pfennig, 1981). The trace element solution (SL7) had the following composition: 25% (w/v) HCl 1.3 mL/L, H_3BO_3 62 mg/L, $CoCl_2$ x 6 H_20 190 mg/L, $CuCl_2$ x 2 H_2O 17 mg/L, $MnCl_2$ x 4 H_2O 100 mg/L, Na_2MoO_4 x 2 H_2O 36 mg/L, $NiCl_2$ x 6 H_2O 24 mg/L and $ZnCl_2$ 70 mg/L (Biebl and Pfennig, 1981). It was supplemented with, 20 mM sodium gluconate (SG) (Fisher Scientific) unless otherwise stated in the results chapters. MM agar was prepared by addition of 15 g/L Technical agar No. 1 (Oxoid) to the broth. MM and its derivatives were sterilised by filtration using 500 mL Rapid-Flow Filter Units 0.2 µm a PES membrane (Scientific Laboratory Supplies).

2.1.3.5 *Li et al.* 2012 minimal medium (L-MM)

The original medium (Li *et al.*, 2012) was modified to match the components that we had in our laboratory. The medium used was composed of three separate stock solutions to mix with a carbon source for growth assays in organo-autotrophic conditions in the microbioreactor (BioLector Pro®) or with no carbon source in autotrophic conditions in bioreactors. The stock solution L-MM (5x) contained $(NH_4)_2SO_4$ 5 g/L, KH_2PO_4 3 g/L, Na_2HPO_4 34 g/L to dissolve in 1L of dH_2O and autoclave (shelf life of 2 months at room temperature). The stock ThiaMix (10x) contained $MgSO_4$ x 7 H_2O 2 g/L, $FeSO_4$ x 7 H_2O 200 mg/L, $CaSO_4$ x 2 H_2O 40mg/L, thiamine hydrochloride 1mg/L to dissolve in 1L of dH_2O and sterilised by filtration using 500 mL Rapid-Flow Filter Units 0.2 µm a PES membrane (Scientific Laboratory Supplies). The trace element solution (SL7; 1000x) had the following composition: 25% (w/v) HCl 1.3 mL/L, H_3BO_3 62 mg/L, $CoCl_2$ x 6 H_20 190 mg/L, $CuCl_2$ x 2 H_2O 17 mg/L, $MnCl_2$ x 4 H_2O 100 mg/L, Na_2MoO_4 x 2 H_2O 36 mg/L, $NiCl_2$ x 6 H_2O 24 mg/L and $ZnCl_2$ 70 mg/L (Biebl and Pfennig, 1981). The carbon source (HCOONa 5.44 g/L) was added if necessary and sterilised by filtration using 500 mL Rapid-Flow Filter Units 0.2 µm a PES membrane (Scientific Laboratory Supplies).

2.1.4 Culture growth

Unless stated otherwise, *E. coli* and *C. necator* H16 strains were grown aerobically in a shaking incubator (Thermo Scientific™ MaxQ™ 8000 Incubated Stackable Shaker) at 37 or 30 °C, respectively, with shaking at 200 rpm. Falcon tubes (Greiner Bio-one) and shake flaks (Duran® baffled flask 250 mL) were filled to a 10 % of their total volume to allow for good oxygenation. Culture growth was monitored by measuring optical density at 600 nm (OD_{600}) with a BioMate™ 3S Spectrophotometer (ThermoFisher Scientific).

2.1.5 Bacterial storage

All *E. coli* and *C. necator* H16 strains were stored at -80 °C using Microbank™ vials (Pro-Lab Diagnostics). A liquid culture of the strain of interest grown overnight (O/N) was used to inoculate three Microbank™ vials, colour-coded green, yellow and red. For recovery, a single bead was used to directly streak on to solid medium or dropped into an appropriate liquid medium.

2.2 Antibiotics

The following antibiotics were used in this study:

Antibiotic	Stock solution	Concentration for *E. coli* (µg/mL)	Concentration for *C. necator* (µg/mL)
Tetracycline	12.5 mg/mL in 100% EtOH	12.5	15
Gentamicin	10 mg/mL in H_2O	NA	10
Chloramphenicol	50 mg/mL in H_2O	25 in plates; 12.5 in liquid	N/A
Kanamycin	250 mg/mL in H_2O	50	N/A

Table 2.2: List of antibiotics used in this study

All antibiotics were prepared as stock solutions and stored at -20° *C*. They were added to the growth medium to give the final concentrations as described in the table above.

2.3 DNA manipulation

2.3.1 Plasmids

All plasmids used in this study are listed in Table 2.3.

Plasmid	Genotype / Description	Reference / Source
pMTL70641::*sacB*	Suicide vector used for knocking-out genes *in C. necator* H16 Modular vector system. pBR322 TetR resistence marker, pBR322 OriT replicon, *sacB* selection marker.	CRG strain collection (Ehsaan M, unpublished)
pMTL70641::P_{cbbL} *cbbR*::P_{j5}	Derivative pMTL70641 vector for knocking-out *cbbR* and promoter of *cbbL* (*cbbR*: H16_RS25680 promoter of *cbbL*: H16_RS25675) in *C. necator* H16 and replace it with P_{j5} constitutive promoter.	This study
pMTL70641::P_{cbbL} *cbbR*::P_{trp}	Derivative pMTL70641 vector for knocking-out *cbbR* and promoter of *cbbL* (*cbbR*: H16_RS25680 promoter of *cbbL*: H16_RS25675) in *C. necator* H16 and replace it with P_{trp} constitutive promoter.	This study

pMTL70641::P*cbbL* *cbbR*:: P*BAD*	Derivative pMTL70641 vector for knocking-out *cbbR* and promoter of *cbbL* (*cbbR*: H16_RS25680 promoter of *cbbL*: H16_RS25675) in *C. necator* H16 and replace it with P*BAD* arabinose inducible promoter.	This study
pMTL70641::*cbbR*::*cbbR* A167V	Derivative pMTL70641 vector for knocking-out *cbbR* and replace it with the mutant *cbbR* A167V.	This study (built by Arenas L. C)
pMTL71301	Medium-high-copy-number plasmid used for heterologous expression. Modular vector system. pBBR1 rep/mob+ replicon with silenced NotI site, pBBR1 TetA/R resistance marker.	CRG strain collection (Ehsaan M, unpublished)
pMTL71301::P*phaC*	Medium-high-copy-number plasmid used for constitutive heterologous expression. Modular vector system. pBBR1 rep/mob+ replicon silenced NotI site, pBBR1 TetA/R resistance marker, P*phaC* promoter.	(McGregor C, unpublished)
pMTL71301::P*phaC*-eYFP	Medium-high-copy-number plasmid used for constitutive heterologous expression. Modular vector system. pBBR1 rep/mob+ replicon silenced NotI site, pBBR1 TetA/R resistance marker, P*phaC* promoter, fluorescence labelling with *eyfp*.	CGR strain collection
pMTL71301::P*BAD*	Medium-high-copy-number plasmid used for inducible heterologous expression. Modular vector system. pBBR1 rep/mob+ replicon silenced NotI site, pBBR1TetA/R resistance marker, P*BAD* promoter.	CGR strain collection
pMTL71301::P*BAD* *rfp*	Derivative pMTL71301::P*BAD* vector. Medium-high-copy-number plasmid used for inducible, heterologous expression using the *rfp* reporter gene.	This study
pMTL71301::P*BAD* *tcdR* P*tdcB* *rfp*	Derivative pMTL71301 P*BAD* vector. Medium-high-copy-number plasmid used for inducible heterologous expression, orthogonal expression cassette as promoter for fluorescence labelling with *rfp*.	This study

pMTL71301::P*BAD* // P*tdcB rfp*	Derivative pMTL71301 P*BAD* vector. Medium-high-copy-number plasmid used for inducible heterologous expression, orthogonal expression cassette lacking *tcdR* promoter (control plasmid), fluorescence labelling with *rfp*.	This study
pMTL71301::P*BAD tcdR // rfp*	Derivative pMTL71301 P*BAD* vector. Medium-high-copy-number plasmid used for inducible heterologous expression, orthogonal expression cassette lacking P*tcdB* promoter (control plasmid), fluorescence labelling with *rfp*.	This study
pMTL71301::P*BAD tcdR* P*tdcB //*	Derivative pMTL71301 vector. Medium-high-copy-number plasmid used for inducible heterologous expression, orthogonal expression cassette lacking *rfp* promoter (control plasmid).	This study
pHnCBS1D	Medium to high-copy-number plasmid used for inducible heterologous expression. pBR322 OriT replicon, CmR resistance marker, P*lac* promoter, *H. naepolitanus cso* operon.	Addgene
pMTL71301::*panC*	Derivative pMTL71301, Medium-high-copy-number plasmid used for heterologous expression. Modular vector system. pBBR1 rep/mob+ replicon with silenced NotI site, pBBR1 TetA/R resistance marker, *panC* selection marker.	(McGregor C, unpublished)
pMTL71301::P*BAD Hn*CB	Derivative pMTL71301::P*BAD* vector. Medium to high-copy-number plasmid used for inducible heterologous expression, *H. naepolitanus cso* operon in *C. necator* H16.	This study
pMTL71301::*panC*::P*BAD*	Derivative pMTL71301::P*BAD* vector. Medium to high-copy-number plasmid used for inducible heterologous expression, *panC* selection marker.	This study
pMTL71301::*panC*::P*BAD Hn*CB	Derivative pMTL71301::*panC*::P*BAD* vector. Medium to high-copy-number plasmid used for inducible heterologous expression, *H. naepolitanus cso* operon in *C.*	This study

	necator H16, *panC* selection marker.	
pMTL71301::P_{phaC} *Hn*CB	Derivative pMTL71301::P_{phaC} vector. Medium to high-copy-number plasmid used for constitutive heterologous expression of the *H. naepolitanus cso* operon in *C. necator* H16.	This study
pMTL71301::*panC*::P_{phaC}	Derivative pMTL71301::P_{phaC} vector. Medium to high-copy-number plasmid used for constitutive heterologous expression, *panC* selection marker.	This study
pMTL71301::*panC*::P_{phaC} *Hn*CB	Derivative pMTL71301::*panC*::P_{phaC} vector. Medium to high-copy-number plasmid used for constitutive heterologous expression of the *H. naepolitanus cso* operon in *C. necator* H16.	This study
puC57::CbbQO	Medium to high-copy-number plasmid used for synthetic gene construction. pBR322 OriT replicon, KanR resistance marker, *cbbO* and *cbbQ* genes	GeneScript
pMTL71301::*panC*::P_{phaC} *Hn*CB Act	Derivative pMTL71301::*panC*::P_{phaC} *Hn*CB vector. Medium to high-copy-number plasmid used for constitutive heterologous expression of the *H. naepolitanus cso* operon with activases (*cbbO* and *cbbQ*) in *C. necator* H16.	This study
pMTL71301::P_{phaC} *Hn*CbbLS	Derivative pMTL71301::P_{phaC} vector. Medium to high-copy-number plasmid used for constitutive heterologous expression of *cbbLS* from *H. naepolitanus*	This study (built by Arenas L. C)
pMTL71301::P_{phaC} *Cn*CbbLS	Derivative pMTL71301::P_{phaC} vector. Medium to high-copy-number plasmid used for constitutive heterologous expression of *cbbLS* from *C. necator* H16.	This study (built by Arenas L. C)
pUC57::*cbbL* mut	Medium to high-copy-number plasmid used for synthetic gene construction. pBR322 OriT replicon, KanR resistance marker, *cbbL* mut gene	GeneScript

pMTL71301::P$_{phaC}$ *Mut*CbbLS	Derivative pMTL71301::P$_{phaC}$ vector. Medium to high-copy-number plasmid used for constitutive heterologous expression of mutated *cbbLS* from *C. necator* H16	This study (built by Arenas L. C)
pMTL71301::*panC*::P$_{phaC}$ *Cn*CbbLS *HnCB*	Derivative pMTL71301::*panC*::P$_{phaC}$ vector. Medium to high-copy-number plasmid used for constitutive heterologous expression of the chimeric *H. naepolitanus cso* operon with *cbbLS* from *C. necator* H16.	This study
pMTL71301::*panC*::P$_{phaC}$ *Mut*CbbLS *HnCB*	Derivative pMTL71301::*panC*::P$_{phaC}$ vector. Medium to high-copy-number plasmid used for constitutive heterologous expression of the chimeric *H. naepolitanus cso* operon with mutated *cbbLS* from *C. necator* H16.	This study

Table 2.3: List of plasmids used in this study.

2.3.2 Oligonucleotide primers

All primers used in this study are listed in Table 2.4.

Primer	Sequence (5' to 3')	Function
Hifi Assembly and traditional cloning primers		
pMTL70641::P_{cbbL} *cbbR*::P_{j5}		
GT11_CBB1_LHA_F	ATGACCATGATTACGAATTC GAGCTCCATGGTGCCGGCG GTCAC	Amplify left homology arm from *C. necator* H16, P_{j5} sequence overhang
GT55_CBB1A_P_{j5}_LHA_R	ATTTAGAATATACTGTTAGT AAACCTAATGGATCGACCTT TTTAAGAAGGAGATATACAT ATGAACGCACCTGAATCGG TCC	
GT16_CBB1_RHA_R	TCTCCATGGACGCGTGACG TCGACTCTAGAACTGCACG AGTATCGTTTGTTCTGTT	Amplify right homology arm from *C. necator* H16, P_{j5} sequence overhang
GT54_CBB1A_P_{j5}_RHA_F	GGTTTACTAACAGTATATTC TAAATTTCCACCTGTGTCAA TAACGGTTTTTATATCCGCT GGTTCCAGGAGGGTTGGCT GGGATT	
pMTL70641::P_{cbbL} *cbbR*::P_{trp}		
GT120_CBB6_P_{trp}	GGACCGATTCAGGTGCGTTC ATATGTCTTTCTCCTTGTCTA TGAGGAGTCTGTGCGTACTA GTTAACTAGTTCGATGATTAA TTGTCAACAGCCATGGGGTT CCAGGAGGGTTGGCTGGG	Primer coding for P_{trp} sequence
GT105_CBB6_LHA_F	TTTATCAGGAAACAGCTATG ACCGCGGCCGCACATGGTG CCGGCGGTCA	Amplify left homology arm from *C. necator* H16
GT119_CBB6_LHA_R	AGGAGAAAGACATATGAACG CACCTGAATCGGTCC	
GT121_CBB6_RHA_F	TCAACAGCCATGGGGTTCCA GGAGGGTTGGC	Amplify right homology arm from *C. necator* H16
GT110_CBB6_RHA_R	TGCCAAGCTTGCATGTCTGC AGGCCTCGAGACTGCACGAG TATCGTTTGTTC	
pMTL70641::P_{cbbL} *cbbR*:: P_{BAD}		
GT107_CBB7_P_{BAD}_F	CAGGTGCGTTCATATGTATA TCTCCTTCTTAAAAG	Amplify *rrnBT2* terminator *araC* and P_{BAD} from pKTrfp
GT125_CBB7_P_{BAD}_R	CCCTCCTGGAACCGGTACCA AATTAAGCAGAAG	
GT124_CBB7_RHA_F	TTAATTTGGTACCGGTTCCA GGAGGGTTGGCT	Amplify right homology arm from *C. necator* H16

GT110_CBB7_RHA_R	TGCCAAGCTTGCATGTCTGC AGGCCTCGAGACTGCACGAG TATCGTTTGTTC	
pMTL71301::P*BAD* *rfp*		
GT69_*araC*_F	GGAAACAGCTATGACCGCGG CCGCGTTATGACAACTTGAC GGCTACATCA	Amplify *araC* P*BAD* *rfp* from pLO3_PBAD_rfp_for _phaCA
GT70_*rfp*_R	AGCTTGCATGTCTGCAGGCC TCGAGTTAAGCACCGGTGGA GTGACGAC	
pMTL71301::P*BAD* *tdcR* P*tdcB* *//*		
GT5_CBB2A_*araC*_R	AAGACTTTTGCATATGTATA TCTCCTTCTTAAAAGATCTT TTG	Amplify *rrnBT2* terminator *araC* and P*BAD* from pKTrfp
GT18_CBB2A_*araC*_F	TCTTTAAATTCATTAAGCTG GGTACCAAATTAAGCAGAA	
GT17_CBB2A_*tcdB*_F	ATGACCATGATTACGAATTC GAGCTCATTTTCTCCTTTAC TATAATATTT	Amplify P*tcdB* from WOK16 vector system (CRG culture collection)
GT17_CBB2A_*tcdB*_R	TAATTTGGTACCCAGCTTAA TGAATTTAAAGAAATATTTA C	
GT6_CBB2A_*tcdR*_F	AGGAGATATACATATGCAAA AGTCTTTTTATGAATTAATTG	Amplify *tcdR* from WOK16 vector
GT6_CBB2A_*tcdR*_R	TGGACGCGTGACGTCGACT CTAGAGGATCCTTACAAGTT AAAATAATTTTCATAGTCTT TTTTTATTTTTTTAAATG	system (CRG culture collection)
pMTL71301::P*BAD* *tdcR* P*tdcB* *rfp*		
GT50_CBB2A_*rfp*_R	AGCTTGCATGTCTGCAGGC CTCGAGTTACAAGTTAAAAT AATTTTCATAGTCTTTTTTT ATTTTTTTAAATG	Amplify orthogonal expression cassette from pMTL71301::P*BAD* *tcdR* P*tdcB* *//*
GT122_CBB2A_*rfp_tdcB* _F	TACTCGCCATACTAGTATTTT CTCCTTTACTATAATATTTTT ATTGAATATTTTTAC	
GT39_CBB2A_*rfp*_F	AAACAGCTATGACCGCGGCC GCTTAAGCACCGGTGGAGTG	Amplify *rfp* from pKTrfp
GT123_CBB2A_*rfp*_R	GTAAAGGAGAAAATACTAGT ATGGCGAGTAGCGAAGACGT TATC	
pMTL71301::P*BAD* *//* P*tdcB* *rfp*		
GT51_CBB2A_*rfp_tdcB*_ F	TACTCGCCATATTTTCTCCT TTACTATAATATTTTTATTG AATATTTTTAC	Amplify *rrnBT2* terminator *araC* P*BAD* and P*tdcB* from pMTL71301::P*BAD* *tdcR* P*tdcB* *//*
GT104_CBB2C_P*BAD*_R	AGCTTGCATGTCTGCAGGC CTCGAGATGTATATCTCCTT CTTAAAAGATCTTTTG	
GT39_CBB2A_*rfp*_F	AAACAGCTATGACCGCGGC CGCTTAAGCACCGGTGGAG TG	Amplify *rfp* from pKTrfp

GT123_CBB2A_*rfp*_R	GTAAAGGAGAAAATACTAG TATGGCGAGTAGCGAAGAC GTTATC	
pMTL71301::P$_{BAD}$ *tcdR // rfp*		
GT50_CBB2A_*rfp_tcdR*_R	AGCTTGCATGTCTGCAGGC CTCGAGTTACAAGTTAAAAT AATTTTCATAGTCTTTTTTT ATTTTTTTAAATG	Amplify from pMTL71301::P$_{BAD}$ *tcdR* P$_{tdcB}$ //
GT65_CBB2A_F	TACTCGCCATGCTGGGTACC AAATTAAGCAGAA	
GT39_CBB2A_*rfp*_F	AAACAGCTATGACCGCGGC CGCTTAAGCACCGGTGGAG TG	Amplify *rfp* from pKTrfp
GT64_CBB2A_R	ATTTGGTACCCAGCATGGC GAGTAGCGAAGACGTTATC	
pMTL71301::P$_{BAD}$ *Hn*CB		
GT23_aCB1C_*araC*_F	TGACCGCGGCCGCTGTATC CATATGTTATGACAACTTGA CGGCTACATCA	Amplify *rrnBT2* terminator *araC* and P$_{BAD}$ from pKTrfp
GT23_aCB1C_*araC*_R	CTATACTTTTTAACTGCCATA TGTATATCTCCTTCTTAAAAG ATCT	
GT24_aCB1C_aCB1_F	TTTAAGAAGGAGATATACAT ATGGCAGTTAAAAAGTATA GTGCTGG	Amplify first fragment of *cso* operon from
GT8_aCB1_R	GCTCGGCAAGTGCCCGGCTC GTTTCCAGTCAC	pHnCBS1D
GT9_aCB2_F	GGAAACGAGCCGGGCACTTG CCGAGCAGTCAC	Amplify second fragment of *cso*
GT9_aCB2_R	TAGCCCAACCATCGCTGCG GTTTGCGTTGAG	operon from pHnCBS1D
GT10_aCB3_F	CGCAAACCGCAGCGATGGTT GGGCTAAAGGAAATG	Amplify third fragment of *cso*
GT10_aCB3_R	TCTCCATGGACGCGTGACG TCGACTCTAGATTAGAACCC TTCAGCGCG	operon from pHnCBS1D
pMTL71301::P$_{phaC}$ *Hn*CB		
GT22_aCB1B_aCB1_F	AGTGACGGCAGAGAGACAA TCAACATATGGCAGTTAAAA AGTATAGTGCTGG	Amplify first fragment of *cso* operon from
GT8_aCB1_R	GCTCGGCAAGTGCCCGGCT CGTTTCCAGTCAC	pHnCBS1D
GT9_aCB2_F	GGAAACGAGCCGGGCACTT GCCGAGCAGTCAC	Amplify second fragment of *cso*
GT9_aCB2_R	TAGCCCAACCATCGCTGCG GTTTGCGTTGAG	operon from pHnCBS1D
GT10_aCB3_F	CGCAAACCGCAGCGATGGTT GGGCTAAAGGAAATG	Amplify third fragment of *cso*
GT10_aCB3_R	TCTCCATGGACGCGTGACG TCGACTCTAGATTAGAACCC TTCAGCGCG	operon from pHnCBS1D

pMTL71301::*panC*::P$_{phaC}$ *Hn*CB Act
pMTL71301::*panC*::P$_{BAD}$ *Hn*CB Act

GT132_aCB6_F	TCTGAATAGCTAATGTCTAG TAGGGAAGATGCGC	Amplifying the *cbbO* and *cbbQ* fragment
GT133_aCB6_R	GCTTGCATGTCTGCAGGCC TCGAGTTATTTGGTCAATGC AATGAACAGC	

pMTL71301::P$_{phaC}$ *Hn*CbbLS

CAL61_P$_{phaC}$_*cbbLS*-Hn_F	CGGCAGAGAGACAATCAAC ATATGGCAGTTAAAAAGTAT AGTGCTGGTGTAAAAG	Amplify fragment *cbbLS* from pMTL71301::P$_{phaC}$ *Hn*CB
CAL62_P$_{phaC}$_*cbbLS*-Hn_R	GGATCCCGGGCCCGCGGTA CCTTAGTTGCCGCGGTAGA CCACGAA	

pMTL71301::P$_{phaC}$ *Cn*CbbLS

CAL57_P$_{phaC}$_cbbLS-Cn_F	CGGCAGAGAGACAATCAAC ATATGAACGCACCTGAATC GGTCCAAGCCAAGCC	Amplify fragment *cbbLS* from *C. necator* gDNA.
CAL58_P$_{phaC}$_cbbLS-Cn_R	GGATCCCGGGCCCGCGGTA CCTCAGTAGCGGCTGCCCT CGGGGCG	

pMTL71301::P$_{phaC}$ *Mut*CbbLS

CAL59_P$_{phaC}$_cbbL-MUT_R	GCGCCAGGGCTCAGGCCAC CGACGCGGTGGGCGC	Amplify fragment mutant *cbbL* from pUC57::*cbbL* mut
CAL57_P$_{phaC}$_cbbLS-WT_F	CGGCAGAGAGACAATCAAC ATATGAACGCACCTGAATC GGTCCAAGCCAAGCC	
CAL60_P$_{phaC}$_cbbS-WT_F	GGTGGCCTGAGCCCTGGCGC GACGCCACCAAAGC	Amplify fragment *cbbS* from *C. necator* gDNA
CAL58_P$_{phaC}$_cbbLS-Cn_R	GGATCCCGGGCCCGCGGTA CCTCAGTAGCGGCTGCCCT CGGGGCG	

pMTL71301::*panC*::P$_{phaC}$ *Cn*CbbLS *HnCB*

GT150_*cbbLS*_F	AGTGACGGCAGAGAGACAA TCAACATATGAACGCACCTG AATCGGTCCAAGCCAAGCC	Amplify *cbbLS* from *C. necator* gDNA
GT151_*cbbLS*_R	TGACAGACTTGACTCAGTA GCGGCTGCCCTCGGGGCG	
GT152_aCB2_F	CAGCCGCTACTGAGTCAAG TCTGTCATTGCGCTGTGC	Amplify second fragment of the *cso* operon from pMTL71301::P$_{phaC}$ *Hn*CB
GT8_aCB1_R	GCTCGGCAAGTGCCCGGCT CGTTTCCAGTCAC	
GT9_aCB2_F	GGAAACGAGCCGGGCACTT GCCGAGCAGTCAC	Amplify third fragment of the *cso* operon from pMTL71301::P$_{phaC}$ *Hn*CB
GT9_aCB2_R	TAGCCCAACCATCGCTGCG GTTTGCGTTGAG	
GT10_aCB3_F	CGCAAACCGCAGCGATGGT TGGGCTAAAGGAAATG	Amplify fourth fragment of the *cso*

GT10_aCB3_R	TCTCCATGGACGCGTGACG TCGACTCTAGATTAGAACCC TTCAGCGCG	operon from pMTL71301::P*phaC* *Hn*CB
pMTL71301::*panC*::P*phaC* *Mut*CbbLS *HnCB*		
GT150_*cbbLS*_F	AGTGACGGCAGAGAGACAA TCAACATATGAACGCACCT GAATCGGTCCAAGCCAAGC C	Amplify *cbbLS* from pMTL71301::P*phaC* *Mut*CbbLS
GT151_*cbbLS*_R	TGACAGACTTGACTCAGTA GCGGCTGCCCTCGGGGCG	
GT152_aCB2_F	CAGCCGCTACTGAGTCAAG TCTGTCATTGCGCTGTGC	Amplify second fragment of the *cso*
GT8_aCB1_R	GCTCGGCAAGTGCCCGGCT CGTTTCCAGTCAC	operon from pMTL71301::P*phaC* *Hn*CB
GT9_aCB2_F	GGAAACGAGCCGGGCACTT GCCGAGCAGTCAC	Amplify third fragment of the *cso*
GT9_aCB2_R	TAGCCCAACCATCGCTGCG GTTTGCGTTGAG	operon from pMTL71301::P*phaC* *Hn*CB
GT10_aCB3_F	CGCAAACCGCAGCGATGGTT GGGCTAAAGGAAATG	Amplify fourth fragment of the *cso*
GT10_aCB3_R	TCTCCATGGACGCGTGACG TCGACTCTAGATTAGAACCC TTCAGCGCG	operon from pMTL71301::P*phaC* *Hn*CB
Screening primers		
Genomic replacement		
GT60_CBB_ChII_S_F	ACGTAGCAGAAGAACTGCT CG	Screening primers for mutants annealing
GT61_CBB_ChII_S_R	ACGTTCTAGGGATTCTGGC	externally to the replacement location on *C. necator* Chr II
GT67_CBB_cbbR_F	AGATCTTCGTCACCGTGG	Screening primers for
GT68_CBB_RHA_R	CTTTGTCATCCGAAGGCC	mutants annealing internally to the replacement location on *C. necator* Chr II, fw primer anneals in a unique location only present in *cbbR* on Chr II
GT27_CBB2A_S_araC_R	CGACGCTGGCGATATCAAA ATTGCT	Screening primers for mutants, rv
GT60_CBB_ChrII_S_F	ACGTAGCAGAAGAACTGCT CG	annealing in the P*BAD* region and fw on *C. necator* H16 Chr II
Genomic deletions		
GT138_*panC*_Ex_F	CCGCAGGACCTGTACATCC	Screening primers for
GT139_*panC*_Ex_R	GTATATCGGCCTGAGCCTT G	*panC* deletion annealing externally to the deletion location
GT140_*panC*_Int_F	CCAGCTTGGCGGCAGTCAG	

GT141_*panC*_Int_R	CAACTGCGCGGCCAGAAC	Screening primers for *panC* deletion annealing internally to the deletion location
GT142_*cbbLS*_Ex_chr_F	GCAGTGTGCTCCAGCAAATAGGC	Screening primers for *cbbLS* deletion from Chr II annealing externally to the deletion location
GT143_*cbbLS*_Ex_chr_R	GCTTGCATGGCTTACTCCTTGGC	
GT144_*cbbLS*_Ex_m_F	GAAGGTTTCGTGGCGTAGTTCTTGCAGATCG	Screening primers for *cbbLS* deletion from megaplasmid annealing externally to the deletion location
GT145_*cbbLS*_Ex_m_R	GTAGTAGGCCTCGTCGATGAAGAGCACCC	
GT146_*phaCAB*_Ex_F	TATGCGGCCGCGCTGCCTCACTCGTCCTT	Screening primers for *phaCAB* operon deletion annealing externally to the deletion location
GT147_*phaCAB*_Ex_R	TATCTCGAGCGTAGAAGCGGATGATCTG	
Expression vectors carrying cso operon		
GT19_aCB_F	TTATCAGGAAACAGCTATGACCGCGGC	Screening primers for P_{phaC} *Hn*CB and P_{phaC} *Hn*CB Act constructs
GT126_aCB_P_{phaC}_CB_R	TTCCGGCATCCAATAGGTCTGCC	
GT19_aCB_F	TTATCAGGAAACAGCTATGACCGCGGC	Screening primers for P_{BAD} *Hn*CB construct
GT23_aCB_P_{BAD}_CB_R	CTATACTTTTTAACTGCCATATGTATATCTCCTTCTTAAAAGATCT	
GT19_aCB_F	TTATCAGGAAACAGCTATGACCGCGGC	Screening primers for P_{phaC} *Cn*CbbLS *Hn*CB and P_{phaC} *Mut*CbbLS *Hn*CB
GT153_aCB_R	GTTCAGGCAGTATTCAAGCTGGCTG	
Sequencing primers		
Genomic integration vectors		
M13_Rev	CAGGAAACAGCTATGAC	Sequencing primers annealing on pMTL70641 based vectors
M13_Fw	GTAAAACGACGGCCAG	
GT60_CBB1_S_LHA_F	ACGTAGCAGAAGAACTGCTCG	Sequencing primer annealing on the left homology arm of *C. necator* H16 Chr II
GT61_CBB1_S_RHA_R	ACGTTCTAGGGATTCTGGC	Sequencing primer annealing on the right homology arm of *C. necator* H16 Chr II
GT27_CBB2A_*araC*_S_F	CGACGCTGGCGATATCAAAATTGCT	Sequencing primer annealing on *araC*

Expression vectors carrying orthogonal expression system and controls

M13_Rev	CAGGAAACAGCTATGAC	Sequencing primers annealing on pMTL71301 based vectors
M13_Fw	GTAAAACGACGGCCAG	
GT27_CBB2A_*araC*_S_F	CGACGCTGGCGATATCAAAATTGCT	Sequencing primer annealing on *araC*
GT51_CBB2A_*rfp_tdcB*_F	TACTCGCCATATTTTCTCCTTTACTATAATATTTTTATTGAATATTTTTAC	Sequencing primer annealing across *rfp* and P_{tcdB}
GT65_CBB2A_F	TACTCGCCATGCTGGGTACCAAATTAAGCAGAA	Sequencing primer annealing across *rfp* and *rrnBT2*

Expression vectors carrying cbbLS genes

M13_Rev	CAGGAAACAGCTATGAC	Sequencing primers annealing on pMTL71301 based vectors
M13_Fw	GTAAAACGACGGCCAG	
GT136_F	TGACCTGTCGCTGTTCGAGG	Sequencing primers annealing across *Cn* and *MutcbbL*
GT137_F	TTCCGCGTGATTGCCAAATGG	
GT22_aCB1B_aCB1_F	AGTGACGGCAGAGAGACAATCAACATATGGCAGTTAAAAAGTATAGTGCTGG	Sequencing primers annealing across *HncbbL*
GT30_aCB1A_S_1.1_F	AGACGCGACCGAAACTGCTGAAGCC	

Expression vectors carrying cso operon

M13_Rev	CAGGAAACAGCTATGAC	Sequencing primers annealing on pMTL71301 based vectors
M13_Fw	GTAAAACGACGGCCAG	
GT27_CBB2A_S_araC_F	CGACGCTGGCGATATCAAAATTGCT	Sequencing primer annealing on *araC* gene
GT28_CBB2A_S_P_{BAD}_F	CATAAGATTAGCGGATTCTACCTGA	Sequencing primer annealing on P_{BAD} promoter region
GT63_S_ P_{BAD}_F	CATATTGCATCAGACATTGCC	Sequencing primer for *Hn*CB construct annealing on P_{BAD} promoter region
GT30_aCB1A_S_1.1_F	AGACGCGACCGAAACTGCTGAAGCC	Sequencing primer for *Hn*CB construct annealing on the first fragment of *cso* operon
GT31_aCB1A_S_1.2_F	AACCATGGCTGAAATGCAGGATTAC	Sequencing primer for *Hn*CB construct annealing on the first fragment of *cso* operon

GT32_aCB1A_S_1.3_F	TTCTGTAAATGAAGTGCCCACTGAT	Sequencing primer for *Hn*CB construct annealing on *csoS2* in the first fragment of *cso* operon
GT9_aCB2_F	GGAAACGAGCCGGGCACTTGCCGAGCAGTCAC	Sequencing primer for *Hn*CB construct annealing on *csoS2* in the second fragment of *cso* operon
GT33_aCB1A_S_2.1_F	ACCAGGTTCATGCTCTCGGGTTACA	Sequencing primer for *Hn*CB construct annealing on *csoS2* in the second fragment of *cso* operon
GT34_aCB1A_S_2.2_F	AAAAGCCTGAGTTGGTATGTGCACC	Sequencing primer for *Hn*CB construct annealing on *csoS2* in the second fragment of *cso* operon
GT35_aCB1A_S_2.3_F	AACAGCAGCCACGGAACAATTCAGT	Sequencing primer for *Hn*CB construct annealing on *csoS3* in the second fragment of *cso* operon
GT10_aCB3_F	CGCAAACCGCAGCGATGGTTGGGCTAAAGGAAATG	Sequencing primer for *Hn*CB construct annealing on *csoS3* in the third fragment of the *cso* operon
GT36_aCB1A_S_3.2_F	GTTCCGACCTAATCGCAACACGCA	Sequencing primer for *Hn*CB construct annealing on *csoS4B* the third fragment of *cso* operon
GT62_aCB_S_3.3_F	GTCGTCAATTTGTTGGTGGC	Sequencing primer for *Hn*CB construct annealing on *csoS1A* the third fragment of *cso* operon

Expression vector carrying cso operon with activases

GT132_aCB_csoS1D_S_F	TCTGAATAGCTAATGTCTAGTAGGGAAGATGCGC	Sequencing primer for *Hn*CB Act construct annealing on the third fragment of *cso* operon
GT127_aCB6_S_F	AGCTCATGTGACTCTGATCGAT	Sequencing primer for *Hn*CB Act construct annealing on the third fragment of *cso* operon
GT128_aCB6_S_F	ACGAAGGCATTTCAACCCGTC	Sequencing primer for *Hn*CB Act construct annealing on *cbbQ*

GT129_aCB6_S_F	TCTGGTCGATAACGAGCGAA	Sequencing primer for *Hn*CB Act construct annealing on *cbbO*
GT133_aCB6_R	GCTTGCATGTCTGCAGGCCTCGAGTTATTTGGTCAATGCAATGAACAGC	Sequencing primer for *Hn*CB Act construct annealing on pMTL71301 vector
Expression vectors carrying Cn and MutCbbLS with cso operon		
M13_Rev	CAGGAAACAGCTATGAC	Sequencing primers for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing on pMTL71301 based vectors
M13_Fw	GTAAAACGACGGCCAG	
GT150_cbbLS_F	AGTGACGGCAGAGAGACAATCAACATATGAACGCACCTGAATCGGTCCAAGCCAAGCC	Sequencing primer for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing across P_{phaC} and *cbbL* in the first fragment of the constructs
GT137_F	TTCCGCGTGATTGCCAAATGG	Sequencing primer for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing across *cbbL* in the first fragment of the constructs
GT154_F	GGTGGCCTGAGCCCTGGCGCGACGCCACCAAAGC	Sequencing primer for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing in the intercitronic region between *cbbL* and *cbbS* in the first fragment of the constructs
GT32_aCB1A_S_1.3_F	TTCTGTAAATGAAGTGCCCACCTGAT	Sequencing primer for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing on *csoS2* in the second fragment of the constructs
GT9_aCB2_F	GGAAACGAGCCGGGCACTTGCCGAGCAGTCAC	Sequencing primer for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing on *csoS2* in the third fragment of the constructs
GT33_aCB1A_S_2.1_F	ACCAGGTTCATGCTCTCGGGTTACA	Sequencing primer for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing

		on *csoS2* in the third fragment of the constructs
GT34_aCB1A_S_2.2_F	AAAAGCCTGAGTTGGTATGTGCACC	Sequencing primer for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing on *csoS2* in the third fragment of the constructs
GT35_aCB1A_S_2.3_F	AACAGCAGCCACGGAACAATTCAGT	Sequencing primer for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing on *csoS3* in the third fragment of the constructs
GT10_aCB3_F	CGCAAACCGCAGCGATGGTTGGGCTAAAGGAAATG	Sequencing primer for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing on *csoS3* in the fourth fragment of the constructs
GT36_aCB1A_S_3.2_F	GTTCCGACCTAATCGCAACACGCA	Sequencing primer for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing on *csoS4B* in the fourth fragment of the constructs
GT62_aCB_S_3.3_F	GTCGTCAATTTGTTGGTGGC	Sequencing primer for *Cn* and *Mut*CbbLS *Hn*CB constructs annealing on *csoS1A* in the fourth fragment of the constructs

Table 2.4: List of oligonucleotides used in this study.

2.3.3 Isolation of plasmid DNA

Plasmid DNA isolation was performed using the QIAprep® Spin Miniprep Kit (Qiagen Group) in accordance with manufacturer's protocol. 3-5 mL of an overnight liquid culture was centrifuged, and plasmid DNA purified.

2.3.4 Isolation of chromosomal DNA

Genomic DNA extraction was performed using the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich) in accordance with manufacturer's protocol. 1.5 mL of an overnight liquid culture was centrifuged and DNA purified.

2.3.5 PCR analysis of DNA

Oligonucleotide primers from Sigma-Aldrich and Eurofins Scientific were used unmodified for PCR amplification. Primers were provided lyophilised and resuspended in nuclease-free H_2O (Ambion) to a final concentration of 100 pmol/µL.

PCRs were performed using DreamTaq PCR Master Mix (x2) (Thermofisher Scientific) and Q5® High-Fidelity 2X Master Mix (NEB) following the manufacturers' protocols. DreamTaq PCR Master Mix was used for screening of clones by colony PCR whereas Q5® polymerase was used for amplification of sequences for cloning. For both procedures, the reaction mix contained the Master Mix, 1 µM of each primer and was supplemented with 5 % DMSO. PCR amplifications were carried out in TProfessional Trio Thermocycler (Biometra, Analytikjena) and analysed by agarose gel electrophoresis.

2.3.6 Restriction analysis of plasmid DNA

Restriction enzymes used in this study were FastDigest enzymes (Thermofisher Scientific) used according to manufacturer's protocols. Reactions generally contained 0.100 – 1 µg DNA, 0.5 Units of each enzyme (approximately 1 µL), 2 µL of 10x Buffer and H_2O up to 20 µL. The digestions were performed at 37 °C for 1.5 h and analysed by agarose gel electrophoresis. (Unit definition: One unit is defined as

the amount of enzyme required to digest 1 μg of λ DNA in 1 h at 37°C in a total reaction volume of 50 μL).

2.3.7 Agarose gel electrophoresis

DNA electrophoresis was performed in 1% (w/v) agarose gels (with SYBR Safe DNA Gel Stain 1 μL/10 mL agarose, Thermofisher Scientific), and was used to check restriction digestion reactions and PCR amplicons in a Compact XS/S Horizontal Gel Electrophoresis Apparatus (Biometra). To load the gel, 5 μL of 2-Log DNA ladder (0.1–10.0 kb; NEB) was used as DNA molecular weight marker and 20 μL of each sample, restriction digest or PCR amplicon, were pipetted per well. 2 μL of 10X FastDigest Green Buffer were added to NEB restriction digestions and Q5® PCR amplifications prior to loading; samples digested with FastDigest enzymes and DreamTaq PCR amplifications could be loaded directly onto the gel. The gels were run in 1x TAE buffer (PanReac AppliChem, TAE buffer 50X, diluted to 1X following manufacturer's protocol) at 100 V for approximately 1 h and then visualized using a Gel Doc™ XR+ (Bio-Rad).

2.3.8 Purification of PCR products *via* gel extraction

PCR amplicons and relevant digestion bands were excised from 1% (w/v) agarose gels when needed and recovered with the Zymoclean™ Gel DNA Recovery Kit (Zymo Research) following manufacturer's protocol.

2.3.9 DNA quantification

DNA concentration and purity were assessed using the SimpliNano microvolume spectrophotometer (GE Healthcare Life Sciences).

2.3.10 HiFi DNA Assembly

The assembly of constructs was achieved using HiFi DNA Assembly (NEB) according to manufacturer's protocol. As recommended, a total of 0.03–0.2 pmols of DNA fragments was used during the assembly of 1 or 2 fragments and 0.2–0.5 pmols were used when 4–6 fragments were assembled. For optimal assembly conditions the number of pmols of each fragment needs to be calculated, based on fragment length and weight.

NEBiocalculator tool was used for the task following the formula:

$$\textbf{pmols} = \frac{\textbf{(weight in ng) x 1,000}}{\textbf{base pairs x 650 daltons}}$$

As suggested, DH5-α competent *E. coli* cells were transformed with 2-5 µL of the assembled product, following the transformation protocol (section 2.4.2 or 2.4.3).

2.3.11 Sequencing of DNA constructs

DNA sequencing was performed by the Eurofins MWG Custom DNA Tube Sequencing Service using the cycle sequencing technology (dideoxy chain termination / cycle sequencing) on ABI 3730XL sequencing machines.

2.4 Transformation of bacterial strains

2.4.1 Preparation of *E. coli* electrocompetent cells

A total volume of 100 mL LB broth was inoculated with 1 mL fresh *E. coli* overnight culture and incubated in a 500 mL conical Duran flask at 37 °C at 200 rpm until mid to late exponential phase (OD_{600} of 0.5-1.0). The flask was incubated on ice for 15 min and after centrifugation at 4,000 x g for 15 min. The supernatant was discarded, and the pellet resuspended in 100 mL dH_2O (pre-chilled on ice) then centrifuged as in the previous step. As before, the supernatant was discarded, and the pellet resuspended in 2 mL 10% glycerol (v/v) (pre-chilled on ice) then centrifuged as in the previous step. As final step, the supernatant was discarded and the pellet resuspended in 300 µL

of 10% (v/v) glycerol (pre-chilled on ice). Aliquots of 50 µL were prepared in 1.5 mL Eppendorf tubes and stored at -80 °C.

2.4.2 Electroporation of *E. coli* electrocompetent cells

100 ng of purified plasmid DNA was added to a 50 µL aliquot of electrocompetent *E. coli* cells. Once mixed, cells were transferred into pre-chilled 2 mm electroporation cuvettes (Scientific Laboratory Supplies, SLS), electroporated at 2.5 kV using a MicroPulser Electroporator (Bio-Rad) and resuspended in 950 µL of SOC medium (ThermoFisher Scientific). The samples were incubated at 37° C at 200 rpm for 1 h. The cells were then spread,100 µL to 200 µL volumes, on selective LB agar plates and incubated O/N at 37° C.

2.4.3 *E. coli* transformation using $CaCl_2$

Chemically competent *E. coli* cells for transformation were prepared by diluting an O/N culture 1:100 in 100 mL warm LB and growing it in a 1 L conical Duran flask at 37° C / 200 rpm until OD_{600} 0.5. The culture was then cooled down on ice and 50 ml of the culture were centrifuged at 4000 rpm for 6 min at 4 °C. The pellet was resuspended in 16.6 mL ice-cold 100 mM $CaCl_2$ and centrifuged again. The pellet was then resuspended in 16.6 mL of the same ice-cold solution and left in ice-water for 20 min before centrifuging again. The final pellet was resuspended in 10 mL ice-cold 100 mM $CaCl_2$, 10 % glycerol, and aliquots of 200 µL were prepared in 1.5 mL Eppendorf tubes and stored at -80° C.

For the transformation, 50 µL of the cells were mixed with 5 µL of plasmid DNA, incubated on ice for 30 min and then heat-shocked at 42 °C for 1 min, prior to addition of 950 µL SOC medium. Cells were incubated for 1 h at 37 °C and 200 rpm. After this stage, the cells were spread in 100 µL to 200 µL volumes on selective LB agar plates and incubated O/N at 37 °C.

2.4.4 Preparation of *C. necator* H16 electrocompetent cells (Ehsaan M, unpublished)

A 10 mL culture of Hanahan's Broth (SOB Medium, H8032 Sigma) was inoculated with a heavy loop of fresh, plate-grown *C. necator* H16 cells and vortexed for thorough resuspension. The initial 10 mL culture was serially diluted to 10^{-1}, 10^{-2} and 10^{-3} in 10 mL SOB medium each and incubated in an aerobic shaker at 30 °C and 200rpm O/N (approximately 10-14 h). The following day, the OD_{600} of the dilutions was measured and the culture with the lowest OD used to inoculate 50 mL of SOB broth in a 250 mL baffled flask to an OD_{600} of 0.055-0.075. The culture was incubated for 2 hours at 30 °C and 250 rpm until OD_{600} reached 0.2 to 0.3. 25 mL of the culture were centrifuged at 7000 rpm for 10 min at 4 °C in a pre-cooled benchtop centrifuge. The harvested cells were washed with 10 mL of pre-chilled, filter sterilised 1mM HEPES buffer, pH 7.0 (adjusted with NaOH) and then centrifuged as the previous step. The wash was repeated with 5 mL of the previous solution and cells centrifuged as described before. The final pellet was resuspended in the same solution up to an OD_{600} of 5. Aliquots of 100 µL cell suspension in 1 mM HEPES buffer / 10 % glycerol (w/v) were prepared in 1.5 mL Eppendorf tubes and stored at -80 °C.

2.4.5 Electroporation of *C. necator* H16 electrocompetent cells (Ehsaan M, unpublished)

100 ng–500 ng of DNA were added to a 100 µL aliquot of C. necator H16 electrocompetent cells in a pre-chilled electroporation cuvette of 0.2 cm gap width and incubated on ice for 5 min. After ice incubation, the cell suspension was electroporated at 2.5 kV, 25 µF, and 200 Ω settings in a MicroPulser Electroporator (Bio-Rad). Immediately after electroporation, 900 µL of SOC medium were added to the cell suspension and transferred into a 50 mL polypropylene Falcon tube. Cells were incubated at 30 °C, 250 rpm for 2 h. After this stage, the cells were spread in 100 µL to 200 µL volumes on selective LB agar

plates and incubated at 30 °C. Transformants appeared after 48 h of incubation.

2.4.6 Mobilisation of plasmid from *E.coli* to *C. necator* H16 by conjugation

5 mL seed cultures of *C. necator* wild type or deletion strains grown in LSLB medium and *E. coli* S-17-1 λpir carrying the plasmid to be transferred in LB medium were prepared and incubated O/N for about 20 h at 30 °C, 200 rpm and 16 h at 37 °C, 200 rpm, respectively. For the conjugation procedure, recipient and donor cells were mixed at a ratio of 1:1 (v/v). 1 mL of each culture was centrifuged for 4 min at 5,000 rpm, separately. The supernatant was discarded, and the resulting pellet resuspended in 1 mL LB and centrifuged as in the previous step. Once the supernatant was discarded, the pellet was resuspended in 500 µL of LB, mixed and centrifuged as the previous step. For the mating, the pellet was resuspended in 30 µL LB, spotted on LB agar plates and incubated at 30 °C for 6 h. Controls of recipient and donor cells were treated similarly. Following the mating period, the cells were removed from the plate with a sterile loop, resuspended in 1 mL phosphate buffered saline (PBS) and 10-fold serially diluted. 100 µL of dilutions 10-4 and 10-5 were plated on MM selective agar plates. Transconjugants appeared after 2 to 3 days of incubation at 30 °C and were re-streaked twice onto new MM selective agar plates. Transconjugants were inoculated in LSLB sucrose 15 % medium without antibiotic and grown overnight. Approximately 10^8 cells per plate (100 µL of dilutions 10^{-4} – 10^{-6}) were spread on LSLB plates containing 15% sucrose. Sucrose–resistant colonies appeared after 48-72 h of incubation at 30 ºC. Individual sucrose–resistant clones were picked and purified in LB agar plates, and replica plating on LB Tet (15 µg/mL) plates to confirm the loss of the pMTL70641 plasmid. Confirmation of transconjugants was performed by colony PCR as stated in section 2.3.5.

2.4.7 Whole genome sequencing

Further confirmation of the deletion/insertion mutant strains was achieved performing a whole genome sequencing analysis. Genomic DNA extraction was performed as described in section 2.3.4, quantified as described in section 2.3.9 and sent for analysis at MicrobesNG spin-out from the University of Birmingham.

2.5 Plasmids and strain construction

2.5.1 Construction of deletion and insertion plasmids

The suicide plasmid pMTL70641 from the modular vector system designed by Muhammad Ehsaan was used to carry out deletions and simultaneous insertion of genes in *C. necator* H16. This plasmid contains a tetracycline resistance marker and the *sacB* negative selection marker that allows for counter-selection in sucrose (Gay *et al.*, 1985; Lenz *et al.*, 1994).

The construction of the plasmids consisted of two homology arms, of 750 and 1000 bp, flanking the region of interest and fused via HiFi DNA Assembly (NEB) following manufacturer's protocol. A series of oligonucleotide primers containing homology regions of the vector and target region were designed using NEBuilder® Assembly Tool (Table 2.4). The inserts were amplified by PCR (section 2.3.5), separated in agarose gel electrophoresis and purified as described in section 2.3.8. The pMTL70641 vector was prepared through restriction enzyme digestion and recovered as described above for the inserts. For final assembly, digested vector and purified homology arms were mixed following the formula from section 2.3.10 in a final volume of 5 µL. 5 µL of 2X NEBuilder HiFi DNA Assembly Master Mix was added and the samples were incubated at 50 °C for 1 h. The product of the assembly reaction was transformed in *E. coli* DH5α following the procedure described in section 2.4.2. The assembled vectors were verified by sequencing (section 2.3.11) and electroporated in *E. coli* S17-1 λpir donor strains (section 2.4.2).

2.5.2 Construction of *C. necator* H16 allelic mutants

The conjugative plasmids containing the desire sequences and *sacB* gene were transferred from *E. coli* S17-1 λpir to *C. necator* H16 by conjugation as described in section 2.4.6. Transconjugants for first cross over were selected by plating on sodium gluconate (0.4%) -MM Tet (15 mg/mL) agar plates, and colonies appeared after 48-72 h of

incubation at 30 °C. Tet resistant single colonies were picked and streaked twice on the same selective agar plates to get rid of possible *E. coli* carry over from the mating step. Purified transconjugants were inoculated in LSLB 15% sucrose medium without antibiotic selection and grown O/N at 30 °C at 200 rpm. O/N cultures were serially diluted and 100 µL of dilutions 10^{-4} – 10^{-6} were spread on LSLB 15% sucrose agar plates to obtain single colonies. Sucrose – resistant colonies following a second crossover appeared after 48-72 h of incubation at 30 °C. Individual sucrose – resistant clones were picked and purified in LB agar plates, and by replica plating on LB Tet (15 µg/mL) to confirm the loss of the pMTL70641plasmid. Screening of the clones was performed by colony PCR and subsequent sequencing confirmation (oligonucleotide primers in Table 2.4).

2.5.3 Construction of expression plasmids

2.5.3.1 Construction of expression plasmids for the evaluation of an inducible orthogonal expression system in *E. coli* and C. *necato*r H16

In order to compare gene expression in response to different induction levels, a modular, broad-host range vector (designated as pMTL70000 series) containing the arabinose-inducible gene expression system was assembled (pMTL71301 P_{BAD}). Subsequently, this plasmid was used to generate all other constructs.

The plasmid pMTL71301 from the modular vector system designed by Muhammad Ehsaan was used as expression system in *E. coli* DH5α and *C. necator* H16. Five vectors were designed for the evaluation of the performance of the inducible orthogonal expression system in the two organisms. The first vector contained the inducible orthogonal system with reporter gene *rfp* (pMTL71301::P_{BAD} *tdcR* P_{tcdB} *rfp*). The *rfp* gene encodes for a monomeric red fluorescent protein (mRFP1) firstly published in 2002 (Campbell *et al.*, 2002). The pBBR1-2USER-LYC1 plasmid (CRG culture collection) was used as template for *rfp* sequence amplification. The other three vectors have

the same architecture as the latter, but they lack one component of the system. The second vector lacks the promoter region P_{tcdB} (pMTL71301::P_{BAD} *tdcR // rfp*), the third vector lacks *tcdR* gene (pMTL71301::P_{BAD} // P_{tdcB} *rfp*) and the fourth lacks the reporter gene *rfp (*pMTL71301::P_{BAD} *tdcR* P_{tdcB} //). The fifth vector carries the inducible promoter P_{BAD} driving the expression of *rfp* (pMTL71301::P_{BAD} *rfp*) and it was used as internal inducible expression reference.

The architecture of the inducible, orthogonal expression system contains the inducible promoter P_{BAD} driving the expression of the alternative sigma factor *tdcR*, which will titrate the promoter region P_{tcdB} in response to the arabinose induction levels. The coding sequences for the alternative sigma factor and *rfp* are arranged in opposite direction. To prevent transcriptional read-through, the coding sequence of *araC* is terminated by a terminator sequence (TrrnB2), and both *tcdR* and *rfp* genes are flanked by a double terminator region on the plasmid.

Oligonucleotide primers (Table 2.4) were used to amplify the arabinose-inducible promoter and the *rfp* reporter gene from pKTrfp; the alternative sigma factor *tcdR* with its promoter region P_{tdcB} from WOK16 modular system (CRG culture collection). Purification of inserts, digestion of the vectors and assembly were carried out following procedures described above for the deletion vectors (section 2.5.1). The resulting assembly reactions were transformed in *E. coli* DH5α following the procedure described in section 2.4.3. The constructs were verified by sequencing and electroporated in *C. necator* H16 (section 2.4.5).

2.5.3.2 Construction of plasmids for the heterologous expression of *cso* operon from *H. naepolitanus* in *E. coli* and *C. necator* H16

To attempt the heterologous expression of the *cso* operon from *H. naepolitanous* two modular broad-host range vector (designated as

pMTL70000 series) were tested. One containing the arabinose-inducible expression promoter (pMTL71301 P_{BAD}) and the other one containing the mild constitutive promoter P_{phaC} (pMTL71301 P_{phaC}-eYFP). The plasmids contain pBBR1 replicon, which allows replication in Gram-negative bacteria and TetA/R resistance marker. To avoid read-trough a double terminator on the plasmid flanks the operon region. Subsequently, these plasmids were used to generate the other constructs.

Oligonucleotide primers (Table 2.4) were used to amplify three fragments of the open reading frame of the *cso* operon from pHnCBS1D (Addgene). Purification of inserts, digestion of the vectors and assembly were carried out following procedures described above for the deletion vectors (section 2.5.1). The resulting plasmids were transformed in *E. coli* DH5α following the procedure described in section 2.4.3. Constructs were verified by sequencing, transformed in *E. coli* S17-1 λpir and conjugated with *C. necator* H16 and *C. necator* Δ*phaCAB* strains (section 2.4.6).

To assess the ability of the native *C. necator* and heterologous *H. neapolitanus* RubisCO-encoding gene to functionally complement the *C. necator* ΔΔ*cbbLS* mutant, these *cbbLS* genes from the two organisms were cloned in the pMTL71301 modular vector (oligonucleotide primers in Table 2.4), under the control of the constitutive promoter P_{phaC} and transformed in *C. necator* ΔΔ*cbbLS* and *C. necator* Δ*phaCAB* ΔΔ*cbbLS* strains after sequencing.

To increase the plasmid stability in our host organism, a plasmid addiction system was created by cloning the wild type *panC* gene, encoding the enzyme pantothenate synthetase, under the control of its native promoter in the pMTL71301::P_{phaC} vector. A *C. necator* pantothenate auxotroph, CbbLS null mutant strain was generated by deleting the *panC* gene (Callum McGregor and Marco Garavaglia unpublished) and, subsequently, the RubisCO-encoding genes were

also knocked out in this strain, thus obtaining the *C. necator* ΔΔ*cbbLS* Δ*panC* mutant. This strain can survive in a pantothenate-free culture media only if transformed with the pMTL71301::*panC* plasmid, since this provides *in-trans* complementation of the Δ*panC* mutation. The same process was applied to the *C. necator* Δ*phaCAB* background strain. The *cso* operon was then traditionally cloned in pMTL71301::*panC*::P_{phaC} backbone using NdeI and XbaI sites (restriction enzymes: Thermo Fisher; T4 DNA ligase: New England BioLabs), downstream of the P_{phaC} promoter. The vector was then confirmed by sequencing and conjugated with *C. necator* ΔΔ*cbbLS,* Δ*panC* and *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* strains.

For the construction of the *cso*-operon with RubisCO activases (*cbbO, cbbQ*), the plasmid pCU57::CbbOQ carrying *cbbO* an *cbbQ* genes was synthesised by GeneScript (New Jersey, United States). Oligonucleotide primers (Table 2.4) were used to amplify one fragment. The plasmid pMTL71301::*panC*::P_{phaC} *Hn*CB Act was created using traditional cloning method where pMTL71301:*panC*::P_{phaC} *Hn*CB and the PCR fragment were digested with XbaI and NheI restriction enzymes (ThermoFisher) and ligated using T4 DNA ligase (New England BioLabs). The construct was sequenced and conjugated with *C. necator* ΔΔ*cbbLS* Δ*panC* strain.

2.5.3.3 Construction of expression plasmids with chimeric *cso* operon replacing *cbbLS* genes from *H. neapolitanus* with *C. necator* wild type and mutant form

From the *in-silico* comparison study between the surface charges of *C. necator* and *H. neapolitanus* RubisCO in dimer (*cbbLS*) and octamer $(cbbLS)_4$ conformation, it appeared that the electrostatic potential of *Hn*RubisCO was negatively charged compared to *Cn*RubisCO in the outer backbone of the octamer. To assess the relevance of the surface charge in the sequestering mechanism of RubisCO during carboxysome assembly, three mutation sites carrying seven aminoacidic mutations were selected in the *C. necator*

cbbL (Thr33, Leu35, Leu76, Asp79, Asp92, Gly96 and Asp97) to modify the electrostatic potential of the outer backbone towards a negative value similar to *Hn*RubisCO. The resulting synthetic *cbbL* gene was synthesised by GeneScript (New Jersey, United States)

Firstly, to evaluate the ability of the native *C. necator* (*Cn*RubisCO) and mutant (*CnMut*RubisCO) RubisCO-encoding gene to functionally complement the *C. necator* ΔΔ*cbbLS* mutant, these *cbbLS* genes were cloned in the pMTL71301 modular vector (oligonucleotide primers in Table 2.4), under the control of the constitutive promoter P_{phaC} and transformed in *C. necator* ΔΔ*cbbLS* after sequencing.

While, to attempt the expression and *in-vivo* activity assay of the chimeric *cso* operon either wild type or mutant form of *cbbLS* genes from *C. necator* were cloned upstream the *cso* operon in the auxotrophic selection vector (pMTL71301::*panC*) under the constitutive promoter P_{phaC} (oligonucleotide primers in Table 2.4). Following sequencing, the constructs were then conjugated in *C. necator* ΔΔ*cbbLS* Δ*panC* strain.

2.6 Protein expression and analysis

2.6.1 Growth conditions for protein analysis

Screening for protein expression was performed in *E. coli* DH5α, *C. necator* H16 wild type, *C. necator* Δ*phaCAB* and derivative deletion strains. Sterile falcon tubes were inoculated from single colonies of *E. coli* and *C. necator* strains. The cells were grown overnight in LB medium at 200 rpm 37 °C or 30 °C, respectively before being diluted into 5 mL of fresh LB media. At this point, different procedures were carried out depending on the nature of the promoter (constitutive or inducible) driving the expression of the carboxysomes. For the cells carrying the plasmid with the constitutive promoter, samples were collected after 6, 12 and 24 h after inoculation. Cells carrying the plasmid containing the inducible promoter were tested with a range of induction levels (2.5, 5 and 10 mM arabinose), and after the stock inducer was added at an OD_{600} of 0.5, samples were collected after 6, 12 and 18 h after induction. Modifications of overnight media and sampling time compared to the standard procedure were reported in the relevant sections.

2.6.2 Denaturing polyacrylamide SDS-PAGE gel electrophoresis analysis

Protein expression analysis was carried out through denaturing SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The OD_{600} of the cultures was measured, 1 mL samples taken and centrifuged at full speed (14,000rpm) for 1 min. Supernatants were discarded and pellets frozen at -20 °C. To normalize the protein concentration and obtain the necessary volume of 1X BugBuster® Protein Extraction Reagent (Merck Millipore), the following formula was used: $[(OD_{600}/0.2)*45]/2$. BugBuster (Merck Millipore) reagent mix was prepared adding 1:10000 Benzonase® (Merck Millipore); 1:50 Protease Inhibitor Cocktail Set VII (Merck Millipore), 1:1000 rLysozyme™ Solution (Merck Millipore) and brought up to volume with PBS. Pellets were

resuspended in the calculated volume and incubated at room temperature (RT) for 20 min with gentle shaking (50 rpm).

After the incubation step, protein samples were centrifuged at 14,000 rpm for 10 min and 10 µL of the supernatant (soluble fraction) added to 10 µL of sample buffer (5 µL NuPAGE™ LDS Sample Buffer (4X) (ThermoFisher Scientific), 3 µL nuclease-free water, 2 µL NuPAGE™ Sample Reducing Agent (10X) (ThermoFisher Scientific)). Samples were boiled at 100 °C for 5 min and then loaded onto precast NuPAGE™ 4-12% Bis-Tris Protein Gels (ThermoFisher Scientific).

To load the gel, 5 µL of PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa (ThermoFisher Scientific) was used as protein molecular weight marker and 10 µL of each sample. The gels were run in an XCell SureLock™ Mini-Cell Electrophoresis System (ThermoFisher Scientific) with 1X NuPAGE™ MES SDS Running Buffer (20X) buffer at 150 V for approximately 1 h. Once run, gels were taken out from the cast, washed briefly with water 3 times and stained with Comassie Blue.

2.6.3 Coomassie Blue staining

SDS-PAGE gels were stained using GelCode™ Blue Stain Reagent (ThermoFisher Scientific) following manufacturer's protocol until bands were visible and for no longer than 2 h. Once stained, gels were washed once with water and left destain overnight.

2.6.4 MALDI-TOF mass spectrometry analysis

Relevant bands were excided from the SDS-PAGE gel and collected in clean Eppendorf tubes. The samples were sent for protein identification to the University of York, Metabolomics and Proteomics Department. The analysis was performed using MALDI-MS Bruker ultraflex III MALDI-TOF/TOF, which allow to obtain higher-confidence identifications by including tandem spectral data in Mascot software (Matrix Science) searches.

2.7 Transmission electron microscopy (TEM) analysis

2.7.1 Sample preparation using iPGel Kit (GenoStaff®, Japan) and TEM imaging

Transmission electron microscopy was used to observe the correct assembly of the caboxysomal shell proteins in *E. coli*, *C. necator* H16 and *C. necator ΔphaCAB* background strains. The cells were grown overnight in LB medium at 200 rpm 30 °C, respectively. 1 mL of each culture was centrifuged for 4 min at 5,000 rpm, separately. The supernatant was discarded and the resulting pellet resuspended in 1 mL PBS and centrifuged as in the previous step. Under the flow hood 1.5 mL of fixative solution (6% glutaraldeyde in 0.1 M cacodylate buffer) was added and incubated at 50 rpm room temperature for 1 h. The samples were centrifuged as in the previous step and resuspended in 1.5 mL of 0.1 M cacodylate buffer for 5 min at 50 rpm. The previous step was repeated twice and the samples were kept in the buffer at 4 °C for 2 days. Previous embedding the cells iPGel Kit (GenoStaff, Japan) was used according to manufacturer's protocol. Solidified samples were washed with 1.5 mL of 0.1 M cacodylate buffer. The supernatant was discarded and the samples resuspended in 1.5 mL of 1% aqueous Osmium Tetroxide (OsO_4) for 1 h. The samples were centrifuged as in the previous steps, the supernatant removed and the samples resuspended in 1mL 0.1 M cacodylate buffer twice. The samples were centrifuged as in previous steps, the supernatant removed and washed with 1 mL of distilled water twice. The samples were progressively dehydrated in graded ethanol series 50% to 100% and in the last step in 100% acetone. For each of the ethanol series 1 mL of ethanol was added to the previously centrifuged samples and kept dehydrate for 10 min. The samples were centrifuged at 5,000 rpm for 2 min and each step was repeated twice. The samples are ready to be infiltrated with a progressive series of acetone/resin solution (1:3 and 1:1). For the 1:3 acetone/resin mix

samples were previously centrifuged as in previous steps and resuspended for 30 min in 1 mL of solution, while for the 1:1 acetone/resin mix samples were left infiltrating O/N. Then, the samples were infiltrated in pure resin for 1 h at 400 mbar twice and as last step embedded in pure resin for 2 days at 70 °C. 70 nm to 90 nm sections were cut using an UC6 ultra microtome (Leica), post-stained with saturated uranyl acetate in 50% ethanol followed by lead citrate stain. The images were captured using a Tecnai G2 12 BioTWIN Transmission Electron Microscope (TEM; FEI company, Eindhoven, The Netherlands) operated at an accelerated voltage of 100 kV and Gatan SIS Megaview IV camera system.

2.7.2 Sample preparation and Cryo-TEM tomography analysis

Pellet from 5 mL of cell culture grown overnight in LB media at 37 or 30 °C, 200 rpm were centrifuged at 5,000 rpm and resuspended in 3 mL PBS. Three microliters were applied to each grid (graphene oxide on holy grid, Quantifoil Micro Tools, Jena, Germany), before plunge freezing into ethane/ propane (Tivol *et al.*, 2008), using gravity plunger (blot time of 1–2 s). The standard EM tilt-series should have been performed on a JEOL 2100F FEG TEM cryo-microscope, equipped with a 2K x 2K CCD camera Gatan Tridiem Filter Spectrometer (Gatan, Pleasanton, CA). Cellular tomograms were planned to be collected from -70 ° to +70 °, with an angular step of 1°. The imaging was never completed due the pandemic outbreak of Covid-19 and subsequent lack of access to the Nanoscale and Microscale Research Centre (nmRC, University of Nottingham) facilities.

2.8 RFP Fluorescence assays

The protocol for fluorescence assay was used according to Hanko et al. 2017 (Hanko *et al.*, 2017) with slight changes to the procedure. For the growth and fluorescence time course experiments, single colonies

of *E. coli* DH5α or *C. necator* H16 were used to inoculate 5 mL of SG-MM in 50 mL culture tubes. They were diluted to 0.2 OD_{600} in 142.5 µL fresh minimal medium into a 96-well clear bottom plate (Greiner Bio One International, Germany; Cat. no 655090) and incubated at 200 rpm and 30 °C until an OD_{600} between 0.3 and 0.5 was reached. Then 7.5 µL of stock inducer at the desired concentration were added to each sample. Fluorescence was measured using an Infinite® M1000 PRO (Tecan, Switzerland) micro plate reader with 585 nm as excitation wavelength and an emission wavelength of 620 nm. Fluorescence and absorbance were measured every 5 min for 16 h in *E.coli* and for 20 h in *C. necator* H16. The gain factor for fluorescence was set manually to 80%. To determine the background autofluorescence, a culture was included in a separate well containing a control strain transformed with the empty plasmid (pMTL71301 P_{BAD}). Prior calculation, fluorescence and absorbance values were corrected by medium fluorescence and absorbance. Additionally, the OD_{600} was recorded to normalize fluorescence by optical density.

2.9 Growth assay in organo-autotrophic conditions in microbioreactor (BioLector Pro®)

Single colonies of *C. necator* strains of choice were inoculated in sterile falcon tubes containing either 4 mL of 0.4% (w/v) SG minimal medium with 15µg/mL tetracycline (only if harbouring expression vectors) or 4 mL of 80 mM Na-formate minimal medium and incubated overnight at 30 °C, with shaking at 200 rpm. The OD_{600} of the cultures was measured in the morning and cultures were washed twice with 3mL-5mL of PBS depending on the size of the pellet. Then, the cultures were normalised to an OD_{600} of around 0.15–0.25 in 1 mL of 80 mM Na-formate minimal medium and inoculated in a microtiter plate (MTP-48-BOH 1) compatible with the BioLector Pro® microbioreactor (m2p Labs, Germany). Scattering light was recorded every 30 min for 48 to 72 h to assess the growth of the bacterial strains. Control of

moisture was set to 85% and the anti-evaporation seals (m2p Labs) were used to reduce the evaporation of samples during the run. The oxygen levels were set at 21% for the best oxygen transfer within the MTP flower plate and the CO_2 at 10%.

2.10 Growth assay in autotrophic conditions in serum bottles

The serum bottle experiment was carried out in autotrophic conditions (H_2 : CO_2 : Air 60 : 20 : 20 v/v/v %) with a working volume of 25 mL. The minimal medium (MM) pH 7 used for the experiment was filtered sterilised and moved into the bottles after autoclaving. The serum bottles were sealed with butyl rubber stoppers and crimp-sealed prior to gas filling. The headspace was subsequently flushed with a 60% H_2, 20% CO_2 gas and 20% air. The seed cultures for each strain were prepared from fresh plates (0.4% SG MM gentamicin 10 μg/mL) scooping up a loop of culture into a sterile 250 mL Duran baffled shake flask containing 20 mL of 0.4% (w/v) SG minimal medium gentamicin (10 μg/mL), with tetracycline (15 μg/mL; only if harbouring expression vectors) and incubated overnight at 30 °C, with shaking at 200 rpm. The serum bottles were inoculated with the seed culture, previously washed in PBS to get rid of residual carbon sources, at a final OD_{600} of 0.2. Samples were taken regularly every 24 h for monitoring biomass growth.

2.11 Fermentation in autotrophic and mixotrophic conditions in DASGIP® Parallel Bioreactor System

The fed-batch fermentation in autotrophic conditions (H_2 : CO_2 : Air 78 : 3 : 19 v/v/v %) was carried out in a 1.3 L DASGIP® Parallel Bioreactor System (Eppendorf, Germany) with a working volume of 0.75 L. The pH two-point calibration was carried out before the vessel sterilisation. The minimal medium (MM) used for the fermentation was filtered into the vessel after autoclaving, followed by the

calibration of the DO probe at 0%, 10% and 21% using nitrogen, nitrogen with air and pure air, respectively. All the gasses (H_2, CO_2 and Air) were supplied during the fermentation at a total flow rate of 1 vvm with a percentage composition of 3%, 19% and 78% (v/v/v), respectively. In particular, the hydrogen was supplied through an attached Red-y-mass flow controller (Icenta, UK). A pH of 7 was maintained throughout the experiment using HCl 1.5 M and potassium hydroxide 5 M as acid and base buffers. A cascaded agitation between 800 and 1600 rpm was used to keep the DO at a minimum of 5%. The primary cultures for each strain were prepared from fresh plates (0.4% SG MM gentamicin 10 μg/mL) scooping up a loop of culture into 5 mL of LB gentamicin (10 μg/mL) medium in a sterile falcon tube and grown overnight at 30°C, 200 rpm. A secondary culture was prepared with dilution 1:50 from the first seed culture, in a total volume of 100 mL of LB gentamicin (10 μg/mL) medium in a 250 mL Duran shake flask and grown overnight at 30°C, 200 rpm. The bioreactor was inoculated with the secondary culture, previously washed with PBS, to a final OD_{600} of 0.2 in a working volume of 0.75 L. Off-gases were collected in a separate vessel with allocated DO sensor to monitor the oxygen concentration in the bioreactor headspace to ensure it was kept below the limiting oxygen concentration (LOC) for hydrogen explosion of 4 %. Subsequently, they were split and redirected to a RAMAN spectrometer (Raman Laser Gas Analyser RLGA, Atmosphere Recovery Inc, Minnesota, USA) for an accurate percentage measurement of the off-gasses composition. Samples were taken regularly for each bioreactor for monitoring growth, PHA and VFA concentration.

Fermentation in mixotrophic conditions was carried out with a similar bioreactor set up as described above, with addition of 10 mM VFAs mix at the start of the fermentation process. The mixture of VFAs contained an equal concentration of C1 to C6 fatty acids

namely: formic acid, acetic acid, propionic acid, butyric acid, valeric acid and hexanoic acid.

In case of the mixotrophic fermentation for the co-consumption experiment, the bioreactor set up was kept the same. Although, the cultures were grown in auto-trophic conditions until injection in exponential phase at 24 h with a known concentration of VFAs mix (5 mM). VFAs concentration in the supernatant was analysed using liquid chromatography.

2.12 HPLC sample preparation and analysis

Culture samples were centrifuged at 14,000 rpm for 2 min and the supernatants used for product analysis. Also, an appropriate linear range of standard solutions using 3-HP (Apollo Scientific) covering the range of concentrations expected was prepared in the medium used for the corresponding experiment. All standards, samples and blanks were passed through 0.2 µm syringe filters, and 150 µL of the filtered solution mixed with 150 µL of diluent solution, vortexed thoroughly and added into HPLC vials containing inserts and closed with split caps. Samples were run at a flow rate of 0.5 mL min^{-1} at 35°C in 0.005 M H_2SO_4 mobile phase for 30 min on the Dionex UltiMate 3000 HPLC system (ThermoFisher Scientific) using a Rezex ROA-Organic Acid H + (8%) 150 mm × 4.6 mm × 8 µm column (Phenomenex) and a Diode Array detector (UV–VIS 210 nm). An Aminex HPX-87H 300 mm × 7.8 mm × 9 µm column (Bio-Rad) was used when a better separation of peaks was needed, and samples were run for 55 min.

2.13 PHA quantification using GC-MS

The content of intracellular PHA was determined by GC-MS using previously described method (Lo *et al.*, 2009; Jung *et al.*, 2019) with modifications. This was done by subjecting approximately 25 mg of lyophilised cells to methanolysis in Teflon capped glass vials with 2

mL of 5% sulphuric acid, 2 mL chloroform and benzoic acid added as internal standard. The vials were then incubated for 4 hours at 100°C and subsequently cooled to room temperature. 4 mL of 1 M NaCl solution in HPLC grade H_2O (Sigma Aldrich) was added to the vials and the mixture was vortexed for approximately 30 s. After phase separation, 1 mL of the organic phase containing methylated monomers, was transferred to GC vials and injected into a gas chromatograph (GC) (Agilent 6890N with 7673 Autosampler), with split injection (10:1) and a constant temperature of 250 °C. The GC is equipped with a J&W DB-WAX column (Agilent), with hydrogen as the carrier gas at 0.6 mL min^{-1}. In our protocol, the oven was held at 40 °C for 5 minutes, then ramped at 20 °C/min to 240 °C, which was then held for 6 minutes. The GC to mass spectrometer (MS) interface held at 250 °C. The MS has a 5973 MSD EI^+ detector (Agilent), operating in scanning mode from 40 to 500 m/z, with 3 min solvent delay and EMV delta 0. PHA contents were estimated from Standard curves of hydroxybutyrate using commercial P-3HB (Sigma Aldrich, UK). Concentrations of Hydroxyvalerate (HV) was estimated by calculating the peak area relative to HB.

2.14 Bioinformatics

2.14.1 Plasmid map analysis

Geneious 11.1.2 software was used to conduct sequence alignments, restriction sites analysis and design of plasmid maps.

2.14.2 Sequence databases

The databases at the National Centre for Biotechnology Information (NCBI) were used to identify and access nucleotide sequences. The Kyoto Encyclopedia of Genes and Genomes (KEGG) was used for identifying pathways and annotating genes.

2.14.3 BLAST analysis

BLAST tool, available on the Geneious 11.1.2 software, was used to analyse sequences and identify homology. Nucleotide BLAST (blastn) optimised for highly similar sequences (megablast) was used to identify homologous genes based on DNA sequence; blastp (protein-protein blast) was used to identify homologous proteins.

2.14.4 Next generation Sequencing (NGS) analysis

Geneious 11.4 software was used to conduct NGS analysis from Illumina sequencing data sets. The in-built tool was used for alignment of NGS reads, identification and annotation of SNP variant analysis.

2.15 Molecular modelling computational methods

2.15.1 Model systems

Two sets of dimeric (bigger and smaller subunits) models were prepared, using the wild type RubisCO from *C. necator* H16 (*Cn*RubisCO) and *H. naepolitanus* (*Hn*RubisCO). The structures of the wild type enzymes were taken from the RSCB Protein Data Bank (PDB) using the 1BXN (Hansen *et al.*, 1999) and 1SVD (Kerfeld *et al.*, 2004) codes, respectively. The protonation states of the titratable residues were determined at neutral pH 7 utilizing the H++ (Gordon

et al., 2005; Myers *et al.*, 2006; Anandakrishnan *et al.*, 2012) server following the validation by visual inspection of the individual sidechain surroundings. Additionally, a third model was created by mutating the residues to Thr33, Leu35, Leu76, Asp79, Asp92, Gly96 and Asp97 of the wild type *Cn*RubisCO (*CnMut*RubisCO) to modify the surface charge of the enzyme to mimic the negative behaviour of the *Hn*RubisCO.

2.15.2 Electrostatic potential

The electrostatic potential (ESP) was obtained from the semiempirical AM1 calculations performed on the crystal structure of the *Hn*RubisCO protein (PDB: 1svd), the *Cn*RubisCO (PDB: 1bxn) and the created mutant using EMPIRExy software (Hennemann and Clark, 2014). The positive and negative regions of the ESP are shown in 50 to -100 kcal/mol, respectively. The isovalue of the grid is set to 0.001.

2.15.3 Force filed parameters

The ff14SB force field (Maier *et al.*, 2015) was employed to describe standard protein residues. All models were solvated in a truncated octahedron of approximately 15 Å using the sodium chloride (NaCl) salt concentration of 20, 150 and 1000 mM. To achieve the desired concentration of c(NaCl) = 20 mM, 16 sodium and 9 chloride ions were added in the case of the *Cn*RubisCO enzyme, 18 sodium and 7 chloride ions were added for *CnMut*RubisCO enzyme, while 17 cations and 7 anions were added to the system containing the *Hn*RubisCO enzyme. For the physiological c(NaCl) = 150 mM, the number of added cations was 90, 91 and 90, while the number of anions was 82, 80 and 81 in the case of the *Cn*RubisCO, *CnMut*RubisCO and *Hn*RubisCO enzymes, respectively. For the higher c(NaCl) = 1000 mM, the number of added sodium ions was 580, 575 and 579, while the number of chloride ions was 571, 562 and 569 in the case of the *Cn*RubisCO, *CnMut*RubisCO and *Hn*RubisCO enzymes, respectively. We utilized

the SLTCAP software to calculate the number of ions necessary to achieve the desired salt concentration (Schmit *et al.*, 2018).

2.15.4 Molecular dynamic (MD) simulations

First, the 10 000 steps of the steepest descent energy minimization were performed applying the harmonic position restraints (10 kcal/mol) on the protein atoms. The temperature equilibration (NVT) was performed by heating the system during 100 ps to a target temperature of 300 K in 1000 ps using the Langevin thermostat (Farago, 2019) restraints (10 kcal/mol) on the protein atoms with a collision frequency of 2 ps^{-1}. The constant pressure simulations (NPT) of the system was achieved by using a Berendsen barostat (Lin *et al.*, 2017) with an average pressure of 1 atm restraints (5 kcal/mol) on the protein atoms. We run a short NPT with the same condition but without positional restrains on the protein for 800 ps. All simulations were propagated using the 2 fs time step and writing the output trajectory every 5 ps. The molecular dynamics simulation of two independent replica for each condition were carried out for a total length of 550 ns each. The Long-range electrostatic interactions were modelled using the particle-mesh Ewald (PME) method, with a 12 Å cut-off. All bonds involving hydrogen atoms were constrained with the SHAKE algorithm (Miyamoto and Kollman, 1992). All MD simulations were performed using the GPU accelerated *pmemd* module of the AMBER18 software package (Götz *et al.*, 2012; Le Grand *et al.*, 2013; Salomon-Ferrer *et al.*, 2013)

2.15.5 Analysis

The postprocessing of the trajectories including the root mean squared deviation (RMSD), the root mean square fluctuation (RMSF) and the calculation of the radial distribution function (RDF) and the average density of the ions around the solute was carried out using the *cpptraj* analysis tool of the AMBER18 package(Götz *et al.*, 2012;

Le Grand *et al.*, 2013; Salomon-Ferrer *et al.*, 2013). All structures were visualized using VMD 1.9.3 (Humphrey *et al.*, 1996).

Chapter 3 - Characterisation of the *Halothiobacillus neapolitanus cso* operon in *Cupriavidus necator* H16

The model organism *C. necator* H16 is able to produce biodegradable bioplastics by growing on CO_2 and H_2 as the sole carbon and energy sources. CO_2 assimilation is mediated by the endogenous Calvin-Benson-Bassham (CBB) cycle. However, this is known for being highly energy-demanding and inefficient, mostly because of O_2 competition with CO_2 for the RubisCO catalytic site. As an attempt to improve CO_2 assimilation efficiency, the *cso* operon from *H. neapolitanus* was introduced in *C. necator* H16. This operon encodes self-assembling, proteinaceous prokaryotic organelles (α-carboxysomes) involved in the carbon concentrating mechanism (CCM).

In this chapter, synthetic operons harbouring the *H. neapolitanus cso* genes under the control of either an inducible or a constitutive promoter were constructed. In addition, these were introduced into a RubisCO null mutant of *C. necator* H16 to test whether the heterologous expression of α-carboxysomes was sufficient to restore this strain's ability to grow under autotrophic conditions. Finally, the effects of introducing the RubisCO activase-encoding genes *cbbQ* and *cbbO* from *H. neapolitanus* were also investigated.

3.1 Introduction

Membrane-bound organelles for compartmentalization of specific functions are a key feature of all eukaryotic cells. Bacteria also possess similar organelles, however, these are not conserved throughout the domain. They are, instead, irregularly distributed amongst species and frequently provide key evolutionary advantages for the successful colonization of specific niches (Gonzalez-Esquer *et al.*, 2016). More than 50 years ago, polyhedral bodies composed of only proteins were discovered in electron micrographs in members of the Cyanobacteria (Drew and Niklowitz, 1956) and, later on, in chemoautotrophs (Shively *et al.*, 1970). These structures were subsequently identified as proteinaceous shells encapsulating enzymes required for carbon fixation and, for this reason, they were named carboxysomes (CBs) (Shively *et al.*, 1973). Through sequence comparison of the encapsulating shell proteins from different microbial species, it was shown that carboxysomes are part of a family of related organelles, known as bacterial microcompartments (BMCs) (Axen *et al.*, 2014). Despite the functional diversity of organelles belonging to the BMC family, two of its members, the α and β carboxysomes, are both known to play a role in CO_2-fixation and autotrophic metabolism, by participating in the Calvin-Benson-Bassham cycle (Kerfeld and Melnicki, 2016).

The carboxysome is the core component of what is known as CO_2–concentrating mechanism (CCM) (Badger and Price, 1992; Kaplan and Reinhold, 1999; Price *et al.*, 2008), which works together with active CO_2 and HCO_3^- uptake transporters to accumulate inorganic carbon, primarily as HCO_3^-, in the cytoplasm of the cell (Figure 1.9 B) (Price *et al.*, 1998; Dobrinski *et al.*, 2005). HCO_3^- then diffuses through the proteinaceous shell layer and accumulates in the lumen, where carbonic anhydrase (CA) enzymes convert it to CO_2 (Figure 1.9). Subsequently, CO_2 is fixed by RubisCO into a three-carbon molecule

(3-phosphoglycerate; 3-PGA) that can freely diffuse through the proteinaceous barrier and enter the CBB cycle.

The putative mechanism for efficient carboxysomal C_i fixation involves the sequestration of RubisCO and CAs within a selectively permeable protein shell layer (Kerfeld and Erbilgin, 2015). The molecular mechanism by which CBs enhance the catalytic performance of RubisCO is dependent on the co-localisation of this enzyme with CA activity. Indeed, the CA localisation model has shown that CA nuclei are present throughout the shell structure (Cot *et al.*, 2008) and that these are surrounded by RubisCO zones (Rae *et al.*, 2013). Moreover, it has been experimentally proven that α-carboxysome shells provide CO_2 diffusion resistance (Reinhold *et al.*, 1991; Cai *et al.*, 2009), preventing leakage of CO_2 and oxygen influx, while allowing the transit of HCO_3^-, Mg^{2+}, ribulose-1,5-bisphosphate (RuBP) and 3-PGA. Sequestration of this rather inefficient enzyme into a microcompartment allows carboxysomes to function as CO_2 traps that provide RubisCO with a substrate-enriched environment, thus enhancing C_i fixation rates.

Because of the advantages offered by BMCs, which include spatial segregation from central metabolism and enhancement of inefficient enzyme activity, these structures have recently started to gain interest as possible tools for improving carbon fixation. In this regard, the heterologous expression of α- and β-carboxysomes has already been described in the model organism *E. coli* (Bonacci *et al.*, 2012; Cai *et al.*, 2016). The ability of α-carboxysomes from *H. neapolitanus* to incorporate chimeric and heterologous RubisCO species has also been demonstrated (Menon *et al.*, 2008). Even though the heterologous expression of CBs has so far been attempted mostly in *E. coli*, a successful expression of *H. neapolitanus cso* operon in *Corynebacterium glutamicum,* a Gram-positive model species, has been reported recently (Baumgart *et al.*, 2017). These findings

support the overall modularity of the α-carboxysomes system and the feasibility of its heterologous expression. However, more work needs to be done to optimise carboxysomes functionality in different host organisms.

Further research has suggested the existence of putative accessory genes that may belong to the recently identified carboxysomes superloci (Axen *et al.*, 2014). They are located downstream of RubisCO genes in several operons (Badger and Bek, 2008b) and have been shown to code for proteins involved in RubisCO activation (Wheatley *et al.*, 2014; Turmo *et al.*, 2017). CsoCbbQ was the first of these accessory proteins to be shown to interact with α-carboxysome shells. It forms an hexameric ring and contains an AAA+ domain with ATPase activity, which is essential for the release of inhibitory molecules, such as RuBP, from the RubisCO active site (Sutter *et al.*, 2015). CsoCbbQ was also shown to interact with CsoCbbO (Tsai *et al.*, 2015). This protein contains a von Willebrand Factor A (VWA) domain, which is well known for mediating protein-protein interaction via a carboxylate side-chain (Xiong *et al.*, 2002; Santelli *et al.*, 2004). This domain appears to be essential for the functionality of the CsoCbbQCbbO complex, since it has been suggested to promote the interaction between RubisCO and CsoCbbQ (Sutter *et al.*, 2015; Tsai *et al.*, 2015).

In the past decade, several studies have highlighted the potential of *Cupriavidus necator* strain H16 (formerly *Ralstonia eutropha* or *Alcaligenes eutrophus*) as a model organism for facultative chemo-lithoautotrophic bacterial species and as a chassis for biotechnological applications, particularly for the production of bioplastics (Abascal *et al.*, 2005; Brigham *et al.*, 2012; Grousseau, Lu, *et al.*, 2014; Chen *et al.*, 2015). Sequencing of the *C. necator* genome suggested a high degree of flexibility in its carbon and energy metabolism (Pohlmann *et al.*, 2006; Cramm, 2009). Its potential for scalability at industrial

level and genetic amenability make of *C. necator* an attractive host for the production of valuable bioproducts, such as bioplastics and biofuels (Shimizu *et al.*, 2015; Torella *et al.*, 2015).

The aim of this study was to express the *cso* operon from *H. neapolitanus*, by its own and in combination with the *cbbQO* genes, encoding for RubisCO activases, in *C. necator* and to test CBs expression, formation and *in vivo* activity in this bacterial species. Overall, the data presented here contribute to further extend our understanding and knowledge of the essential protein components of CBs and paves the way for future applications of these structures as bioengineering tools for unlocking the potential of *C. necator* to be used as a microbial cell factory at industrial scale.

3.2 Results

3.2.1 Heterologous expression and formation of carboxysomes in *E. coli* and *C. necator* H16

In *H. neapolitanus,* all the genes known, to date, to be essential for the α-carboxysome assembly are clustered in one genomic locus (Heinhorst *et al.*, 2006; Bonacci *et al.*, 2012), known as the *cso* operon (Cannon *et al.*, 2003). Previous studies have shown that expression of this operon resulted in the assembly of CBs in heterologous hosts (Bonacci *et al.*, 2012; Baumgart *et al.*, 2017). Therefore, to establish if active CBs can form in *C. necator* H16 the *cso* operon from *H. neapolitanus* was cloned into the modular vector pMTL71301, under the control of either an inducible (P_{BAD}) or a constitutive (P_{phaC}) promoter and the resulting plasmids were transformed in *C. necator* H16 (section 2.5.3.2). Expression of the *cso* genes and assembly of the CB structures was also tested in *E. coli* as a positive control, since the heterologous expression of these genes in *E.coli* has been reported in the literature (Bonacci *et al.*, 2012).

3.2.1.1 Expression of *cso* operon in *E. coli*

To test expression of the *H. neapolitanus cso*-operon in *E. coli* DH5α, using the P_{BAD} inducible promoter, culture samples were collected after 6, 12 and 18 h post-induction with 10 mM L-arabinose. SDS-PAGE analysis of the corresponding protein samples showed that all the major components of the α-carboxysome (CbbL, CbbS, CsoS2 and CsoS1ACB) are produced after 6 h of induction and can be detected in the membrane fraction (Figure 3.1). Expression of the *cso* operon was also tested in *E.coli* DH5α using P_{phaC}, a constitutive promoter from *C. necator*. Protein samples were collected after 6, 12 and 24 h from re-inoculation of overnight cultures in fresh culture medium (Figure 3.2). Similarly, to the results seen with the inducible promoter, production of the main components of the carboxysomes was observed after 6 h in the membrane fraction, with no visible

increment in expression levels over time. The addition of RubisCO activases to the *cso* operon, as seen in Figure 3.2 will be discussed in section 3.2.3.1.

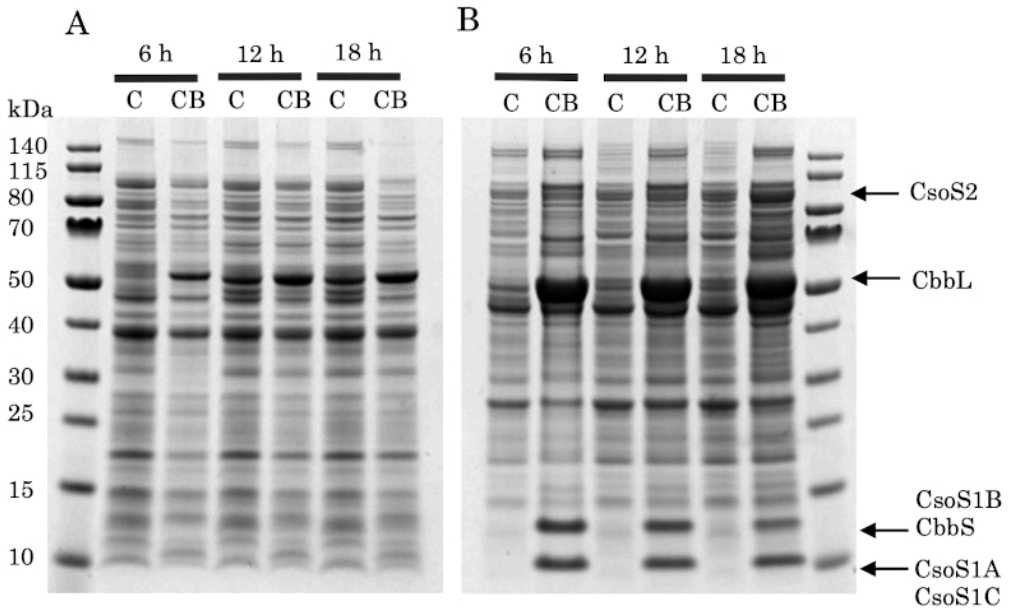


Figure 3.1: Expression of *cso* operon under inducible promoter in *E. coli*. SDS-PAGE analysis of soluble (A) and membrane fractions (B) of the *E. coli* DH5α strain harbouring the plasmid for inducible heterologous expression of *H. naepolitanous* α-carboxysomes (10 mM L-arabinose). CB = *E. coli*/pMTL71301::P_{BAD} *Hn*CB; C = control *E. coli* /pMTL71301::P_{BAD}.

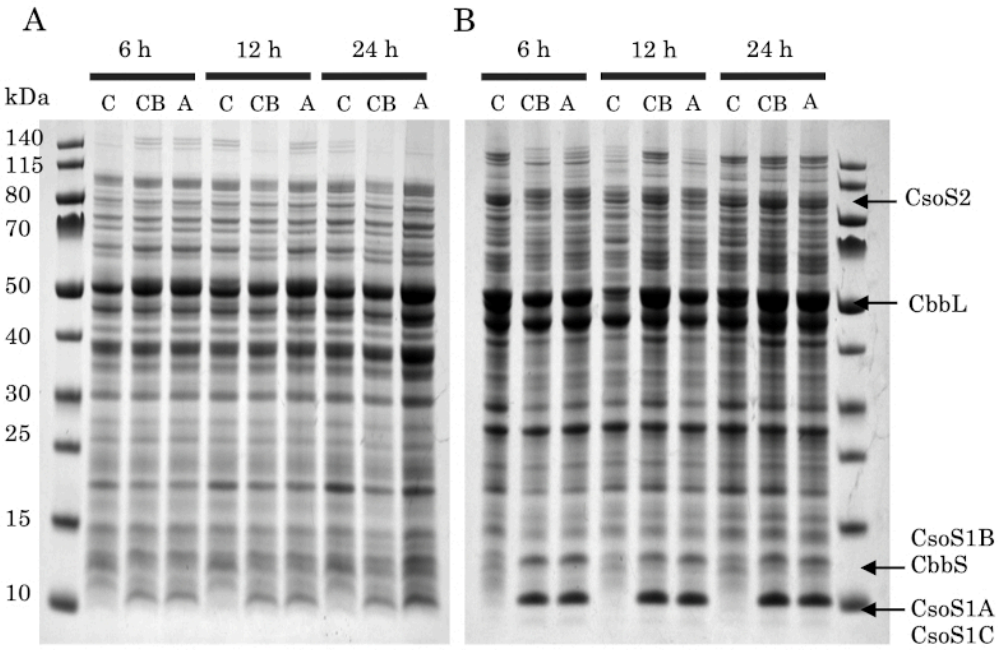


Figure 3.2: Expression of *cso* operon under constitutive promoter in *E. coli*. SDS-PAGE analysis soluble (A) and membrane fraction (B) of *E. coli* DH5α strain harbouring the plasmid for constitutive heterologous expression of *H. naepolitanous* α-carboxysomes, with and without RubisCO activases. Levels of expression after 6, 12 and 24 h. CB = *E.coli*/pMTL71301::P_{phaC} *Hn*CB; A = *E.coli*/pMTL71301::P_{phaC} *Hn*CB Act; C= control *E.coli*/pMTL71301::P_{phaC}-eYFP.

3.2.1.2 Expression of *cso* operon in *C. necator* H16

Heterologous expression of *H. neapolitanus cso* genes was also investigated in *C. necator* H16, using the P_{BAD} arabinose inducible promoter, in the presence of different L-arabinose concentrations (2.5, 5, 10 mM L-arabinose). As in the experiments carried out with *E. coli*, culture samples were collected after 6, 12 and 18 h of induction. The main components of the carboxysome proteins (namely, CbbL, CbbS and CsoS1ACB) could be detected in *C. necator* 12 and 18 h after induction and expression levels increased in an L-arabinose dose-dependent manner. This was particularly evident for the CsoS1A/C shell proteins, 18 h after induction (Figure 3.3). On the other hand, no protein production was detected after 6 h of induction (data not shown). Similar, to what it was observed in *E.* coli, carboxysomal proteins were found mostly in the membrane fraction of the samples. The recovery of the carboxysome components in the membrane fraction suggests their correct assembly. In fact, being them large structures in their closed conformation, they would most likely be retained in the insoluble membrane fraction of the protein samples. The same experiment was repeated in *C. necator* Δ*phaCAB* genomic background to prove consistency between *C. necator* strains expression ability (Appendix 1).

The heterologous expression of caboxysomal proteins was tested in *C. necator* H16 also using a weak constitutive promoter (P_{phaC}). Samples were taken after 6, 12 and 24 h post re-inoculation of overnight cultures in fresh medium. Under these experimental conditions, expression could already be detected at the 6 h time point and further increased after 12 and 24 h of growth (Figure 3.4). Consistently with previous data of expression under inducible promoter, the recovery of the carboxysomal proteins was observed in the membrane fraction, suggesting the correct assembly of the carboxysome shell proteins. These gene expression patterns were consistent with previously published work in *E. coli* (Bonacci *et al.*, 2012) and *Corynebacterium*

(Baumgart *et al.*, 2017). Moreover, the protein abundance shown in the intensity of the bands was consistent with previous data collected in *H. neapolitanus* (Cai *et al.*, 2008) during *cso* operon transcript analysis. The identity of the proteins found in the SDS-PAGE bands was confirmed using MALDI-TOF MS protein identification. The band at 10 kDa was a mixed population of CsoS1A and C, the band at 15 kDa was a mixed population of CsoS1B and CbbS, the band at 55 kDa was recognized as RubisCO large subunit and the band at 100 kDa was identified as CsoS2B. While, the presence of CsoS3 (CA) couldn't be unambiguously detected in the gel, similarly to other studies in *E.coli* (Bonacci *et al.*, 2012; Liu *et al.*, 2018). Identical band patterns were obtained with the *C.necator* Δ*pha*CAB strain (Appendix 2).

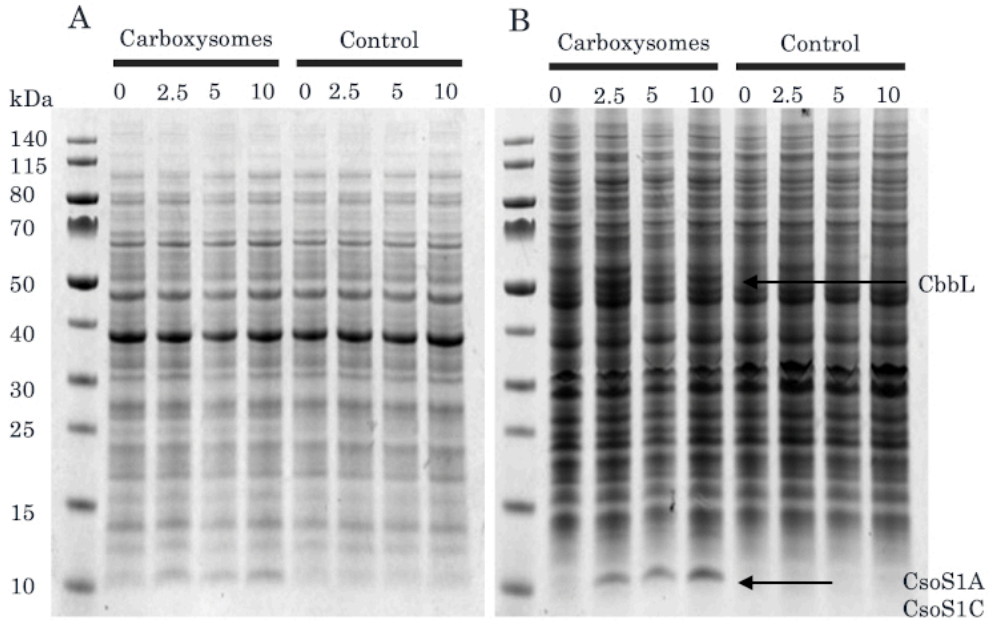


Figure 3.3: Expression of *cso* operon under inducible promoter in *C. necator*. SDS-PAGE of soluble (A) and membrane fractions (B) of the *C. necator* H16 strain harbouring the plasmid for inducible heterologous expression of *H. naepolitanous* α-carboxysomes. Induction levels 0, 2.5, 5, 10 mM L-arabinose. 18 h of induction. Carboxysomes = *C. necator*/pMTL71301::P_{BAD} *Hn*CB; Control = *C. necator*/pMTL71301::P_{BAD}.

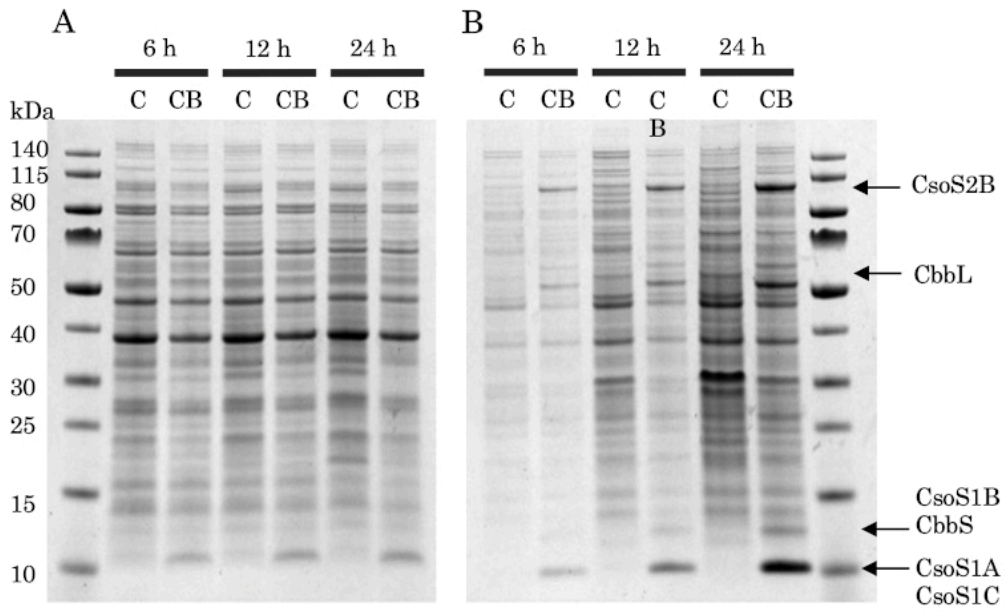


Figure 3.4: Expression of *cso* operon under constitutive promoter in *C. necator*. SDS-PAGE soluble (A) and membrane fractions (B) of the *C. necator* H16 strain harbouring the plasmid for constitutive heterologous expression of *H. naepolitanous* α-carboxysomes. Levels of expression after 6, 12 and 24 h. CB = *C. necator*/pMTL71301::P_{phaC} *Hn*CB; C= control *C. necator*/pMTL71301::P_{phaC}-eYFP.

3.2.1.3 Imaging of *E. coli* strains expressing genes for carboxysomes using transmission electron microscopy (TEM)

The correct structural assembly of the α-carboxysome shell protein in the heterologous hosts *E. coli* and *C. necator* H16 was investigated using whole cell transmission electron microscopy (TEM). As previously described in the literature (Bonacci *et al.*, 2012), we observed the presence of correctly assembled shell proteins in *E. coli* DH5α using an inducible promoter and, for the first time, with a constitutive promoter (P_{phaC}). In our experiments, CB structures were detected after 18 h of induction with 10 mM L-arabinose and after 12 and 24 h after culture re-inoculation, in the case of the constitutive promoter (Figure 3.5). The icosahedral shape of the structures suggests that carboxysomes folded and assembled correctly in *E. coli*. The appearance of filamentous structures at 12 and 24 h in constitutive expressing strains, was consistent with previous data obtained in *H. naepolitanus* (Iancu *et al.*, 2010), where parallel bundles of filaments were detected in between carboxysomes close to the cell membrane. It was hypothesized by the authors that the structures could likely be misassembled shell proteins. Using TEM imaging, it was also possible to measure the diameter of the micro-compartments, which showed an average size of 100 nm, consistent with the previous reports in the literature (Bonacci *et al.*, 2012; Baumgart *et al.*, 2017).

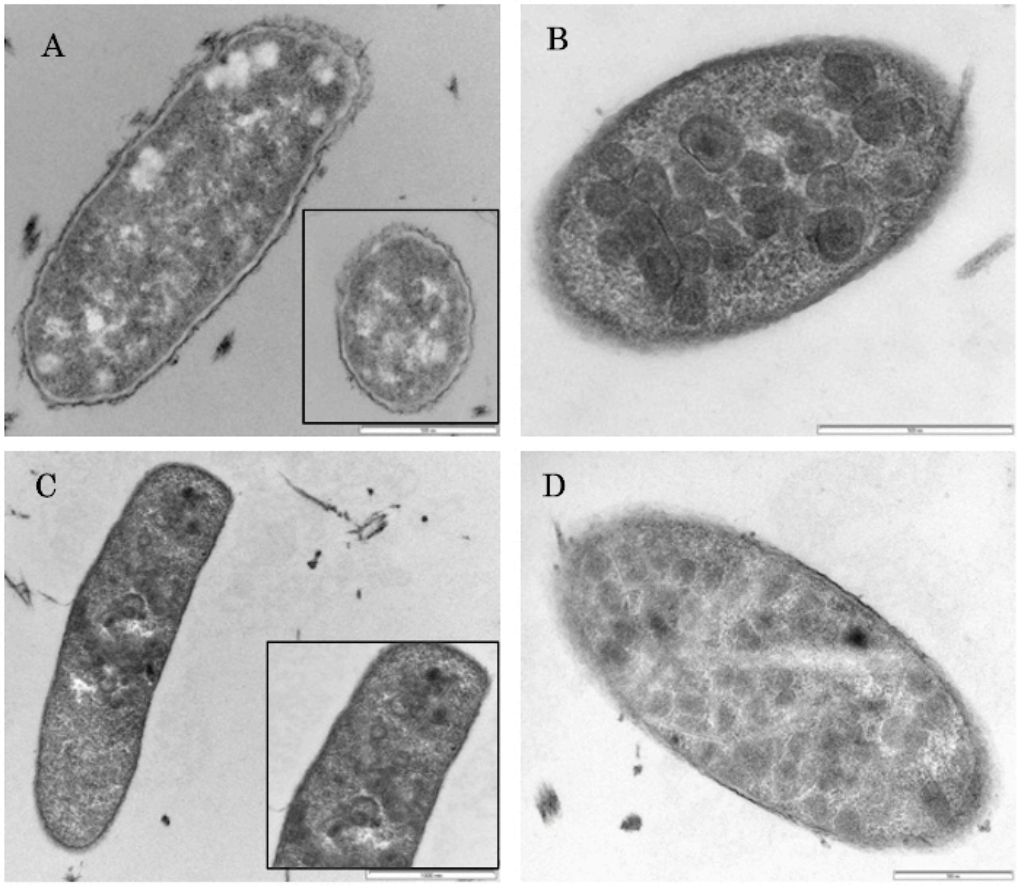


Figure 3.5: TEM imaging of *E. coli* strains expressing *cso* operon under inducible and constitutive promoter. TEM images of carboxysomes *in vivo* in *E. coli* DH5α (A) *E. coli*/pMTL71301::P_{BAD} (Inset) transversal section; (B) Transversal section of *E. coli*/pMTL71301::P_{BAD} *Hn*CB at 12 h after 10 mM L-arabinose induction; (C) *E. coli*/pMTL71301::P_{phaC} *Hn*CB at 12 h after setback (Inset) Icosahedral structures (Scale bar: 1000 nm); (D) *E. coli*/pMTL71301::P_{phaC} *Hn*CB at 24 h after setback. Scale bars: 500 nm.

3.2.1.4 Imaging of *C. necator* strains expressing genes for carboxysomes using transmission electron microscopy (TEM)

TEM imaging of *C. necator* was only attempted with the Δ*pha*CAB deletion strain, which is unable to produce PHA granules, in order to prevent the possible interference that these intracellular structures may have caused during the imaging of the carboxysome shells. An optimised sample preparation method was developed in collaboration with the nmRC center (University of Nottingham), necessary due to the mottled appearance of the cytosolic compartment in *C. necator*. The heterologous expression of carboxysomes using the P_{BAD} arabinose-inducible promoter was observed after 12 h of induction with 2.5 mM and 5 mM of L-arabinose. Interestingly, due to the leakiness of the P_{BAD} inducible promoter a singular carboxysome per cell could be detected even in the non-induced *C. necator* H16 cultures (Figure 3.6). As expected, the quantity of the carboxysomes was dependent on L-arabinose concentration. Indeed, the number of fully formed carboxysomes in the cytoplasm of cells grown in the presence of 5mM L-arabinose was significantly higher, with respect to those supplemented with 2.5 mM of the same inducer. However, high-resolution imaging proved to be very difficult to achieve due to the presence of electron dense areas. Indeed, the production of CBs observed was so high that the entire cytoplasm was filled with the micro-compartments, leading to the formation of inclusion bodies (Figure 3.6 D). Unfortunately, the proportional increase of inclusion bodies with higher arabinose induction levels precluded the opportunity to collect more data on the structural integrity and architecture of the carboxysomes. A similar result was observed following induction with 10 mM L-arabinose (result not shown). Heterologous expression of the *cso* operon-encoded proteins was observed after 12 and 24 h from culture re-inoculation, using the P_{phaC} constitutive promoter (Figure 3.7). Consistently with the data obtained using the P_{BAD} inducible promoter, the number of carboxysomes in the cytoplasm of the imaged cells increased over

time. Even in this case, TEM micrographs showed complete shell structures and correct assembly of the external architecture of the CBs. However, the average CB diameter (74 nm) was slightly smaller in *C. necator*, as compared to that observed in *E. coli* (100 nm). Better resolution of the images was achieved with the constitutive promoter, suggesting that a strong induction (2.5 mM or 5 mM L-arabinose) of the *cso* operon expression may be detrimental for the folding of CB structures, thereby resulting in inclusion bodies formation. On the other hand, a lower but constitutive expression of these genes may allow sufficient time for the CB micro-compartments to self-assemble, without causing an excessive metabolic burden on *C. necator* cells.

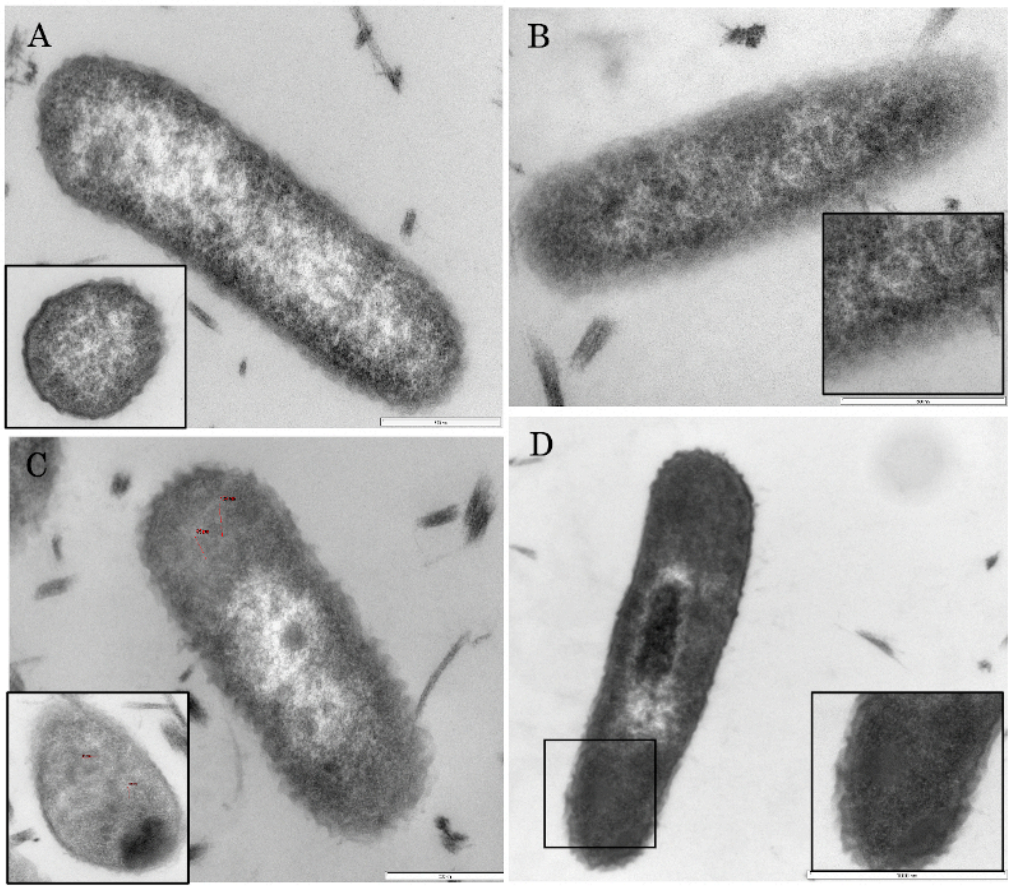


Figure 3.6: TEM imaging of *C. necator* strains expressing *cso* operon under inducible promoter. TEM images of carboxysomes *in vivo* in (A) *C. necator* Δ*phaCAB* /pMTL71301::P_{BAD} (Inset) transversal section; (B) *C. necator*/pMTL71301::P_{BAD} *Hn*CB after 12 h 0 mM L-arabinose induction (Inset) icosahedral structure . (C) Same as B at 2.5 mM L-arabinose (Inset) transversal section with icosahedral structures and inclusion bodies. (D) Same as C, at 5 mM highly electron dense areas (Inset) inclusion bodies. Scale bars: 500 nm.

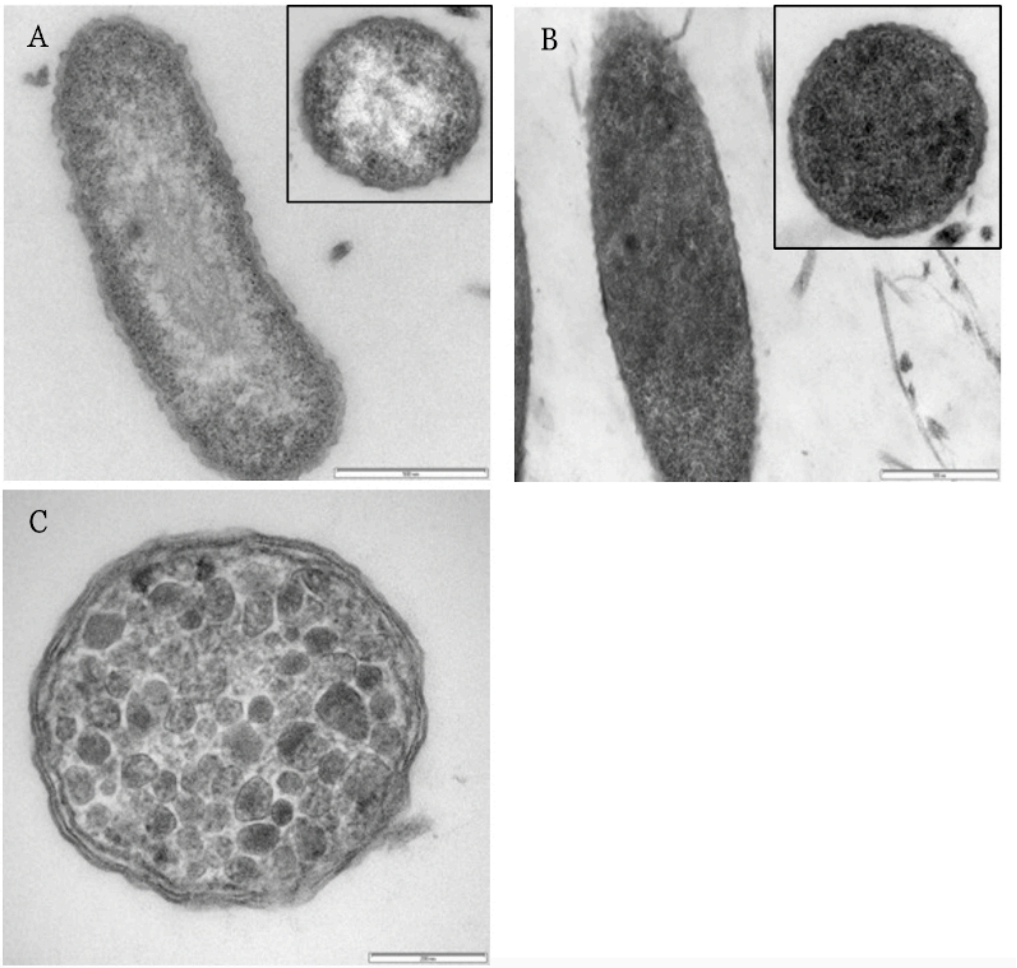


Figure 3.7: TEM imaging of *C. necator* strains expressing *cso* operon under constitutive promoter. TEM images of carboxysomes *in vivo* in (A) *C. necator* Δ*phaCAB*/ pMTL71301::P_{phaC}-eYFP (Inset) transversal section; (B) *C. necator*/pMTL71301::P_{phaC} *Hn*CB after 12 h from set-back (Inset) transversal section icosahedral structures; Scale bars: 500 nm; (C) Same as B transversal section icosahedral structures after 24 h setback Scale bar: 200 nm.

3.2.2 Functional characterisation of carboxysomes in *C. necator* strains

Expression and correct folding of the RubisCO from *H. neapolitanus* (*Hn*RubisCO) represents a major obstacle to the production of functional carboxysomes in a heterologous host. Recent work on carboxysome superloci highlighted the importance of accessory proteins (Axen *et al.*, 2014), which can help protein folding, protein activity and assembling of these structures. In particular, one conserved PCD-like protein, which is encoded by a gene located just downstream of the *cso* operon in *H. neapolitanus* and was renamed α-carboxysome RuBisCO assembly factor (or acRAF), showed the ability to assist *H. neapolitanus* RubisCO folding, in co-operation with the ubiquitous GroEL/ES chaperones in *E.coli* (Wheatley *et al.*, 2014). For this reason, the ability of the *H. neapolitanus* RubisCO to assemble and function in the *C. necator* cytoplasm was investigated in the absence of acRAF.

In order to test the *in vivo* activity of the heterologous *H. neapolitanus* RubisCO in *C. necator*, a *C.necator* mutant, lacking both copies of the RubisCO-encoding genes (ΔΔ*cbbLS)*, which is therefore unable to fix CO_2, was constructed. To assess the ability of the native *C. necator* RubisCO-encoding gene and of that derived from *H. neapolitanus* to complement the ΔΔ*cbbLS* mutant, each of these *cbbLS* gene pairs were cloned in the pMTL71301 modular vector, under the control of the constitutive promoter P_{phaC}, and transformed in *C. necator* ΔΔ*cbbLS*. The resulting strains were then grown in organo-autotrophic conditions (80mM Na-formate MM, in the presence of 10% CO_2) in the BioLector Pro microbioreactor for 48 h. In these experiments, *C. necator* H16 wild type and its isogenic ΔΔ*cbbLS* mutant were used as the positive and negative control, respectively. Surprisingly, we observed that the RubisCO from *H. neapolitanus* was able to restore the growth of the *C. necator* ΔΔ*cbbLS* strain after 4 h of growth under organo-autotrophic conditions (Figure 3.8), with

a doubling time of 11.5 h, compared to the 11 h of the *C. necator* H16 wild type strain. This observation was remarkable, especially considering that the *C. necator* ΔΔ*cbbLS* strain expressing a plasmid-borne copy of the native *C. necator* RubisCO grew slightly slower than the strain expressing *Hn*RubisCo, with a doubling time of 12.9 h. However, at the onset of stationary phase (20 h) we recorded no statistical difference in biomass between the *C. necator* RubisCO deletion mutant strains carrying the *H. naepolitanus* and *C.necator* RubisCO, respectively. Altogether, these data strongly suggest that the ability of *H. neapolitanus* RubisCO to fold and function in *C. necator* is independent of acRAF activity.

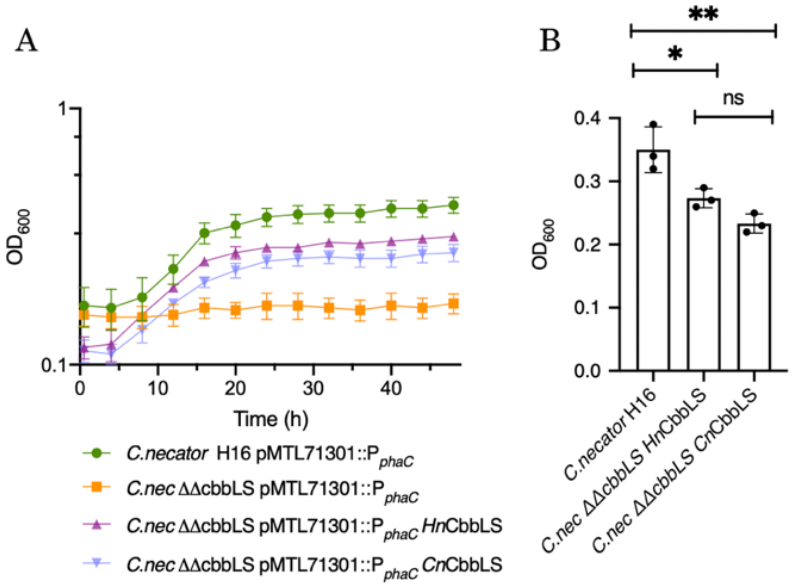


Figure 3.8: Functional complementation of *Hn* and *Cn*RubisCO in *C. necator* ΔΔ*cbbLS* deletion strain. Complementation of the *C. necator* ΔΔ*cbbLS* deletion strain with constitutively expressed CbbLS from either *H. naepolitanus* or *C. necator* (P_{phaC} *Hn*CbbLS and P_{phaC} *Cn*CbbLS, respectively) (B) One-way ANOVA analysis of the biomass produced by *C. necator* H16; *C. necator* ΔΔ*cbbLS*/ pMTL71301::P_{phaC} *Hn*CbbLS and *C. necator* ΔΔ*cbbLS*/ pMTL71301::P_{phaC} *Cn*CbbLS after 20 h of growth under organo-autotrophic conditions (80mM Na-formate MM, in the presence of 10% CO_2).

Similarly, *Hn*RubisCO and *Cn*RubisCO activity was tested in *C. necator* Δ*cbbLS* also lacking the *phaCAB* genes, which are responsible for PHB biosynthesis. This strain was transformed with the plasmids carrying *Hn*RubisCO and *Cn*RubisCO under the control of the constitutive promoter P_{phaC} and the resulting strains were grown in the conditions described above. The data obtained showed that RubisCO from *H. neapolitanus* was again more efficient than *Cn*RubisCO in complementing the growth defects of the *C. necator* Δ*pha*CAB Δ*cbbLS* strain under autotrophic conditions (Figure 3.9). Remarkably, the *C. necator* Δ*cbbLS* Δ*phaCAB*/ pMTL71301::P_{phaC} *Hn*CbbLS (doubling time of 6.5 h) grew even quicker than the positive control strain *C. necator* Δ*phaCAB* (doubling time of 10 h). This observation strongly advocates the ability of *Hn*RubisCO to fold independently from acRAF. Similarly to what observed with *C. necator* Δ*cbbLS*, the *C. necator* Δ*pha*CAB Δ*cbbLS* strain harbouring a plasmid-borne copy of the native *Cn*RubisCO also grew slightly slower than the strain expressing *Hn*RubisCO, which, with a doubling time of 8 h, showed increased fitness with respect to the control strain (*C. necator* Δ*phaCAB*). Moreover, the total amount of biomass produced by the Δ*pha*CAB Δ*cbbLS* deletion strain harbouring *H. naepolitanus* RubisCO at the onset of stationary phase (28 h) was significantly higher, with respect to that observed with both *C. necator* Δ*pha*CAB and the *C. necator* Δ*pha*CAB Δ*cbbLS* deletion strain expressing *C. necator* RubisCO. Altogether, these data further confirm that *Hn*RubisCO is able to fold and properly function in the cytosol of *C. necator*, even in the absence of carboxysomes.

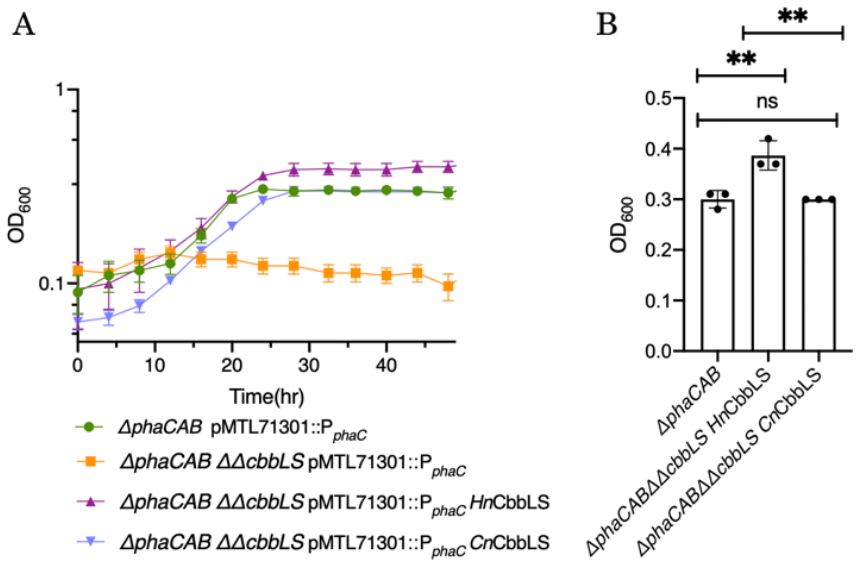


Figure 3.9: Functional complementation of *Hn* and *Cn*RubisCO in *C. necator* Δ*phaCAB* ΔΔ*cbbLS* deletion strain. (A) Complementation of the *C. necator* Δ*phaCAB* ΔΔ*cbbLS* deletion strain with constitutively expressed CbbLS from either *H. naepolitanus* or *C. necator* (P_{phaC} *Hn*CbbLS and P_{phaC} *Cn*CbbLS, respectively) (B) One-way ANOVA analysis of the biomass produced by *C. necator* Δ*phaCAB; C. necator* Δ*phaCAB* ΔΔ*cbbLS*/pMTL71301::P_{phaC} *Hn*CbbLS and *C. necator* Δ*phaCAB* ΔΔ*cbbLS*/pMTL71301::P_{phaC} *Cn*CbbLS after 28 h of growth under organo-autotrophic conditions (80mM Na-formate MM, in the presence of 10% CO_2).

A similar strategy was adopted for testing the *in vivo* activity of the carboxysomes in *C. necator* H16. In order to facilitate a stable expression of the carboxysome genes in *C. necator,* a plasmid addiction system was used. This was achieved by constructing a pantothenate auxotroph, which was obtained by deleting the essential gene *panC*, which encodes the enzyme pantothenate synthetase, from the *C. necator* genome. The *C. necator* Δ*panC* mutant was then transformed with a plasmid harbouring a copy of the wild type *panC* gene, thereby creating a plasmid addiction system which is dependent on the *in-trans* complementation of the pantothenate auxotrophy (Callum McGregor and Marco Garavaglia unpublished). As described above for the RubisCO-encoding genes, the *cso* operon was cloned in the pMTL71301::*panC* vector, under the control of the P_{phaC} promoter and the resulting plasmid was transformed in the *C. necator* ΔΔ*cbbLS* Δ*panC* strain. Also in this case, the resulting strain was grown under organo-autotrophic conditions (80mM Na-formate MM, in the presence of 10% CO_2) in the BioLector Pro microbioreactor, using *C. necator* H16 wild type and its isogenic *C. necator* ΔΔ*cbbLS* as the positive control for growth. However, expression of the *cso* operon was unable to restore the ability of *C. necator* ΔΔ*cbbLS* to grow, under the experimental conditions tested (Figure 3.10 A). Similar results were observed in the Δ*pha*CAB genetic background (Figure 3.10 B).

To further investigate the influence of higher carbon and reduction power availability on strains viability and *in vivo* functional complementation, the *cso* operon was cloned in pMTL71301::panC vector, under the control of either the P_{phaC} promoter or P_{BAD} inducible promoter and conjugated with the auxotrophic ΔΔ*cbbLS* depleted *C. necator* Δ*pha*CAB background strain. The strains were then grown in serum bottles under autotrophic conditions (ratio 60:20:4 of H_2:CO_2:O_2 v/v), using auxotroph *C. necator* Δ*pha*CAB and its isogenic ΔΔ*cbbLS* mutant as the positive and negative control for growth,

respectively. However, also in this condition no restoration of growth was observed over one-week experiment (Figure 3.11).

These observations suggest that, despite the correct assembly of the shell proteins as enclosed structures, carboxysomes do not appear to be functional when expressed in *C. necator*.

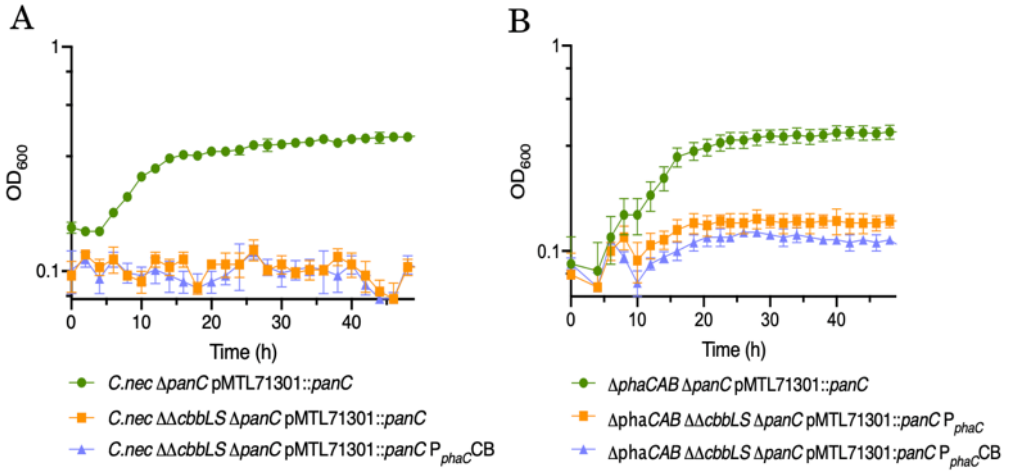


Figure 3.10: Functional complementation of *C. necator* ΔΔ*cbbLS* Δ*panC* and *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* with *cso* operon in organo-autotrophic conditions. Complementation of (A) *C. necator* ΔΔ*cbbLS* Δ*panC* deletion strain and (B) *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* deletion strain, with constitutively expressed caboxysome-encoding genes (P_{phaC} *Hn*CB) on a stable plasmid, under organo-autotrophic conditions (80mM Na-formate MM, in the presence of 10% CO_2) (C) OD_{600} measurements of sterile Na-formate MM over 48 h.

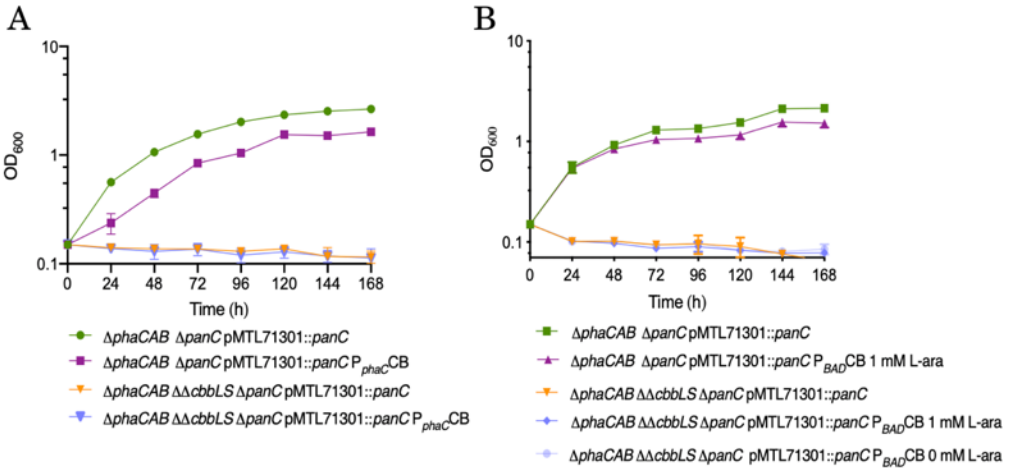


Figure 3.11: Functional complementation of *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* with *cso* operon under constitutive and inducible promoter in autotrophic conditions. Growth curves of *C. necator* Δ*pha*CAB ΔΔ*cbbLS* Δ*panC* deletion strain with (A) constitutively expressed caboxysome-encoding genes on a stable plasmid and (B) a stable plasmid carrying the carboxysome-encoding genes under the control of the P_{BAD} L-arabinose-inducible promoter (P_{phaC} *Hn*CB and P_{BAD} *Hn*CB, respectively) under autotrophic conditions (ratio 60:20:4 of H_2:CO_2:O_2 v/v).

Overall this data might provide additional evidence to support recent theories (Turmo *et al.*, 2017), where a range of accessory proteins may be involved in the correct functionality of the CB micro-compartments, despite being positioned outside of the *cso* operon in the genome. Indeed, even though the data presented here seem to demonstrate that acRAF is not required for the correct folding of *H. neapolitanus* RubisCO when this is expressed in *C. necator* independently of the CBs-encoding genes, there is still the possibility that this accessory protein may play a more crucial role in this process when *Hn*RubisCO is associated with carboxysomal micro-structures.

3.2.3 Heterologous expression and formation of carboxysomes with RubisCO activases in *E. coli* and *C. necator*

As mentioned above, recent reports have suggested the presence of additional essential accessory genes, other than the *cso* operons, which may belong to the recently described carboxysome *superloci* (Axen *et al.*, 2014; Turmo *et al.*, 2017). Some of these genes are located downstream of the RubisCO genes in several operons (Badger and Bek, 2008a) and code for proteins involved in RubisCO activity and activation (Wheatley *et al.*, 2014; Turmo *et al.*, 2017). CsoCbbQ and CsoCbbO, called RubisCO activases, were shown to interact with the shell proteins of α-carboxysomes and with the inactive form of RubisCO (Sutter *et al.*, 2015; Tsai *et al.*, 2015) to allow restoration of the active conformation of this enzyme, by facilitating the dissociation of the inhibitory substrate RuBP from its catalytic site (Tsai *et al.*, 2015). Therefore, it is possible that the inability of the CBs heterologously expressed in *C. necator* to complement the RubisCO null mutant may have been due to the absence of these proteins, which prevented the recycling of active RubisCO within the CB shells. To test this hypothesis, the activase-encoding genes were cloned downstream the *cso* operon, under the regulation of the P_{phaC} promoter. Synthetic RBSs were designed to respect the stoichiometry of the complex $CbbQ_6$:$CbbO_1$ (Tsai *et al.*, 2015). Again, protein expression, correct structural assembly together with *in vivo* activity of the CBs, this time in the presence of the activases, was assessed in the *C. necator* Δ*cbbLS* Δ*panC* and *C. necator* Δ*phaCAB* Δ*cbbLS* Δ*panC* strains.

As highlighted in the literature (Chen *et al.*, 2013; Wheatley *et al.*, 2014), RubisCO's large subunit folding would still be assisted, even in absence of acRAF, by the ubiquitous GroEL/ES ATP-dependent chaperone complex (GroEL: WP_010809963.1, WP_010813046.1 and GroES: WP_010813047.1). Two copies of both the *groEL* and *groES*

genes are present in the *C. necator* H16 genome, all of which are located on chromosome 1. Since the two copies of each gene are 100% identical, these were most likely generated by duplication events.

3.2.3.1 Expression of *cso* operon with RubisCO activases in *E. coli*

The expression of carboxysomes in combination with the RubisCO activases genes was tested in *E.coli,* under the control of the constitutive promoter P_{phaC}. This was to test whether the carboxysome protein expression pattern could still be observed in the presence of the activases. Protein samples were collected after 6, 12 and 24 h from re-inoculation of overnight cultures in fresh culture medium (Figure 3.2). As it can be seen in Figure 3.2 the major components of the carboxysomes were expressed however, no bands could be observed for CbbQ and CbbO (about 32 kDa and 90 kDa, respectively) after 24 h from the set back.

3.2.3.2 Expression of *cso* operon with RubisCO activases in *C. necator*

Expression of heterologous carboxysomal proteins harbouring RubisCO activases was investigated in *C. necator* ΔΔ*cbbLS* Δ*panC* strain, using the constitutive promoter $P_{phaC,}$ to confirm the presence of the carboxysomal expression pattern, as reported in section 3.2.1.2. Samples were collected after 24 and 48 h from re-inoculation of overnight cultures in LB fresh medium. The overnight cultures were tested for *cbbLS* double genomic deletions and plasmid presence. Consistently with previous data, after 24 h the expression of the main components of the carboxysomes were visible on the gel in the membrane fraction for the control sample carrying the *cso* operon without activases. However, the membrane fractions of cultures expressing the *cso* operon with activases showed only one band, corresponding to RubisCO (about 55 kDa). The intensity of this band increased from the 24 h to the 48 h time points (Figure 3.12). Comparing the protein expression patterns of these strains with the positive control, no further bands were detected at the expected

molecular weights for the CbbO and CbbQ proteins (about 32 kDa and 90 kDa, respectively), in either the cytosolic or the membrane fractions. Instead, four unknown bands with sizes of around 15, 40, 50 and 60 kDa and intensities increasing over time were detected in the cytosolic fraction of the *C. necator* strain harbouring the genes coding for the carboxysomes and the activases. None of these bands were observed in either the positive or negative controls. We also observed the intensity of the 55 kDa band, corresponding to RubisCO, increasing over time also in the soluble fraction (Figure 3.12).

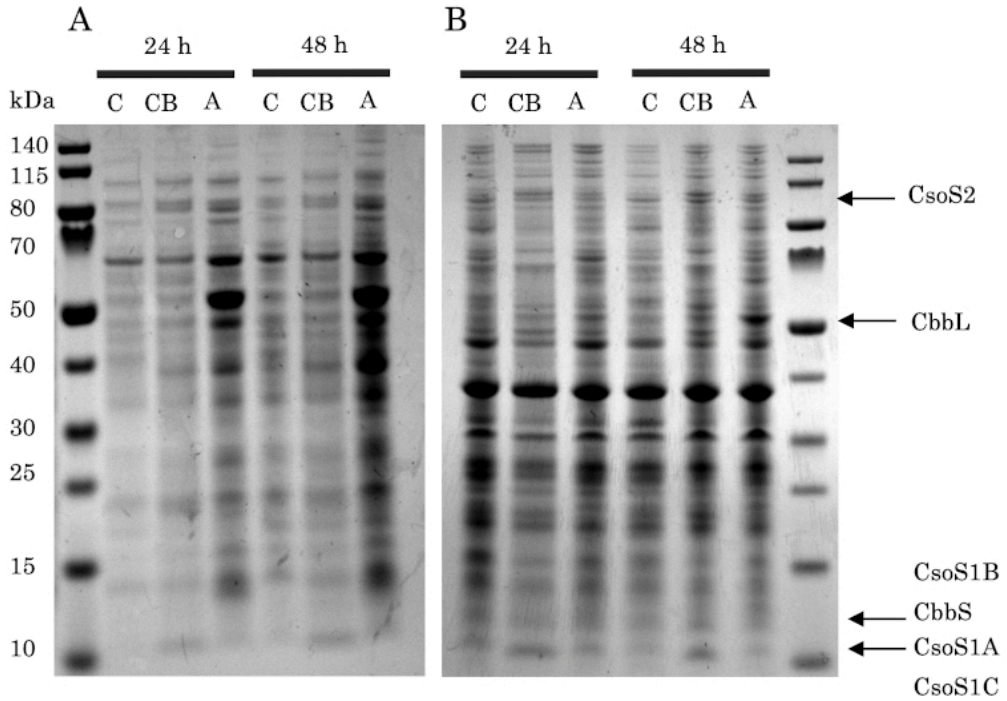


Figure 3.12: Expression of *cso* operon with or without RubisCO activases in *C. necator* ΔΔ*cbbLS* Δ*panC* in LB media. SDS-PAGE of the soluble (A) and membrane (B) fractions of *C. necator* ΔΔ*cbbLS* Δ*panC* harbouring stable plasmids for the constitutive heterologous expression of *H. naepolitanous* α-carboxysomes, with or without RubisCO activases. Protein samples were collected after 24 and 48 h of growth in LB medium. CB = *C.necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} *Hn*CB; A = *C.necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} *Hn*CB Act; C= control *C.necator* Δ*panC* /pMTL71301::*panC*::P_{phaC}.

Two additional growth conditions were tested to assess whether changes in the cultivation media of seed cultures may affect the carboxysome production pattern. Cultures were set up in either MM supplemented with 0.4% SG (heterotrophic conditions) or under organo-autotrophic conditions with 80mM Na-formate and 10% CO_2. Again, the overnight cultures were tested for *cbbLS* double genomic

deletions and plasmid presence. As before, protein samples were collected at 24 and 48 h following re-inoculation in SG MM media. When compared to growth in LB, a reduction in intensity of the bands corresponding to the major carboxysome components, could be observed in the membrane fraction of the *C. necator* strain harbouring the carboxysome-encoding genes, throughout the duration of the experiment. Comparing the protein expression patterns of this strain with that expressing both carboxysomes and activases, a decrease in the expression levels of RubisCO could be observed between the 24 and 48 h time points, in the membrane fraction. Consistent with what was observed during growth in LB medium, no bands associated with CbbO or CbbQ expression could be seen in either the cytosolic or the membrane fractions. Additionally, overexpression of unknown proteins with sizes of about 40, 50 and 60 kDa could be detected in the cytosolic fraction. On the other hand, the RubisCO band at 55 kDa showed a reduction in intensity between the 24 and 48 h time points. However, this may have been an artifact derived from an incorrect normalization of total protein concentration between the two samples (Figure 3.13).

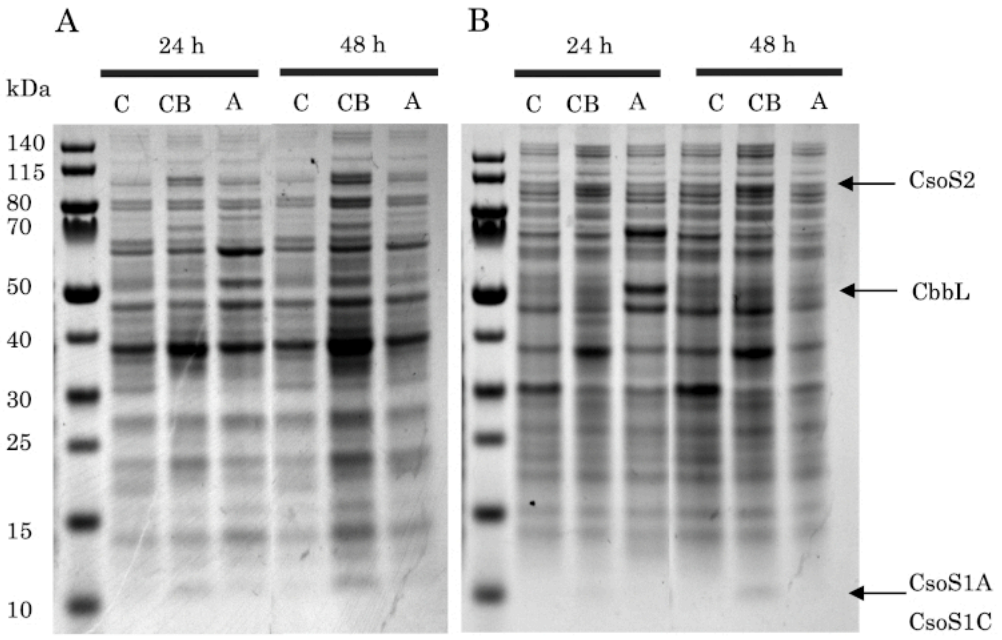


Figure 3.13: Expression of *cso* operon with or without RubisCO activases in *C. necator* ΔΔ*cbbLS* Δ*panC* grown in 0.4% SG MM. SDS-PAGE of the soluble (A) and membrane (B) fractions of *C. necator* ΔΔ*cbbLS* Δ*panC* harbouring stable plasmids for the constitutive heterologous expression of *H. naepolitanous* α-carboxysomes, with or without RubisCO activases. Protein samples were collected after 24 and 48 h of growth in 0.4% SG minimal media. CB = *C.necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} CB; A = *C.necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} *Hn*CB Act; C= control *C.necator* Δ*panC* /pMTL71301::*panC*::P_{phaC}.

In organo-autotrophic conditions, sampling was carried out at 48 and 72 h after re-inoculation, due to the slower growth of *C. necator* cultures in 80mM Na-formate media in the presence of 10% CO_2. The overnight cultures were tested for *cbbLS* double genomic deletions and plasmid presence, in the same way as for the other growth conditions. Since *C. necator* null mutants for RubisCO are not viable under organo-autotrophic conditions, *C. necator* Δ*panC* was the only strain that could be used as a control for growth in these experiments. This strain showed a protein band at about 55 kDa in the membrane fraction. A possible explanation for this is that some of the soluble RubisCO may have been carried-over into this fraction. Alternatively, this strain may be expressing a membrane protein of similar molecular weight as RubisCO (Figure 3.14). The comparison between *C. necator* Δ*panC* harbouring the empty vector pMTL71301::*panC*::P_{phaC} and the *C. necator* ΔΔ*cbbLS* Δ*panC* mutant, complemented with the pMTL71301::*panC*::P_{phaC} *Hn*CB Act plasmid

did not show any major differences in the protein expression patterns of their membrane fractions. Also in this case, no bands with molecular weights comparable with the CbbO and CbbQ proteins were detected in either the cytosolic soluble or the membrane fractions. On the other hand, significantly higher expression levels of RubisCO were observed in the cytosolic fractions of protein samples obtained from the strains expressing the *cso* operon with the RubisCO activases, with respect to the control strain. No further differences could be seen in the overall protein expression patterns of these strains (Figure 3.14).

Overall, the absence of a high level of production of RubisCO in cytosolic fraction of the *C. necator* control strain could also signify that low production of the RubisCO protein is sufficient for survival. In the same way, it is possible that the expression level of the carboxysomal genes, including RubisCO, necessary for survival in organo-autotrophic conditions could be lower than the detection limit and therefore not visible in the membrane faction of the gel.

Protein samples were also collected, under the same three experimental conditions, from the PHB-negative isogenic mutants of the strains described above (*C. necator* ΔΔ*cbbLS* Δ*panC* Δ*phaCAB* genetic background) harbouring plasmids expressing the *cso* operon, either alone or with the activases. The *C. necator* Δ*panC* Δ*phaCAB*/pMTL71301::*panC*::P_{phaC} strain was used as the positive control for growth. Again, the presence of the plasmids and of the in-frame deletions of the two *cbbLS* operons were confirmed by PCR in the overnight cultures used to set up the experiments. Unfortunately, even in this case, CB genes expression could not be observed in any of these strains, under any of the conditions tested (Appendix 3, 4 and 5).

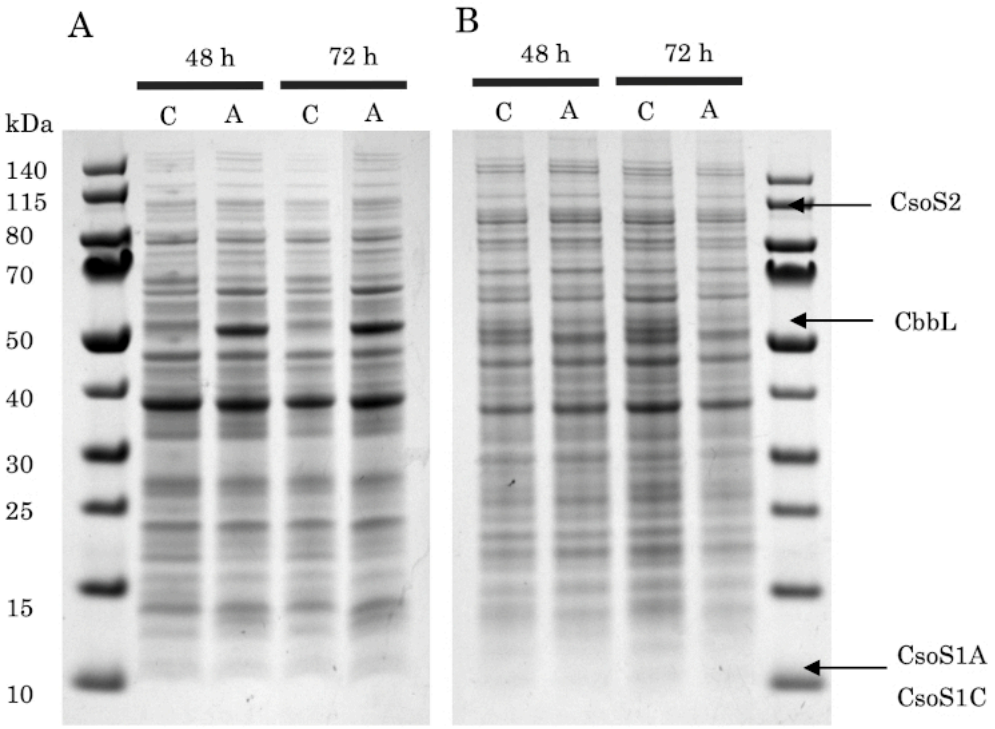


Figure 3.14: Expression of *cso* operon with or without RubisCO activases in *C. necator* ΔΔ*cbbLS* Δ*panC* grown in 80mM Na-formate MM. SDS-PAGE of the soluble (A) and membrane (B) fractions of *C. necator* ΔΔ*cbbLS* Δ*panC* harbouring stable plasmids for the constitutive heterologous expression of *H. naepolitanous* α-carboxysomes, with or without RubisCO activases. Protein samples were collected after 48 and 72 h of growth in 80mM Na-formate MM, in the presence of 10% CO_2. A = *C.necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} *Hn*CB Act; C= control *C.necator* Δ*panC* /pMTL71301::*panC*::P_{phaC}.

3.2.3.3 Imaging of *C. necator* strains co-expressing genes for carboxysomes and RubisCO activases, using transmission electron microscopy (TEM)

3.2.3.3.1 TEM

TEM imaging was attempted on both PHB-producing and PHB-negative (Δ*phaCAB*) *C. necator* strains. This was done to test whether the presence of PHB granules could interfere with the imaging of CB structures. TEM analysis was carried out using the protocol and settings previously described for the detection of carboxysome shells (section 2.7.1). The primary aim of this experiments was to investigate whether CB structures could still form in the presence of the RubisCO activases CbbO and CbbQ, in order to further validate and support the *in vivo* functionality complementation data (discussed in section 3.2.4).

Heterologous expression of the *cso* operon-encoded proteins with the RubisCO activases was observed in the *C. necator* Δ*panC* ΔΔ*cbbLS* strain after 24 h from culture re-inoculation, using the P_{phaC} constitutive promoter (Figure 3.15). Consistently with the data previously obtained in the *C. necator* Δ*phaCAB* genetic background, carboxysome shell structures were detected also in the cytosol of PHB-producing *C. necator* strains expressing the *cso* operon alone, where these were found in the proximity of PHB granules. Unfortunately, the imaging resolution across the samples was not as good as in the previous experiments. A series of pictures were taken across the TEM microgrid for the *C. necator* strain carrying the carboxysomes with RubisCO activases, however high-resolution imaging was difficult to achieve due to the presence of electron-dense areas. These are likely to form as a consequence of the high production of CBs in the cytoplasm and represent a barrier for the electron beam of the TEM, thus compromising the quality of the CB architecture imaging. Moreover, the imaging in these strains proved to be more difficult than previously experienced with the Δ*pha*CAB strain due to the brittle nature of PHA granules, which led to the formation of holes in

the 80 nm microscopy slice, when the electron beam was focused on the sample. Because of this, only the thicker regions of the sample slices could be imaged successfully.

The same experiment was carried out in the *C. necator* Δ*panC* Δ*pha*CAB ΔΔ*cbbLS* strain, harbouring the stable plasmid expressing the CB and activases genes under the control of the P_{phaC} constitutive promoter. Culture samples were collected after 12 and 24 h from re-inoculation (Figure 3.16). Even in this case, carboxysome structures could be detected in the cytoplasm of the control strain carrying only the *cso*-operon, at the 24 h time point. As mentioned above, the imaging resolution was lower than expected and it was not possible to measure the CBs size. Heterologous expression of the CB shell proteins was observed in the Δ*pha*CAB strain carrying the *cso*-operon with activases, between 12 and 24 h after re-inoculation and with levels increasing over time. As for the control sample, the imaging resolution of the structures detected at the 12 h time point was not good enough to visualise the details of the CB microcompartments. The presence of fully formed CB structures could not be confirmed in the 24 h samples either, due to the presence of high electron density areas in the cytoplasm of the analysed bacterial cells.

Overall, the data obtained from TEM analysis seem to suggest that heterologously expressed CB structures may form in *C. necator* even in the presence of the RubisCO activases and in both PHB-producing and PHB-negative strains. However, further improvements of the microscopy technique will be required to obtain a more detailed resolution of the CB microcompartments structure in the cytoplasm of *C. necator*, in a consistent and reproducible manner.

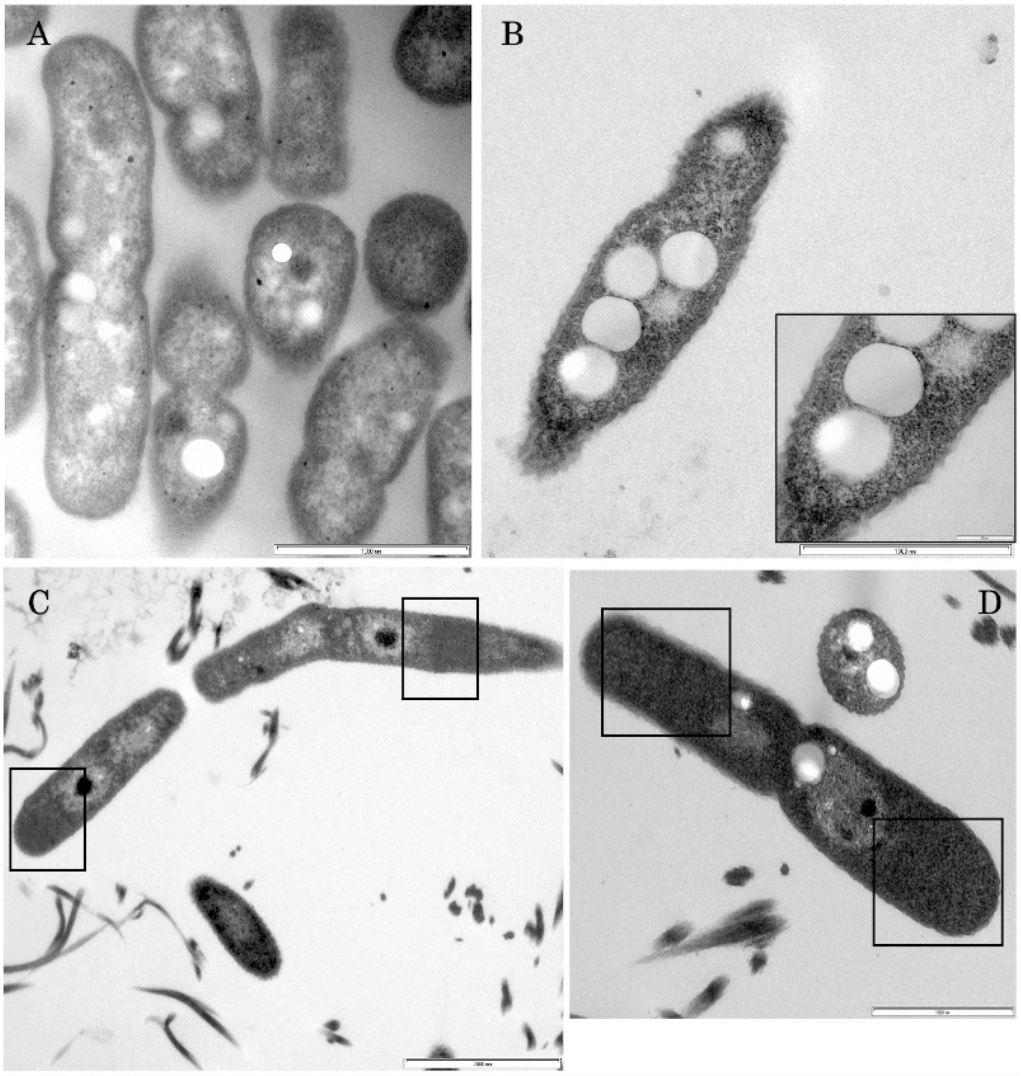


Figure 3.15: TEM imaging of *C. necator* Δ*panC* strains expressing *cso* operon with and without RubisCO activases. TEM images of carboxysomes taken *in vivo* from culture samples collected at 24 h after set back of (A) negative control: *C. necator* Δ*panC*/pMTL71301::*panC*::P_{phaC}; (B) *C. necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} *Hn*CB (Inset) longitudinal section with (Inset) icosahedral structure. Scale bars: 1000 nm. (C) *C. necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} *Hn*CB Act field view and (D) Same as C, with details of highly electron dense areas (black boxes) and icosahedral structures in inclusion bodies. Scale bars: 2000 nm and 1000 nm, respectively.

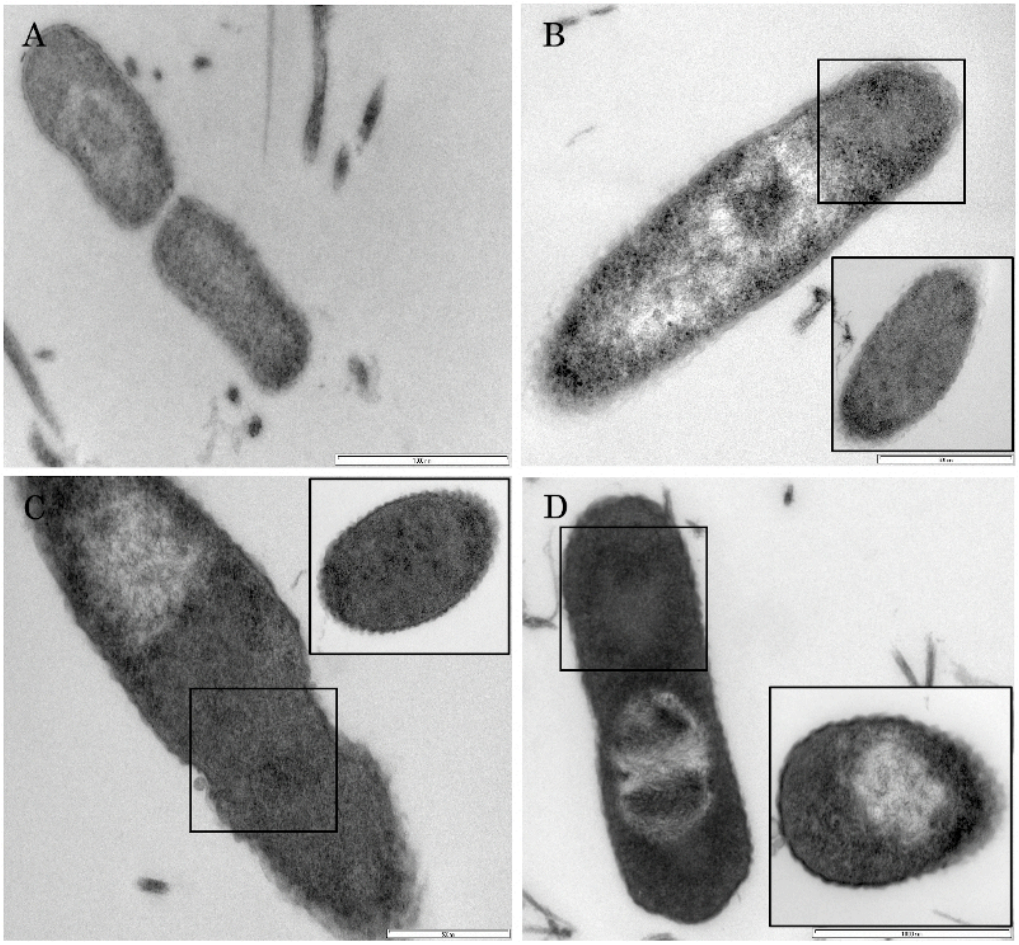


Figure 3.16: TEM imaging of *C. necator* Δ*phaCAB* Δ*panC* strains expressing *cso* operon with and without RubisCO activases. TEM images of carboxysomes taken *in vivo* from culture samples collected at 24 h after set back of (A) negative control: *C. necator* Δ*phaCAB* Δ*panC* /pMTL71301::*panC*::P_{phaC}; (B) *C. necator* Δ*pha*CAB ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} *Hn*CB longitudinal section and (Inset) transversal section with icosahedral structures. Scale bars: 1000 nm and 500 nm, respectively; (C) C. *necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} *Hn*CB Act at 12 hr after set back; (D) Same as C at 24hr after set back. For both images, the icosahedral structures were highlighted by black boxes with insets showing transversal section of the bacteria. Scale bars: 500 nm and 1000 nm, respectively.

3.2.3.3.2 Cryo-TEM tomography

The technical limitations experienced during previous TEM image acquisitions were mostly due to the laborious sample preparation procedure, the limited electron beam intensity (max accelerated voltage 100 kV), and the resolution power of the camera (Gatan SIS Megaview IV) installed on the microscope (Tecnai G2 12 BioTWIN Transmission Electron Microscope). To overcome these problems it was decided that Cryo-TEM tomography would be used. The machine (JEOL 2100F FEG-TEM) used in the nmRC centre was equipped with an electron source able to reach an accelerated voltage of 200 kV, for better sample penetration, and with a tilting platform that allows multiple sample acquisitions over an angle of +/- 70 degrees, which in combination with the camera (Gatan Tridiem Filter Spectrometer and 2K x 2K CCD camera) and image deconvolution software would reduce the background noise level, allowing higher final image resolution. Furthermore, the sample preparation requires a shorter procedure possibly helping to boost consistency. Unfortunately, despite the fact that Cryo-TEM tomography session was planned and the sample preparation carried out, the imaging was never completed due the pandemic outbreak of Covid-19 and subsequent lack of access to the nmRC's facilities.

3.2.4 Function of carboxysomes with activases in a heterologous host

To assess the growth complementation ability of the *cso* operon with the RubisCO activases (CbbQ and CbbO) a *C. necator cbbLS* deletion mutant strain carrying the *cso-csocbbOQ* genes was grown under organo-autotrophic conditions (80mM Na-formate MM, in the presence of 10% CO_2) in the BioLector Pro microbioreactor for 48 h. In this set of experiments, *C. necator ΔpanC* and its isogenic *ΔΔcbbLS* mutant were used as the positive and negative control for growth, respectively. Two media were tested for the overnight cultures: MM with 0.4%(w/v) SG and 80mM Na-formate, in the presence of 10% CO_2. As shown in Figure 3.17 the *cso-csocbbOQ* operon was able to restore the ability of the *C. necator ΔΔcbbLS* deletion strain to grow under organo-autotrophic conditions, independently of the culture media used to set up the seed cultures. However, the *C. necator* RubisCO null mutant carrying the complete *cso-csocbbOQ* operon exhibited a longer lag phase (about 8 h) when this was grown overnight in heterotrophic conditions, with respect to organo-autotrophic conditions (lag phase of about 4 h). This result was somewhat expected since the strain pre-cultured under autotrophic conditions would already be adapted to this environment. Nevertheless, no significant differences in terms of doubling time and μ_{max} values could be observed between the two conditions, during the exponential growth phase (overnight in autotrophic conditions: μ_{max} = 0.104 h^{-1}, T_d =6.6 h; heterotrophic conditions: μ_{max} = 0.118 h^{-1}, T_d =5.85 h). When the same strain was grown again in organo-autotrophic conditions after plating and overnight cultures in heterotrophic conditions, a reduction of the lag phase to about 5 h (Figure 3.18) and a slight reduction in the average μ_{max} to 0.106 h^{-1} could be observed. Interestingly, this was now significantly higher than that recorded for the positive control strain (μ_{max} 0.090 h^{-1}). These results were very different from the previous experiments, where the wild type strain

reached maximum growth rates of $\mu_{max\text{-}hetero}$ = 0.150 h^{-1} and $\mu_{max\text{-}auto}$ = 0.165 h^{-1}, with doubling times of 4.63 h and 4.21 h when re-inoculated from seed cultures grown in heterotrophic and organo-autotrophic conditions, respectively. Biomass accumulation was measured at the onset of the stationary phase (after about 30 h of growth) for the *C. necator* ΔΔ*cbbLS* Δ*panC* mutant complemented with the pMTL71301::*panC*::P_{phaC} *Hn*CB Act plasmid and the positive control *C. necator* Δ*panC*/pMTL71301::*panC*::P_{phaC}. No statistically significant differences could be observed in the biomass produced by these two strains (ANOVA one-way analysis). On the other hand, the *C. necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} (negative control) and *C. necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} *Hn*CB strains failed to grow under organo-autotrophic conditions, as confirmed by the observation that their total biomass did not change throughout the duration of the experiment and was significantly lower ($p<0.001$) than that produced by the positive control strain.

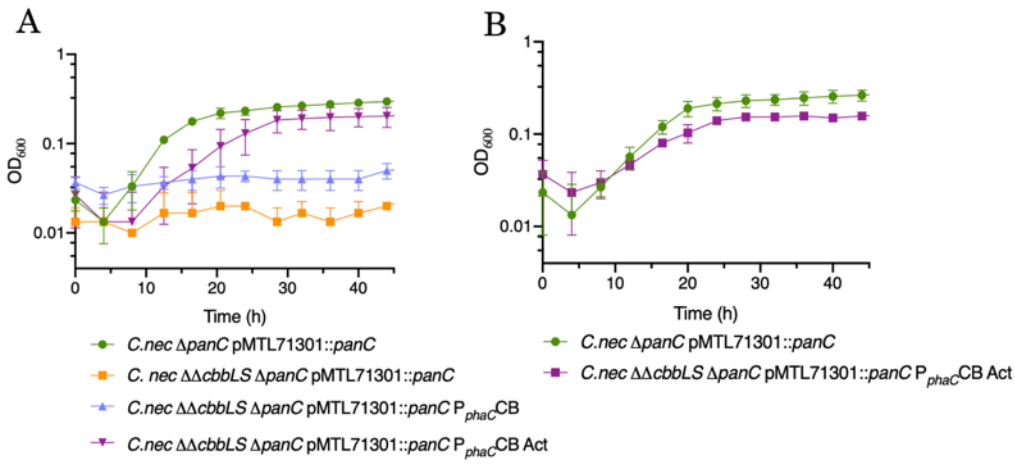


Figure 3.17: Growth curve of *C. necator* ΔΔ*cbbLS* Δ*panC* expressing the *cso* operon with and without RubisCO activases in organo-autotrophic conditions. Growth curves of *C. necator* ΔΔ*cbbLS* Δ*panC* mutants harbouring stable plasmids for the constitutive expression of caboxysome-encoding genes, either alone or in combination with the RubisCO activase genes (P_{phaC} *Hn*CB and P_{phaC} *Hn*CB Act, respectively) under organo-autotrophic conditions (80mM Na-formate MM, in the presence of 10% CO_2) with pre-cultures grown in (A) SG MM or (B) organo-autotrophic conditions

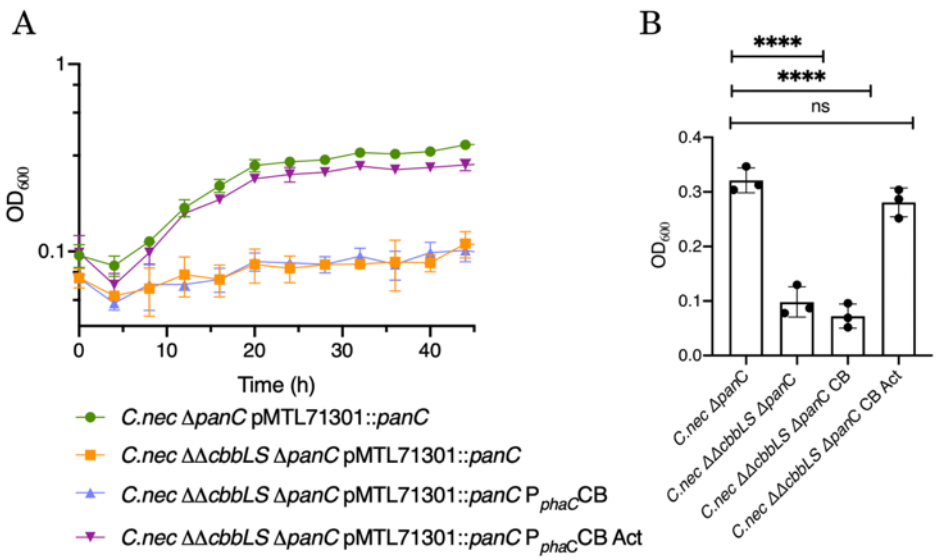


Figure 3.18: Functional complementation of adapted *C. necator* ΔΔ*cbbLS* Δ*panC* expressing the *cso* operon with and without RubisCO activases in organo-autotrophic conditions. (A) Growth curves of *C. necator* ΔΔ*cbbLS* Δ*panC* mutants harbouring stable plasmids for the constitutive expression of caboxysome-encoding genes, either alone or in combination with the RubisCO activase genes (P_{phaC} *Hn*CB and P_{phaC} *Hn*CB Act, respectively) (B) One-way ANOVA analysis of the biomass produced by *C. necator* Δ*panC*/pMTL71301::*panC*; *C. necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*, *C. necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} *Hn*CB and *C. necator* ΔΔ*cbbLS* Δ*panC*/pMTL71301:: *panC*::P_{phaC} *Hn*CB Act, after 30 h of growth under organo-autotrophic conditions (80mM Na-formate MM, in the presence of 10% CO_2).

The experiments previously described were repeated in the Δ*phaCAB* genetic background, using the *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC*/ pMTL71301::*panC*::P_{phaC} *Hn*CB and *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} *Hn*CB Act strains. The growing conditions of the overnight pre-cultures and the BioLector Pro settings were the same as described before, with the only difference being the BioLector running time, which was extended from 48 to 72 h. Plasmid stability for the overnight cultures was assessed using PCR technique targeting *cbbLS* genes. Interestingly, these plasmids were particularly unstable in PHB-negative strains when these were grown in SG MM, even when this was supplemented with tetracycline and despite the presence of the *panC*-based plasmid addiction system. Therefore, seed cultures were also set up under organo-autotrophic conditions to take advantage of the additional selective pressure provided by the *in trans* complementation of RubisCO null mutants of *C. necator*, which are unable to grow autotrophically. After a long lag phase (of approximately 24 h) the *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC*/ pMTL71301::*panC*::P_{phaC} *Hn*CB Act strain grew to an OD_{600} of approximately 3.5. The presence of the plasmid was confirmed by PCR for the seed cultures under heterotrophic and organo-autotrophic growth conditions and these cultures were used to inoculate a BioLector plate. Genomic DNA was also extracted and purified from the seed culture and sent for sequencing together with the DNA sample obtained from its parental strain *C. necator* Δ*phaCAB* ΔΔ*cbbLS*, to investigate whether any mutations were selected during the adaptation to organo-autotrophic conditions. Consistently with previous data, the expression of carboxysome-encoding genes in combination with the RubisCO activases was able to complement the growth defects of RubisCO null mutants under autotrophic conditions also in the Δ*phaCAB* genetic background (Figure 3.19). Again, a longer lag phase (about 40 h) was observed for the strain grown overnight in heterotrophic conditions, as compared

to the ones grown in autotrophic conditions (between 15 and 20 h). As previously seen with PHB-producing *C.necator* strains, no major differences were observed in the doubling times and μ_{max} values of the two cultures (pre-cultures grown in autotrophic conditions: μ_{max} = 0.116 h^{-1}, T_d = 5.9 h; pre-cultures grown in heterotrophic conditions: μ_{max} = 0.111 h^{-1}, T_d =6.2 h). As before, the same strains were grown again in organo-autotrophic conditions after plating and overnight cultures in heterotrophic conditions. Interestingly, this time there was a significant reduction of the lag phase, which went down by over 4-fold to about 8-10 h (Figure 3.20). This reduction in lag phase was also accompanied by an increase in the μ_{max} to 0.120 h^{-1}, which was almost identical to that of the positive control strain (μ_{max} 0.118 h^{-1}). One-way ANOVA analysis of the biomass at the onset of the stationary phase (30 h) showed no statistically significant differences between the biomass produced by the positive control and the *C. necator ΔphaCAB ΔΔcbbLS ΔpanC*/ pMTL71301::*panC*::P_{phaC} *Hn*CB Act strain. On the contrary, the total biomass of the two isogenic *ΔΔcbbLS ΔphaCAB* mutants carrying either an empty plasmid or the *cso*-operon remained constant throughout the duration of the experiment and was significantly lower than that of the positive control ($p<0.001$).

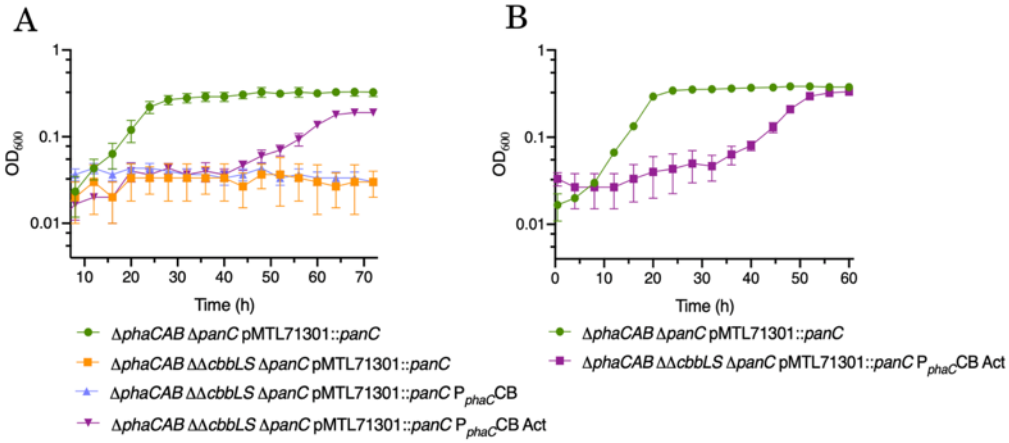


Figure 3.19: Growth curve of *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* expressing the *cso* operon with and without RubisCO activases in organo-autotrophic conditions. Growth curves of *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* mutants harbouring stable plasmids for the constitutive expression of caboxysome-encoding genes, either alone or in combination with the RubisCO activase genes (P_{phaC} *Hn*CB and P_{phaC} *Hn*CB Act, respectively) under organo-autotrophic conditions (80mM Na-formate MM, in the presence of 10% CO_2) with pre-cultures grown in (A) SG MM or (B) organo-autotrophic conditions.

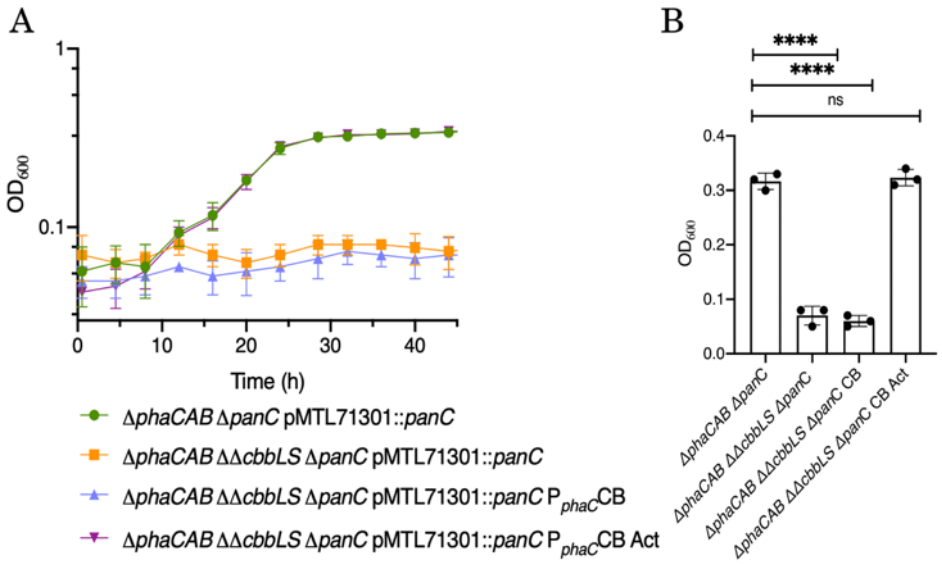


Figure 3.20: Functional complementation of adapted *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* expressing the *cso* operon with and without RubisCO activases in organo-autotrophic conditions. (A) Growth curves of *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* mutants harbouring stable plasmids for the constitutive expression of caboxysome-encoding genes, either alone or in combination with the RubisCO activase genes (P_{phaC} HnCB and P_{phaC} *Hn*CB Act, respectively) (B) One-way ANOVA analysis of the biomass produced by *C. necator* Δ*phaCAB* Δ*panC*/pMTL71301::*panC*; *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*, *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC*/pMTL71301::*panC*::P_{phaC} HnCB and *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC*/pMTL71301::pMTL71301::*panC*::P_{phaC} *Hn*CB Act, after 30 hr of growth under organo-autotrophic conditions (80mM Na-formate MM, in the presence of 10% CO_2).

The analysis of the NGS data showed that there were no differences in the genomic DNA between the *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* parental strain and the one carrying the plasmid expressing the *cso* operon with RubisCO activases. Unfortunately, the analysis highlighted, as well, that the pMTL71301::*panC*::P_{phaC} *Hn*CB Act plasmid integrity was compromised. Indeed, a deletion of the majority of the *cso* operon was observed, starting from 200 bp inside the *cbbL* gene and ending around 800 bp into the *cbbO* gene. Although, the strain selection was done in organo-autotrophic conditions the growth of the strain for genome sequencing was carried out in heterotrophic condition to reach higher genomic DNA yields. Therefore, this outcome is possibly due the plasmid rearrangement operated by the bacteria during the growth under less selective heterotrophic conditions in order to reduce the metabolic burden.

3.3 Discussion

The main goal of this chapter was to test whether α-carboxysomes from *H. neapolitanus* could be heterologously produced and function in *C. necator* H16, in order to improve carbon fixation efficiency in this species. The expression of carboxysome-encoding genes was attempted at first in *E. coli*. This was done as a positive control to ensure that the *cso* synthetic operon was correctly assembled in the expression vector and that carboxysomal proteins could be visualised using SDS-PAGE analysis. Indeed, the heterologous expression of CB genes in *E. coli* has already been reported in the literature (Bonacci *et al.*, 2012). After confirming that CBs could be produced in *E. coli*, the carboxysome-encoding genes were also introduced in *C. necator* H16. The recovery of the carboxysomal components in the membrane fractions of *E. coli* and *C. necator* whole-cell protein samples suggested their correct assembly. Similar CB gene expression patterns were observed across the two species, confirming that CBs were successfully expressed also in *C. necator*. Moreover, the proteins present in the four major bands were identified as carboxysomal components by MALDI-TOF-MS protein identification analysis. Altogether, these results strongly suggest that the *H. naepolitanus cso* operon was successfully produced in *C. necator* H16. To the best of our knowledge, this is the first report in the literature to demonstrate the heterologous production of α-carboxysomes in this bacterial species.

Further light was shed on the ability of CBs to assemble in *C. necator* by TEM imaging. The correct structural assembly of the α-carboxysome shell proteins was easily detected in *E. coli*. This prompted us to carry out the same analysis in the *C. necator* Δ*phaCAB* strain derivatives expressing the *cso* operon. PHB-negative *C. necator* strains were selected for this analysis since the absence of PHB granules should facilitate the imaging of the microbial cytosol.

A higher image definition was achieved with the *C. necator* Δ*phaCAB* strain constitutively expressing the *cso* operon, with respect to the same strain carrying the CB-encoding genes under the control of the P_{BAD} inducible promoter. The transversal sections of this strain's microscopy slices gave a better contrast against the background, allowing a clearer view of the CB structures in their closed conformation (Figure 3.7 C). These results contribute to support the data obtained from the SDS-PAGE analysis of whole-cell protein extracts, in which RubisCO was recovered in the membrane fraction, suggesting that this protein may be sequestered inside of the carboxysomes. These data demonstrate that the proteins encoded by the *H. neapolitanus cso* operon are capable of self-assembling into closed structures when they are heterologously produced in *C. necator* H16.

As previously introduced, the correct assembly and function of RubisCO from *H. neapolitanus* (*Hn*RubisCo) in *C. necator* is of pivotal importance for the expression of functional carboxysomes in this species. For this reason, we sought to investigate the ability of *Hn*RubisCO to assemble in the cytosolic compartment of *C. necator,* in the absence of the *H. neapolitanus*-specific molecular chaperone acRAF (Wheatley *et al.*, 2014). Our data shows that the constitutively expressed *Hn*RubisCO was able to complement the growth defects of *C. necator* ΔΔ*cbbLS* and *C. necator* Δ*phaCAB* ΔΔ*cbbLS* strains in organo-autotrophic conditions. These observations appear to suggest that acRAF is not essential for the correct folding of *Hn*RubisCO in *C. necator* H16. However, there is still the possibility that this accessory protein may play a more crucial role in this process when *Hn*RubisCO is associated with carboxysomal micro-structures. Nevertheless, the observation that expression of the *cso* operon was not sufficient to restore the ability of *C. necator cbbLS* null mutants to grow under autotrophic conditions, showed that, despite being assembled correctly, CBs are ultimately not functional in this species. This result

seems to provide additional evidence in support of the existence of a range of accessory proteins, which may be involved in ensuring the correct functionality of CB micro-compartments, despite being encoded by genes not belonging to the *cso* operon, as it was recently suggested (Turmo *et al.*, 2017). In particular, CsoCbbQ and CsoCbbO were shown to interact with the CB shell proteins (Sutter *et al.*, 2015) and to be involved in restoring and remodelling the active conformation of *Hn*RubisCO (Tsai *et al.*, 2015). Therefore, it was hypothesised that the absence of RubisCO activases may have prevented the recycling of active RubisCO within the CB shells, thus compromising CB functionality. To test this, the genes coding for the RubisCO activases were cloned downstream of the *cso* operon and expressed in *C. necator* RubisCO null mutants. Unlike the derivatives of the *C. necator* ΔΔ*cbbLS* strains expressing the *cso* genes alone, expression of the CB-encoding genes in combination with the RubisCO activases restored the ability of *C. necator* ΔΔ*cbbLS* to grow under organo-autotrophic conditions (Na-formate MM). This result suggests that, when co-expressed with the RubisCO activases CsoCbbO and CsoCbbQ, α-carboxysomes from *H. neapolitanus* appear to function *in vivo* in the heterologous host *C. necator* H16. However, there is still the possibility that this growth complementation may have been the result of RubisCO being expressed in the cytosol of *C. necator* cells, independently of the CBs. To rule out this hypothesis, protein expression was tested in these strains and in the *E. coli* DH5α carrying the synthetic operon formed by the *cso* and RubisCO activases genes. In the latter, expression of the *cso*-encoded proteins could still be observed in the membrane fractions of the protein samples. However, no bands corresponding to the activases CsoCbbO or CsoCbbQ were detected . In the *C. necator* ΔΔ*cbbLS* strains, instead, no bands associated with either the carboxysomal components or the RubisCO activases could be observed in the membrane fractions of the protein samples obtained from

cultures grown under either heterotrophic (LB) or organo-autotrophic conditions. On the other hand, expression of RubisCO, which increased over time, was detected in the soluble fractions of the same protein samples, suggesting that high concentrations of this enzyme may have been present in the cytosol of these strains. This observation may suggest that cloning of the genes coding for the RubisCO activases downstream of the *cso* operon might have affected the expression of the latter, particularly in *C. necator*. These findings also provide additional evidence in support of the hypothesis according to which it may be the cytosolic RubisCO that is responsible for the complementation of the growth defects of *C. necator* RubisCO null mutants, under autotrophic conditions.

However, it must be pointed out that the TEM imaging data seem to suggest the presence of heterologously produced CB structures in both *C. necator* PHB-producing and PHB-negative (Δ*phaCAB*) strains harbouring the *cso* operon in combination with the RubisCO activases genes (Figure 3.15 and Figure 3.16). Unfortunately, the resolution of the images was not detailed enough to assess the structural integrity and the architecture of the CB structures.

To shed light on the RubisCO co-localisation issue within the CB shell proteins, the use of super-resolution microscopy techniques was evaluated, however this was discarded, as its resolution falls around the 100 nm range. Therefore, the technique wouldn't allow the reliable placement of RubisCO either inside or outside of the shell.

Overall, the data in our possess is not conclusive enough to demonstrate the *in vivo* functionality of α-carboxysomes in *C. necator* H16. Indeed, even though growth complementation of *C. necator* RubisCO null mutants was observed under autotrophic conditions, which is a significant achievement, it was not possible to unequivocally demonstrate either the closed conformation of the CB structures or the encapsulation of RubisCO and its activases inside of

the CBs. Further investigations will be needed to shed light on CB architecture, *in vivo* activity and on the process of enzyme encapsulation in these structures in *C. necator* H16.

3.4 Future work

To further investigate the reasons behind the changes in gene expression patterns observed upon addition of the RubisCO activases genes downstream of the *cso* operon, it would be interesting to analyse the expression of the *cso* and RubisCO activases genes at the transcriptional level, taking advantage of molecular techniques such as qPCR or Real-Time RT-PCR. A similar approach would also prove very useful for the analysis of the relative mRNA abundance for each gene in the operon. Indeed, this information could be used to design alternative cloning strategies to fine tune expression of these genes and achieve optimal stoichiometric ratios between the CB components and their accessory proteins. For instance, the cloning of an edited synthetic DNA construct carrying the *cso* operon and the *csocbbOQ* genes under the control of two different promoters should result in increased expression levels of RubisCO activases.

Furthermore, deletion of the *cso* CA gene (*csoSCA*) would help confirm the internal localisation to the α-CB of the active *H. neapolitanus* RubisCO.

In addition, recent studies (Dai *et al.*, 2018) were able to identify individual RubisCO molecules and their assembly intermediates, leading to the formation of carboxysomes, inside of Syn5 cyanophage-infected cyanobacteria (*Synechococcus sp.*) cells, using Cryo-Electron tomography. In light of the recent technological development in the field, similar experimental approaches could be used to further investigate the integrity of the carboxysomal micro-structures, as well as their possible co-localisation with RubisCO, in *C. necator* strains co-expressing the genes coding for the carboxysomes and the RubisCO activases.

Chapter 4 - Biological relevance of RubisCO surface charge for recruitment and incorporation in α-carboxysomes: Insights from computational modelling

Despite several hypotheses about the function of the intrinsically disordered CsoS2 protein being proposed in the last couple of decades, the details behind RubisCO recruitment inside of the α-carboxysome remained unclear until very recently, when the key interactions between the CsoS2 protein and RubisCO were unveiled (Oltrogge *et al.*, 2020).

In this chapter, I studied the relevance of the electrostatic potential at the surface of RubisCO as a driving force for the recruitment of the holoenzyme inside of the carboxysome shells. After a preliminary analysis of the surface charge of a carboxysomal (*H. naepoliatnus*) and cytosolic (*C. necator*) RubisCO, it was decided to create a mutated variant of *Cn*RubisCO (*CnMut*RubisCO)to resemble the electrostatic potential of the *H. naepolitanus* wild type RubisCO (*Hn*RubisCO) encapsulated in the carboxysomes. Furthermore, by employing molecular dynamics (MD) simulations I explored the structural stability of *Hn*RubisCO, *Cn*RubisCO and *CnMut*RubisCO in aqueous solution as well as the pattern of interaction of the proteins with anions and cations at three ionic strengths (20 mM, 150 mM and 1000 mM).

4.1 Introduction

In the last decade, thanks to the advances in technology, significant progress was made in the understanding of structure, function and assembly mechanisms of the BMCs, in general, and more specifically of the α-carboxysome. This knowledge has broadened our understanding of their role in the bacterial metabolism, enabling, at the same time, their use in a variety of synthetic biology applications (Kerfeld *et al.*, 2018). The assembly process of the α-carboxysome has always been more challenging to unravel than their β-carboxysome counterparts, given the absence of recognisable interaction domains in the intrinsically disordered assembly protein, CsoS2 (Cannon *et al.*, 2003; Cai, Dou, *et al.*, 2015). Until very recently, the details of interaction between the RubisCO and CsoS2 were not completely established. Observation of the difference in electrostatic potential between the form IAc *Hn*RubisCO and β-carboxysomal (form IBc) RubisCO and the plant form IB (Zarzycki *et al.*, 2012), led to discussion of the its potential role in the recruitment mechanisms for α-CBs. The positive electrostatic potential of the pore of the hexameric shell proteins, has been suggested to offer a kinetic advantage for bicarbonate diffusion in both α- and β-carboxysomes (Tsai *et al.*, 2007; Tanaka *et al.*, 2009) and to act as a diffusion barrier for the undesirable substrates, CO_2 and O_2. Furthermore, large scale modelling of the electrostatic interactions between protein binding partners, highlighted the importance of these forces for the attraction of partners at long distance and guidance towards their physical binding at short distances (Mihiri Shashikala *et al.*, 2019).

The aim of this study was to investigate the importance of the RubisCO electrostatic potential as a driving force for the recruitment of the catalytically active enzyme inside of the carboxysome shells. Our central hypothesis was that RubisCO surface charge and its electrostatic potential could be involved in the attraction of the

positively charged N-terminal region of the CsoS2 protein, as an assembly mechanism used by the original host of the *cso* operon *H. neapolitanus*. Since *C. necator* H16 is a β-proteobacteria, in which carboxysomes do not occur naturally the electrostatic potential of the carboxysomal *Hn*RubisCO was compared against the *C. necator* RubisCO (*Cn*RubisCO). Furthermore, modifications of the electrostatic potential of the latter were carried out, creating a mutant *Cn*RubisCO (*CnMut*RubisCO) which would resemble the electrostatic potential of *Hn*RubisCO. Resulting trajectories from MD simulations were processed by performing the root-mean-square deviation (RMSD) and RMS fluctuation analysis of the three proteins in water, as well as their interaction with Cl^- and Na^+ ions at three different salt concentrations (20 mM, 150 mM and 1000 mM).

4.2 Results

4.2.1 Electrostatic potential of *H. naepolitanus* and *C. necator* H16 RubisCO in gas phase

To investigate the significance of the surface charge on the recruitment process of the RubisCO complex from CsoS2 during carboxysome assembly, the RubisCO from *Cupriavidus necator* H16 (*Cn*RubisCO) was modified to feature a similar electrostatic potential to the enzyme dimer of *Halothoibacillus naepolitanus* (*Hn*RubisCO). After running a single-point quantum mechanics (QM) calculations at the semi-empirical level of theory using AM1 method (Dewar et al., 1985) the molecular electrostatic potential (MEP) of both RubisCO dimeric proteins (*Cn*RubisCO and *Hn*RubisCO) in the gas phase was determined (Figure 4.1). Å

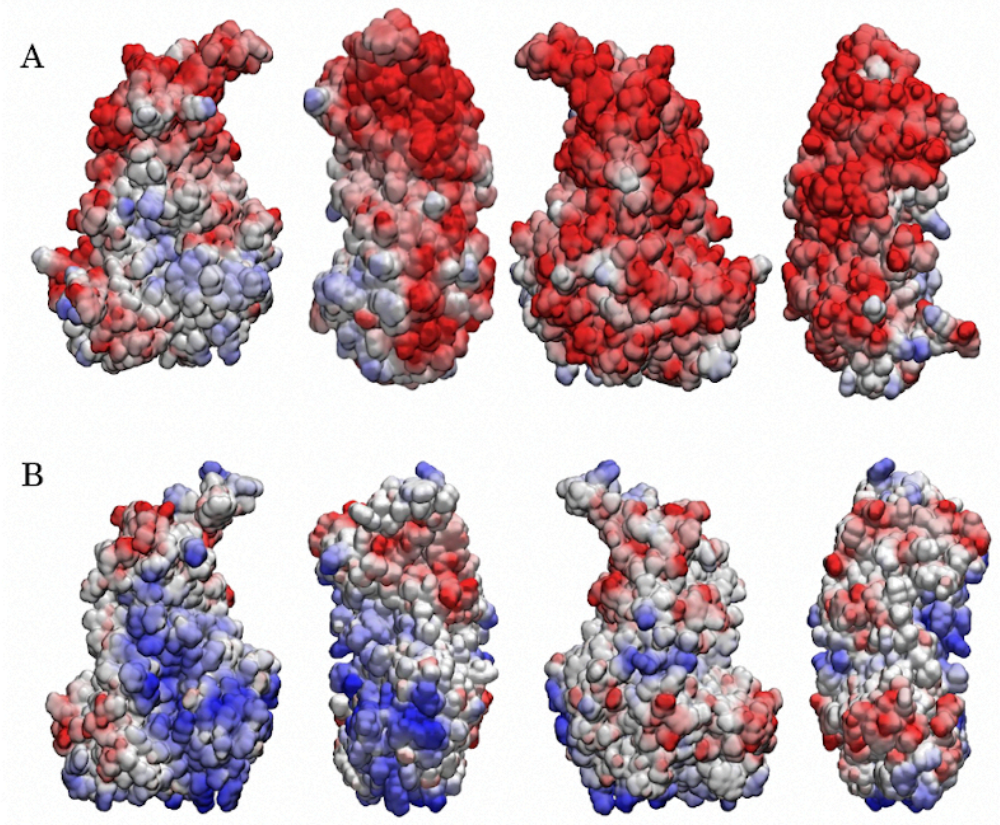


Figure 4.1: Representation of the electrostatic surface potential of *Hn*RubisCO and *Cn*RubisCO large and small subunit calculated in gas phase. (A) *Hn*RubisCO (PBD 1svd) and (B) *Cn*RubisCO (PBD 1bxn) MEP based on gas phase AM1 semiempirical calculation. Colour range is set from -100 to 50 kcal/mol (red to blue) and the surface isovalue is set to 0.001 e/Å^3.

*Hn*RubisCO shows a much more negative surface charge (shown in red) localised at the top and back of the protein, while *Cn*RubisCO overall has a more neutral (shown in white) and positive (depicted in blue) charge distribution. At first, five possible mutation sites, located in flexible loops outside of the catalytic interface were identified to create a redistribution of charges in the overall molecular electrostatic potential of the *Cn*RubisCO dimer. Based on the further investigation, it was noticed that two of the flexible loops selected as mutation sites were located at the interface between opposing RubisCO dimers (CbbLS). The mutation of those regions was not considered feasible, given the possibility of protein misfolding during the formation of the catalytically active hexadecameric structure.

Special care was taken not to modify the loops involved in the catalytic activity of the holoenzyme as well as the interaction with RubisCO activases. Finally, seven mutations (Thr33, Leu35, Leu76, Asp79, Asp92, Gly96 and Asn97) localised in three flexible loops of the large subunit of *Cn*RubisCO were selected and created on the pre-

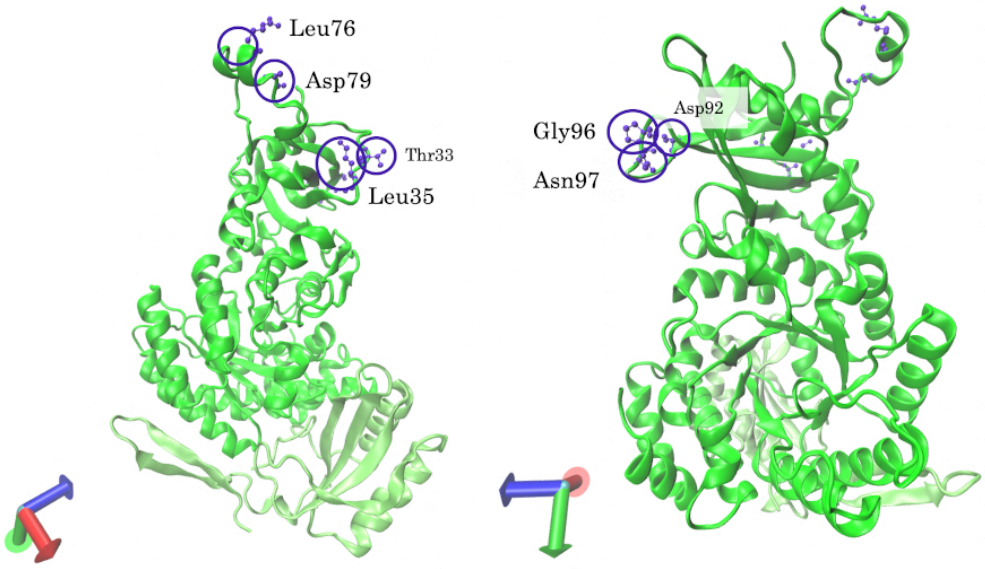


Figure 4.2: *Cn*RubisCO dimeric protein structure with highlighted mutation sites. Mutation sites are displayed in violet CPK representation (Thr33, Leu35, Leu76, Asp79, Asp92, Gly96 and Asn97). Large subunit of *Cn*RubisCO (CbbL) is shown in green, while the small subunit (CbbS) is depicted in lime cartoon representation.

existing X-ray structure of *Cn*RubisCO (PBD 1bxn) (Hansen *et al.*, 1999) using PyMol software (Schrödinger, LLC, 2015) (Figure 4.2).

A new mutant structure was generated (*CnMut*RubisCO) and subjected to a similar semi-empirical AM1 calculations in gas phase as described earlier. As it can be observed from Figure 4.3 the predicted rearrangement of the overall electrostatic potential in the *CnMut*RubisCO structure showed a striking resemblance to the *Hn*RubisCO.

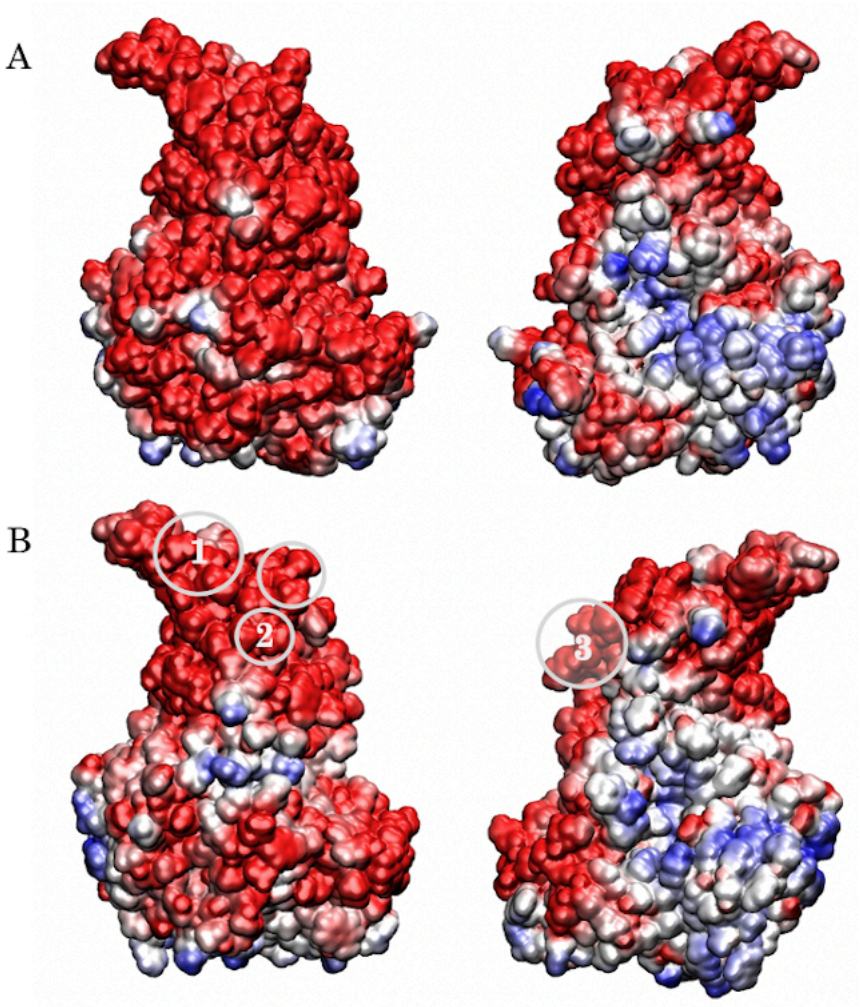


Figure 4.3: Comparison between electrostatic potential of *Hn*RubisCO and *CnMut*RubisCO. (A) *Hn*RubisCO and (B) *CnMut*RubisCO (1bxn structure with mutations in loops 1 to 3) MEP based on gas phase AM1 semiempirical calculation with highlighted mutated loops location. Colour range is set from -100 to 50 kcal/mol and the isovalue is set to 0.001 e/ Å^3.

4.2.2 Molecular modelling of *H. naepolitanus* and *C. necator* H16 wild type and mutant RubisCO

Molecular dynamic simulations of *Hn*RubisCO, *Cn*RubisCO and C*nMut*RubisCO dimers were carried out in aqueous solution at neutral pH 7 and at three ionic strengths 20 mM, 150 mM and 1000 mM NaCl. The simulations were performed at 150 mM as it represents the physiological ionic strength at which the majority of biologically relevant systems are found active (Alberty, 2003). Since phase separation recently emerged as a driving force for the organisation of RubisCO in some carbon concentrating mechanisms, such as algal pyrenoid (Freeman-Rosenzweig *et al.*, 2017; Wunder *et al.*, 2018) and β-carboxysomes (Long *et al.*, 2010; Wang *et al.*, 2019), the behaviour of the three proteins were investigated at low salt concentration (20 mM). Finally, the ionic strength of 1000 mM was chosen to explore the stability of the halophilic *Hn*RubisCO at extremely high salt concentration and compare these results with the *Cn* and *CnMut*RubisCO.

Importantly, the sequence of *Hn* and *Cn* RubisCO are not directly comparable because of the difference in protein length. Specifically, the *H. neapolitanus* CbbL and CbbS amino acidic sequences are 13 and 29 amino acids shorter than the *C. necator*, respectively. In particular, the CbbS subunits from red-type RubisCO, such as *Cn*RubisCO, present two additional β-strands at the C-terminal that are not featured in the green-type RubisCO, such as *Hn*RubisCO (Bracher *et al.*, 2017). When comparing conserved residues, direct alignment of the two amino acid sequences was used to identify correct and comparable residue numbers. In this case, there was an overhang of 9 amino acids of the *C. necator* sequence at the start of the alignment for the CbbL subunits and 8 amino acid overhang of the *H. naepolitanus* sequence for the CbbS subunits (MUSCLE Pairwise Alignment; data not shown). Moreover, the first residue determined in the crystal structure for the *Hn*RubisCO is residue 12,

while for the *Cn*RubisCO is residue 25. Therefore, in all the following analyses these differences were accounted for, especially when comparing regions between the proteins and residues numbers.

Root mean square deviation (RMSD) is used for quantifying the difference between the backbones of a protein from its initial structural conformation to its final position. The stability of the protein relative to its conformation can be determined by the deviations produced during the course of its simulation. The smaller the deviations, the more stable the protein structure. In our case, evaluation of the differences in RMSD between the three protein backbones (*Hn, Cn* and *CnMut*RubisCO) at different ionic strenghts (20, 150 and 1000 mM) were carried out.

The RMSD analysis showed an overall higher fluctuation of the *Hn*RubisCO at 20 mM compared to the *Cn* and *CnMut* RubisCO, with a higher deviation in the first 300 ns of simulation in both replicas and then stabilising at values of about 2.5 Å (Figure 4.4). In contrast, the *Cn*RubisCO has an increased fluctuation of the protein backbone at 150 mM, reaching an average of 2.5 Å shortly after 250 ns of simulation. Surprisingly, the *CnMut*RubisCO did not behave as either the wild type enzyme or the *Hn*RubisCO, rather showing a more rigid structure with a deviation from the X-ray reference structure slightly over 2 Å for the simulation at both 20 and 150 mM of NaCl. Regarding the ionic strength at 1000 mM, all the three proteins showed low average deviation during the 550 ns of simulation, reaching a stable value of around 2 Å after the firsts 100 ns.

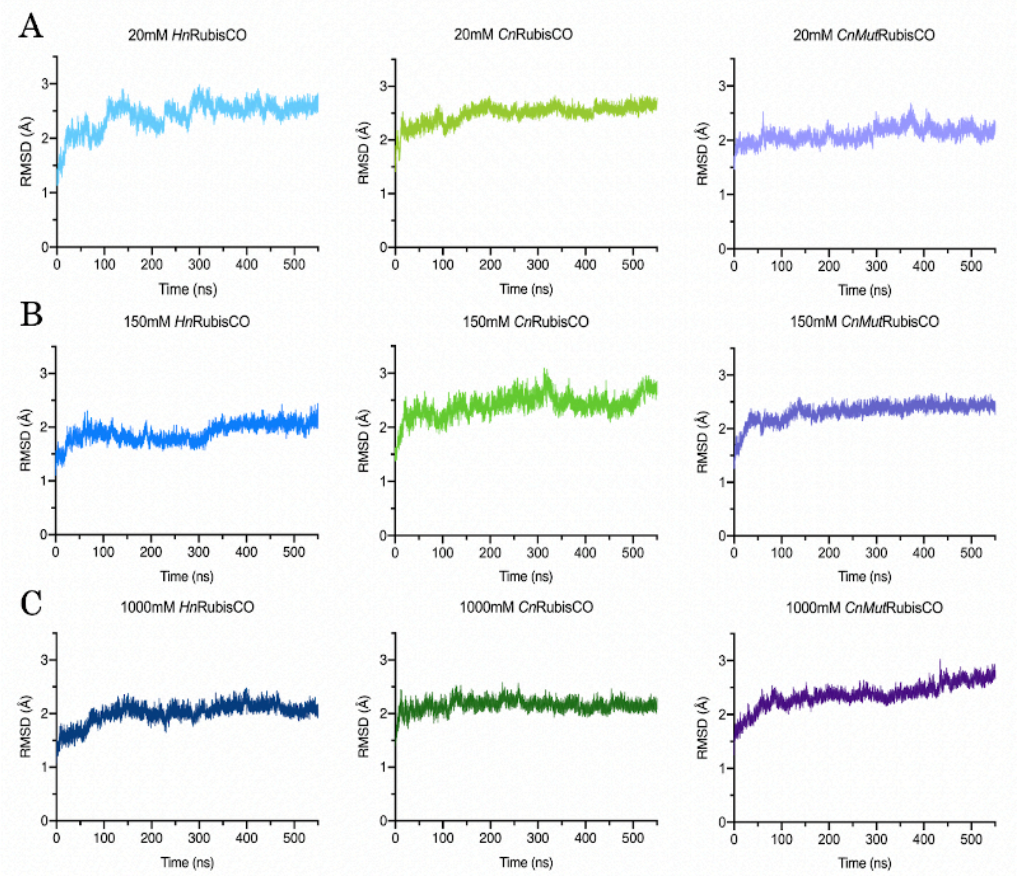


Figure 4.4: RMSD analysis of *Hn*, *Cn* and *CnMut*RubisCO at three ionic strengths (20mM, 150mM and 1000mM). Graphs represent Root Means Square Deviation (RMSD) of the backbone (N, Cα and C) of *Hn* (blue), *Cn* (green) and *CnMut* (purple) RubisCO at (A) 20mM, (B) 150mM and (C) 1000mM, during two independent molecular dynamics simulation replicas (550 ns each) using the crystal structure as reference (displayed mean between the replica).

Subsequently, to gain knowledge about protein regions displaying higher levels of flexibility, the residue-based root-mean square deviation (known as: Root Mean Square Fluctuation; RMSF) against the average position of the protein residues was performed using the same three dimeric protein backbones (*Hn, Cn* and *CnMut*RubisCO) at different ionic strenghts (20, 150 and 1000 mM).

A similar trend to RMSD was observed during RMSF analysis in which the *Hn*RubisCO residues from 57 to 71 were responsible for the majority of the movement (around 7.5 Å) at 20 mM (Figure 4.5). The fluctuation of these residues is found to be decreased below 3 Å at 150 mM and increased again up to 4.4 Å at 1000 mM. While at 150 mM a sharp peak around residue 295 up to 3.8 Å was also observed. For the *Cn*RubisCO, a high mobility of the residues between 68 and 80 was

observed at both 20 and 150 mM, with an RMSF value above 8 Å at 150 mM. The height of the peak was then halved when the ionic strength was increased to 1000 mM. For *CnMut*RubisCO, the highest fluctuations between residues 70 and 83 was observed, reaching up to 5 Å. This feature has a striking similarity to the other two structures; however, no significant changes in the intensity of the peak between the three ionic strengths was seen. When the ionic strength went from 20 mM to 150 mM, an increase in the peak intensity was detected corresponding to residues between 209 and 214, and 300 to 307 reaching a value of 5 Å.

Figure 4.5 confirms that the region after residue 570 (residue 110 of the small subunit) belonging to the small RubisCO subunit, was not present in the *Hn* simulation, while the structure of *Cn*RubisCO extended until residue 596 (residue 129 of the small subunit). This was due to the difference in length between the small subunits of the two variants. As mentioned before, the CbbS from *H. naepolitanus* is 110 amino acids long. In contrast, the CbbS from *C. necator* is 139 amino acids long, 10 of whose position were not determined in the crystal structure. This is possible due to the high fluctuation and movement of the CbbS subunit in its C-terminus. This increased movement was captured, indeed, during our simulations of *Cn*RubisCO, where the last 20 residues proved to be highly mobile at all ionic strengths (Figure 4.5).

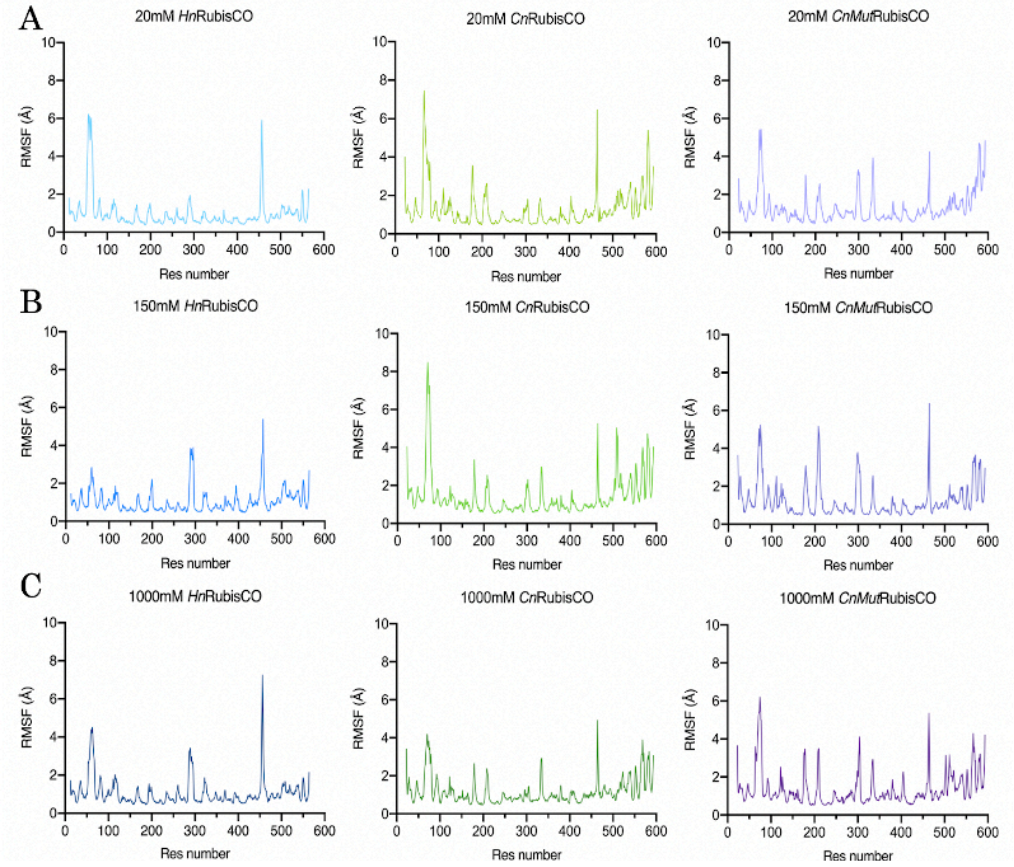


Figure 4.5: RMSF analysis of *Hn*, *Cn* and *CnMut*RubisCO at three ionic strengths (20mM, 150mM and 1000mM). Graphs represent Root Mean Square Fluctuation (RMSF) of *Hn* (blue), *Cn* (green) and *CnMut* (violet) RubisCO at (A) 20mM, (B) 150mM and (C) 1000mM. The analysis was carried out using resulting trajectories from 550 ns of molecular dynamics simulation of two independent replica for each condition using the crystal structure as a reference (displayed mean between replica). *Hn*RubisCO: CbbL (Residue number 16–460) CbbS (Residue number 461–570); *Cn*RubisCO and *CnMut*RubisCO: CbbL (Residue number 25–467) CbbS (Residue number 468–596).

Interestingly, when the highly flexible loop was highlighted in the structure, it was localised in a similar position in the three RubisCO proteins despite the difference in residues numbering. To help the visualisation of the major loop movement in the three structures (residues 57 to 71 for *Hn*RubisCO; residues 67 to 82 for *Cn*RubisCO and *CnMut*RubisCO) Figure 4.6 was created to display a series of conformational changes of the proteins during MD simulation. The *Hn*RubisCO and *CnMut*RubisCO were represented during the simulation at 20 mM, while the *Cn*RubisCO was displayed during the simulation at 150 mM, where the samples showed an increased mobility during the performed MD simulations. The residues involved in the movements were highlighted in orange in Figure 4.6,

displaying the high mobility and flexibility of the protein region during the MD simulation for both *Hn* and *Cn*RubisCO. In contrast, the *CnMut*RubisCO loop appeared to be in a more rigid conformation with a less intense degree of movements, spending the majority of the simulation time between the conformations represented in the second and third frame of the Figure 4.6 C. This behaviour could be the result of the introduced mutation in residues 76 and 79 (Arg76Leu; Ala79Asp) positioned on the same region.

Examination of the loop position in the three-dimensional conformation of the hexadecameric structure (reference structure *Hn*RubisCO from Oltrogge *et al.*, 2020; Figure 4.7 A) and of the residues at the interface between the two catalytic subunits, revealed that this location allows for a wide range of movement *in situ* (view from left and right side Figure 4.7 C and D, respectively). This was particularly relevant since the residues near the flexible region are involved in the catalytic activity of the protein.

Comparing the overall movements and fluctuations of *Hn*RubisCO and *Cn*RubisCO based on the RMSD and RMSF analysis against *CnMut*RubisCO, no significant resemblance in neither residue fluctuations nor protein stability was found.

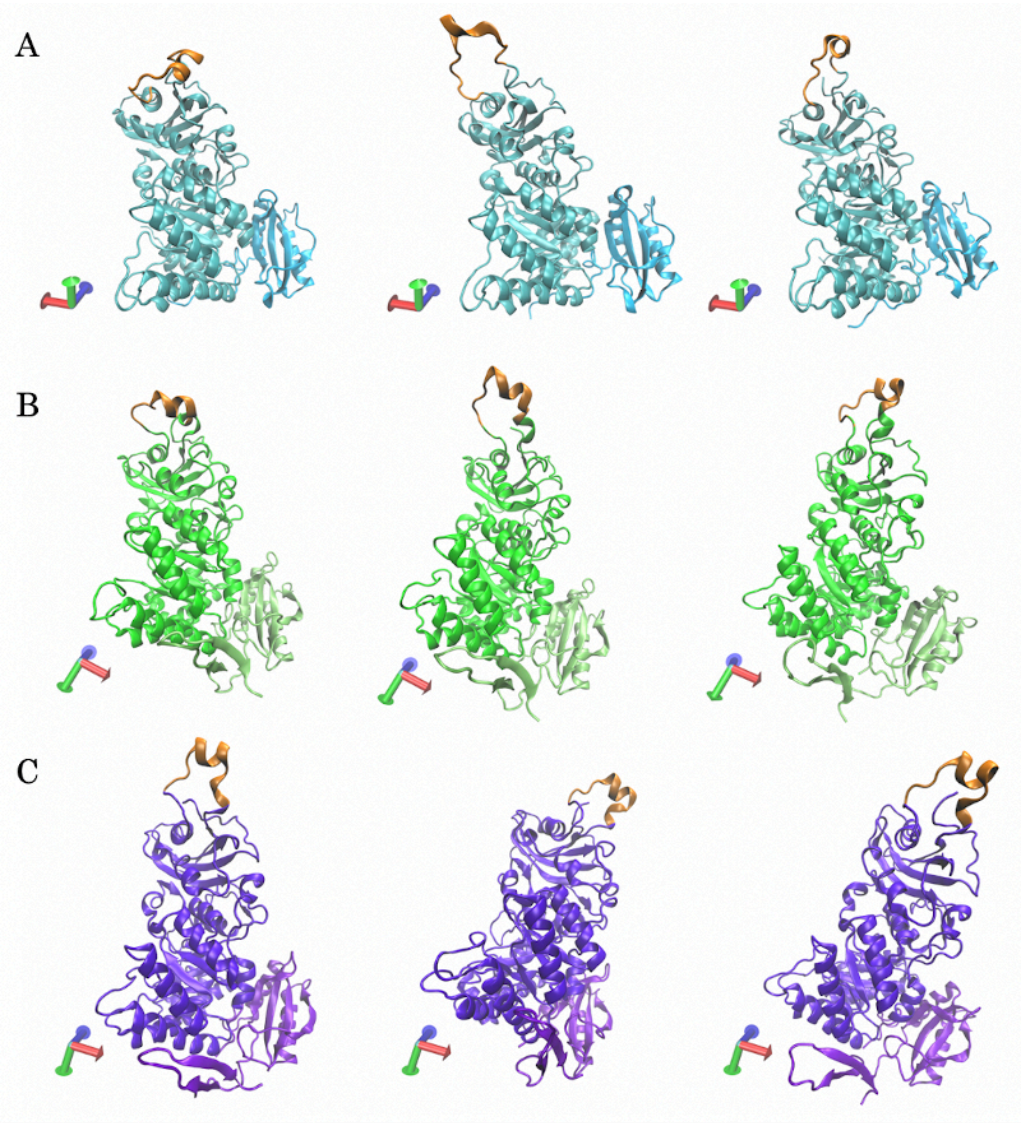


Figure 4.6: Protein structure of *Hn*, *Cn* and *CnMut*RubisCO showing loop flexibility. Figure representing loop movement for (A) *Hn*RubisCO in blue with residues 57 –71 highlighted in orange from MD simulation at 20 mM; (B) *Cn*RubisCO in green with residues 68– 0 highlighted in orange from MD simulation at 150 mM and (C) *CnMut*RubisCO in purple with residues 70–83 highlighted in orange from MD simulation at 20 mM.

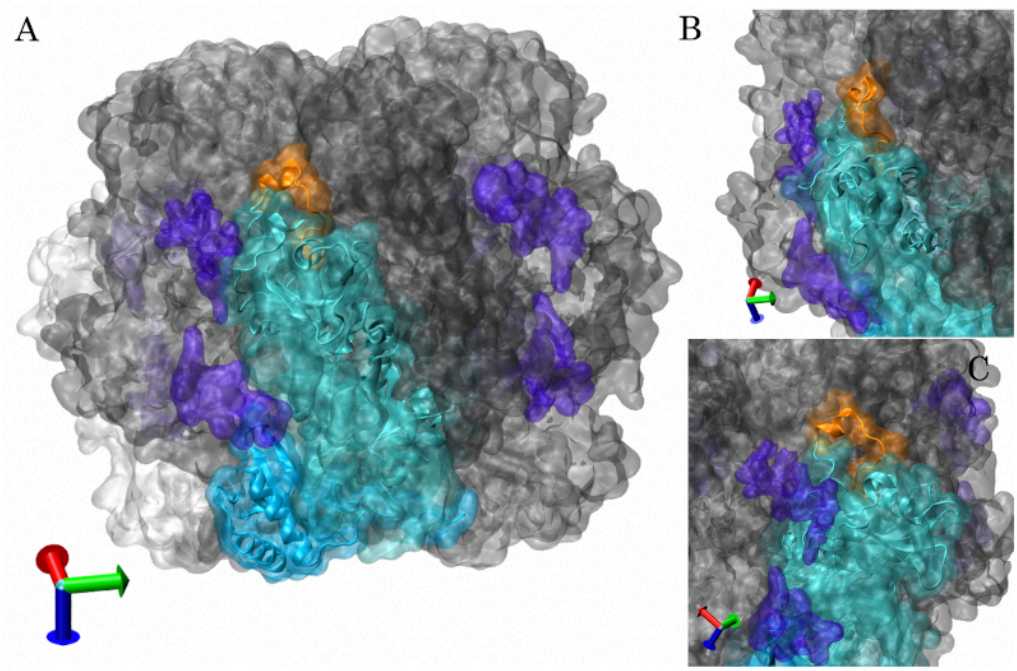


Figure 4.7: Hexadecameric structure of *Hn*RubisCO showing CbbL flexible loop location. (A) Figure represents *Hn*RubisCO (PDB 6uew) where the loop position (in orange) in the RubisCO dimer (large subunit in turquoise; small subunit in blue) is shown as part of the hexadecameric structure (in grey). View rotating 90º (B) to the right and (C) to the left.

To further investigate the interaction between the environment and the RubisCOs (*Hn*, *Cn* and *CnMut*), the distribution of cations (Na^+) and anions (Cl^-) around positively charged (arginine and lysine) and negatively charged (aspartic acid and glutamic acid) residues of the RubisCO proteins was calculated. Radial distribution function (RDF) analysis enabled us to observe the probability distribution of ions at different ionic strengths around the reference residues as a function of radial separation between the two.

The RDFs at 20 mM between anion and Lys (Figure 4.8 A) showed a first peak at 2.5 Å for both *Hn* and *CnMut*RubisCO. The maxima were observed at 5.5 Å for both RubisCOs, with *CnMut*RubiCO reaching double the values of *Hn*RubisCO. Instead, the *Cn*RubisCO's Lys residues interaction with anions showed four distinct peaks at 2.5, 4.3, 6.5 and 8.3 Å, with higher intensities than the ones of *CnMut*RubiCO and *Hn*RubisCO. The interaction between Arg residues on the protein backbone and anions had a similar pattern, again for both *Hn* and *CnMut*RubisCO, showing a small peak at 2.5 Å and then a mild increase up to a plateau at 5 Å. In comparison, *Cn*RubisCO's arginine interactions with Cl^- showed an intense peak at 2.5 Å followed by two double peaks at 4 and 6 Å. With an increase in the ionic strength, the interaction intensity between anions and positively charged residues exhibited an overall reduction for all three proteins. At 150 mM the pattern of interaction of both positively charged residues (Arg and Lys) with the anions appeared to be strikingly similar with regards of shape, intensity and distance distribution between *Hn* and *CnMut*RubisCO (Figure 4.8 B). Both RDFs conserved the first small peak at 2.5 Å, and then slowly increasing up to a plateau starting at 5.5 Å. Overall, a reduction of the interaction intensities to half between Lys residues and anions for *Hn*RubisCO's and up to a third for *CnMut*RubisCO was observed. Surprisingly the interactions between Arg and anions were only mildly affected for both proteins. At 150 mM the *Cn*RubisCO

maintained the four peaks of interaction between Cl^- and Lys, with a reduced intensity to about a fourth of the previous values. The three peaks recorded between the anions and Arg were highly impacted as well, reducing its intensity to a fifth of the previous numbers. At 1000 mM, it was recorded a further reduction in interaction intensities (Figure 4.8 C), where the radial distribution functions of three proteins showed a high resemblance between each other for both positively charged residues and anions. The only exception being *Hn*RubisCO's arginines and anions interaction which appeared to be less intense than the other two cases.

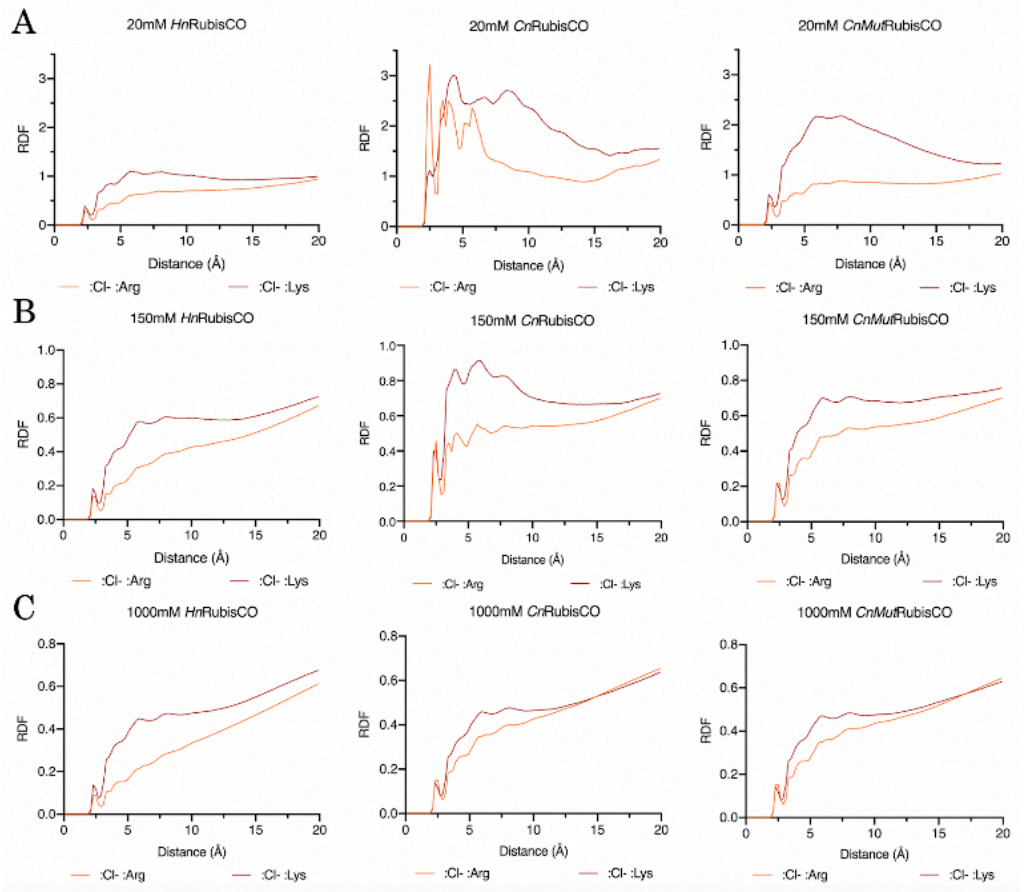


Figure 4.8: RDF analysis between Cl^- and positively charged amino acids of *Hn*, *Cn* and *CnMut*RubisCO in three ionic strengths (20mM, 150mM and 1000mM). Graphs represent intermolecular Radial Distribution Function (RDF) between anion (Cl^-) and positively charged ammino acids arginine (orange) and lysine (red) at (A) 20mM, (B) 150mM and (C) 1000mM, following 550 ns molecular dynamics simulation of two independent replica for each condition.

The RDFs between the negatively charged residues (Asp and Glu) of the three proteins and cations all showed the first shell of interaction at 2.3 Å and a second shell at 4.7 Å (Figure 4.9), where the intensity of the peaks broadly differed depending on the ionic strength. For *CnMut*RubisCO at 20 mM, the negatively charged residues showed an interaction shell with the cations of similar intensity as *H. naepolitanus*'s Asp at both distances (2.3 and 4.7 Å). While an intense peak for the interactions between Asp and cations, and the second peak of lower intensity at 4.7 Å was seen for *Hn*RubisCO (Figure 4.9 A), with interactions between Glu and cations generating peaks at the same distances but halved in size. In contrast, at the same ionic strength, the *Cn*RubisCO, showed minimal interaction between cations and Asp residues and a defined interaction shell with Glu, with the higher first peak at 2.3 Å and the second peak at 4.7 Å.

As observed with the anions, the overall intensity of interactions between ions and the protein backbone decreased when the ionic strength was brought to 150 mM (Figure 4.9 B). This is probably due to the stronger electrostatic screening effect in the higher salt concentrations. Since the binding is driven by the attractive electrostatic interaction, higher salt concentrations tend to weaken the interaction (Zhang *et al.*, 2011; Du *et al.*, 2016). A reduction to about half of the previous values was observed for *Hn* and *CnMut*RubisCO, while the reduction accounted for a quarter in the case of *Cn*RubisCO. The highest interaction intensities between both negatively charged residues and cations was observed for the *CnMut*RubisCO with values slightly higher than *Hn*RubisCO for the two interaction shells. In contrast, the intensity of the same peaks for *Cn*RubisCO barely reached a seventh of the intensity of *CnMut*RubisCO. At 1000 mM, again a further decrease in interaction intensities (Zhang *et al.*, 2011; Du *et al.*, 2016) was observed with halved values for *Hn* and *CnMut*RubisCO in both interaction shells, but an increase to tripled values in intensity for *Cn*RubisCO (Figure

4.9 C). At high ionic strength, the plots were very similar between each other and almost impossible to distinguish based on the intensities and location of the peaks.

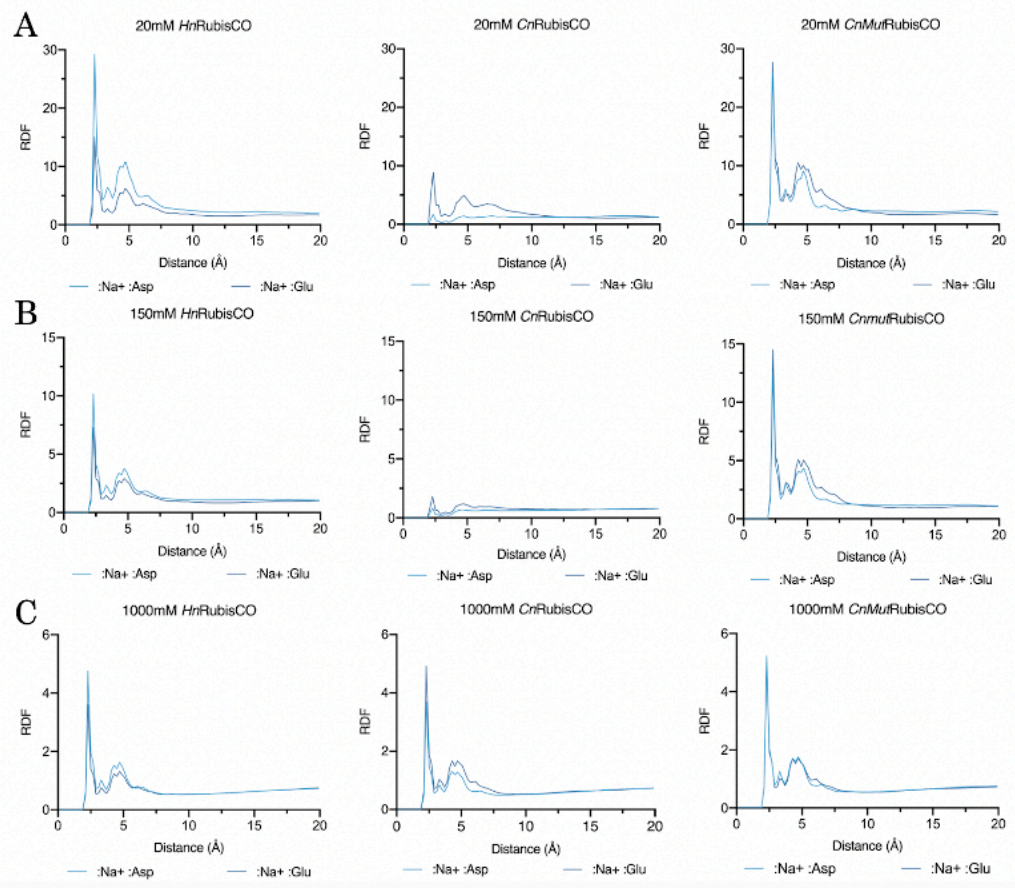


Figure 4.9: RDF analysis between Na^+ and negatively charged amino acids of *Hn*, *Cn* and *CnMut*RubisCO in three ionic strengths (20mM, 150mM and 1000mM). Graphs represent intermolecular Radial Distribution Function (RDF) between cations (Na^+) and negatively charged amino acids aspartic acid (light blue) and glutamic acid (dark blue) at (A) 20mM, (B) 150mM and (C) 1000mM, following 550 ns molecular dynamics simulation of two independent replica for each condition.

As an overall observation, a significant difference in interaction intensities between positively and negatively charged residues with the respective counter-ions was observed and could be caused by the nature of the residues themselves, where the Asp and Glu residues can coordinate cations in a structured way via their carboxylate groups in comparison to Lys and Arg.

From this series of RDF analyses, a remarkable resemblance between *Hn* and *CnMut*RubisCO interaction pattern was observed with both cations and anions, while a large difference could be seen comparing both *Hn* and *CnMut*RubisCO interactions to *Cn*RubisCO.

To have a better understanding of the occupation of ions in three-dimensional space around the three RubisCO proteins at different ionic strengths, the ion density analysis (IDA) was performed. The density map helped to visualise the localisation of the ions during the simulation, highlighting the most likely positions for interaction with the protein structures. At 20 mM a high Na^+ density (in blue) was observed at the top of the structure at either side of the flexible loop (residue 57 to 71) of *Hn*RubisCO (Figure 4.10 A), extending down in a bridge towards the bottom of the large subunit and connecting with the density cloud to the left of the small subunit. By analysing the location of the ions in more detail, a high Na^+ density was observed in the predicted location for interactions between the protein backbone and the CsoS2 recruiting peptides (Oltrogge *et al.*, 2020) (Figure 4.10 A). In fact, a high-density cloud at the right side of the flexible loop was observed around residues 22–28, 69–72 at the top of the structure, where residues Asp26, Asp69 and Tyr72 were highlighted as interacting with the recruiting peptides (Figure 4.10 A; highlighted with purple squares in the first and second frame). In addition, by rotating the structure 90º to the right (Figure 4.10 A, third frame) another cations density cloud was observed around the large subunit protein backbone forming two β-sheets (residues 346–349 and 358–363) and around the small subunit forming a loop (residues 94–98), where the other interactions between the RubisCO and the CsoS2 peptides were described to occur (large subunit residues Phe346 and Asp360 and small subunit Asp94 and Tyr96). From the simulation analysis, two other locations with high Na^+ ion density were highlighted, one located at the left side of the flexible loop (Figure 4.10 A, highlighted with yellow squares in the first and fourth frame), around protein residues 16–18, 42–45 and residues 82 and 85 on the loop. In contrast, the other was positioned towards the lower part of the large subunit near the three α-helices (residues 219–224, 251–253 and 279–280) and bridging across towards the small

subunit loop (residues 49–54; Figure 4.10 A, highlighted with a black square). A similar, distribution of the Na^+ ions was observed during the *CnMut*RubisCO simulation at 20 mM (Figure 4.10 C), but unfortunately, the bridge between the upper and lower ions density clouds was almost completely missing. This lack of negative electrostatic potential might hinder the interaction and recruitment of the *CnMut*RubisCO by the CsoS2 peptides. It was observed also a high similarity in the Na^+ density located at the right side of the flexible loop residues 70–83 (Figure 4.10 C, highlighted with purple squares in the first and second frame), around residues 35–37, 77–83, where the residues Asp36, Asp79 and Try83 could provide the interactions with CsoS2 peptides required for the recruitment inside of the carboxysomes. Interestingly, in MD simulation of *CnMut*RubisCO a very localised and intense cation density was noticed around the large subunit residues 355–357 and 367–370 forming two β-sheets and near the loop of the small subunit (residues 86–87), where the ion bridge was observed in the *Hn*RubisCO protein (Figure 10 C, highlighted with purple square in the third frame). Also, a highly dense cloud was obtained on the left side of the flexible loop (Figure 4.10 C, fhighlighted with yellow square in the first and fourth frame) around the first crystallised residues of the *CnMu*t structure and residues 50–52, 89–92, which in this case merges with the ion cloud originally on the right side of the structure. Overall, in this second interaction location, only three out of four residues highlighted in recent research as interacting with CsoS2 were found (large subunit Tyr356, Asp370 and small subunit Asp86). Moreover, a cation cloud was noticed at the bottom of the large subunit (residues 288–290) and bridging across to the small subunit around residues 39–47 and at the C-terminal similar to *Hn* (Figure 4.10 C, highlighted with black square). Unfortunately, it lacks the larger ion density redistribution across the three α-helixes previously observed in *Hn*RubisCO simulation. In contrast with the simulations of the

previous structures, for *Cn*RubisCO at 20 mM an overall rarefied Na^+ density cloud was observed. It was mainly localised around two positions of the protein structure: one at the top of the large subunit around residues 109–12 (Figure 4.10 B, first frame) and the other one across the large subunit α-helix (residues 150–165) and the small subunit loop (residues 38–47).

The anions density map (Figure 4.10; in red) at 20 mM showed the main cloud located the central groove of the protein structure between two loops for all three structures, with a particularly high density around loops residues 207, 298–303 and 381–383 for both *Cn* and *CnMut*RubisCO. The *Hn*RubisCO density map at the same ionic strength was slightly extended diagonally across the structure between loop residues 331–324 and 372–375 and towards the bottom of the large subunit (residues 166–168).

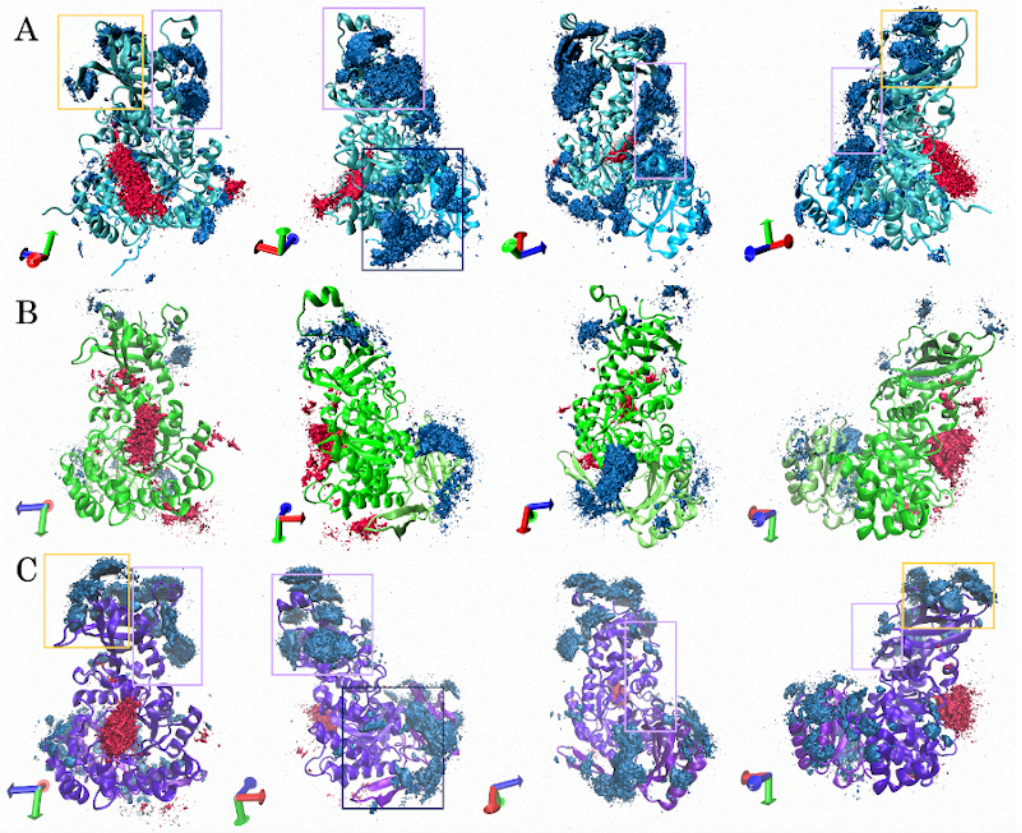


Figure 4.10: Ion Density Analysis of anions and cations during MD simulation of *Hn*, *Cn* and *CnMut*RubisCO at 20 mM. Image representing ion density distribution of Na^+ (blue) and Cl^- (red) around (A) *Hn*, (B) *Cn* and (C) *CnMut*RubisCO structure following 550 ns molecular dynamics simulation of two independent replica for each condition at 20 mM. Isovalue: 4 e/ $Å^3$.

During this analysis, it was observed less intense interactions when the ionic strength was increased to 150 mM (Figure 4.11), with a slight redistribution of the anions and cations on a wider protein surface. For the *Hn* no drastic changes in the Na^+ density cloud were observed, only a thicker volume around the areas highlighted at 20 mM at the top of the protein structure across the bridge and at the left side of the small subunit. Similar, to *Hn* the *CnMutRubisCO* Na^+ density map showed a wider cloud especially at the top of the large subunit and around the small subunit, while in the bridge location the density around residues 355–370 became a bit wider. The only new density area observed on the ions' map for both *Hn* and *CnMut* structures were around the α-helix at the bottom left of the large subunit (Figure 4.11 A, highlighted with green squares in the first and fourth frames; *Hn* residues 427–436 and 450–454; *CnMut* residues 428–445 and 462–467). The *Cn*RubisCO showed an increase in density around the areas highlighted at 20 mM, emphasising again a tangible difference from the *Hn* and *CnMut* samples.

The Cl^- ions density map of the *Hn*RubisCO at 150 mM didn't change in shape dramatically compared against the 20 mM. It expanded upward into the right side of the protein groove towards residues 331–335 (Figure 4.11 A; highlighted with orange square) and showed a further density cloud around residues 203 and 206 (highlighted with magenta square). While the anions density of *Cn* and *CnMut*RubisCO underwent a wider redistribution. It expanded diagonally towards the left side of the groove enveloping the protein residues 298–305 and 312–313 forming an α-helix and a loop (Figure 4.11 C; highlighted with orange square). Moreover, two new density clouds were recorded one under the right side of the large subunit structure (residues 225–227, highlighted with magenta square) and the other one towards the far right (residues 223 and 234, highlighted with grey square).

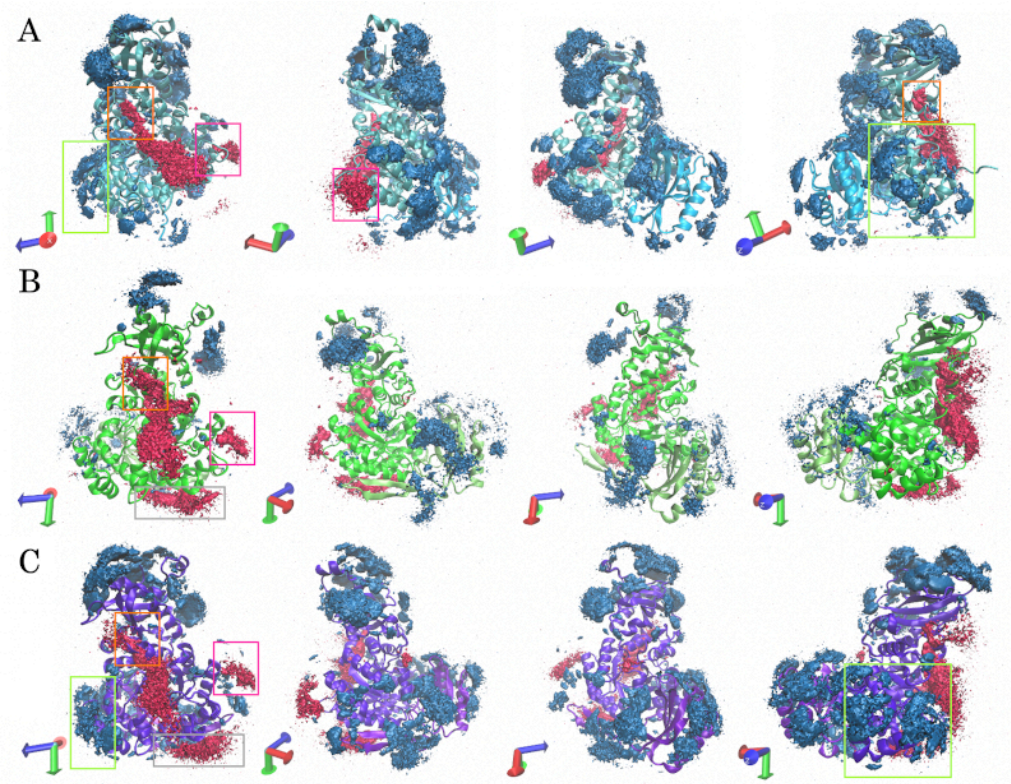


Figure 4.11: Ion Density Analysis of anions and cations during MD simulation of *Hn*, *Cn* and *CnMut*RubisCO at 150 mM. Image representing ion density distribution of Na^+ (blue) and Cl^- (red) around (A) *Hn*, (B) *Cn* and (C) *CnMut*RubisCO structure following 550 ns molecular dynamics simulation of two independent replica for each condition at 150 mM. Isovalue: 7 e/ Å^3.

The increase of the ionic strength to 1000 mM caused a further decrease in the interaction intensity (Appendix 6) and an overall redistribution of the Na^+ density of *Cn*RubisCO to resemble *CnMut* and *Hn*RubisCO density maps. While the Cl^- map for *Hn*RubisCO split into two parts one near the top right of the groove (residue 335) and the other one towards the bottom of the large subunit between residue 166–172 and residue 203–206. The Cl^- density clouds for both *Cn* and *CnMut* remained in the three locations observed at 150 mM: the central groove around the α-helix (residues 298–305) at the bottom and at the right side of the large subunit.

Overall, this data well reflects the interaction intensities previously observed during RDF analysis giving an insight into the similarities between *Hn* and *CnMut* ion density during the simulation.

To understand if the density maps were representative of the interactions between ions and the functionally active holoenzyme in solution, the interaction interface between dimers was analysed including the solvent-accessible surface area of the enzyme and the peptide interaction surface, using the *Hn*RubisCO hexadecameric crystallographic structure from Oltrogge *et al.* study as a reference (Figure 4.12). The posituve density clouds in both *Hn* and *CnMut* IDA analysis compared against the residue locations in the hexadecameric structure were localised near regions involved in the interaction with recruitment peptides, exposed to the solvent and therefore available for interaction with CsoS2 (Figure 4.12 A and B, purple squares) Furthermore, the Na^+ ion density located at the left and right of the flexible loop in *Hn* and *CnMut* structure showed a reliable interaction with the solvent even when incorporated in the hexadecameric structure (Figure 4.12; yellow square). The only exception was represented by the density near residues 101–104 for *Hn*RubisCO and 109–112 for *CnMut*RubisCO since they were in interaction with the adjacent dimer. Additionally, the Na^+ density observed in both *Hn* and *CnMut*RubisCO structures at 20 mM located at the bottom of the large subunit bridging across towards the small subunit were positioned in the hexadecameric structure inside of the inner channel formed during the structure assembly (Figure 4.12; dark blue square). Therefore, the residues were solvent esposed and the ion density would be considered reliable. Interestingly, the Cl^- ion density at both 20 mM and 150 mM for all the proteins was localised near the dimer interface with the other large subunit at the edge of the catalytic site of the RubisCO (Figure 4.12; orange square). These residues positioned in the hexadecameric structure, exposed to the solvent and near to the catalytic site might hold a strategic advantage to attract anions from the solution. Unfortunately, the Cl^- density predictions nearby residues 203 and 206 for *Hn* and 249–251 for *Cn* and *CnMut*RubisCO were not realistic because these residues are

embedded in the structure and in contact with the large subunit of the catalytic dimer.

Overall, the use the MD simulation and IDA of the dimers produced a realistic depiction of the possible interactions between ions and RubisCO proteins.

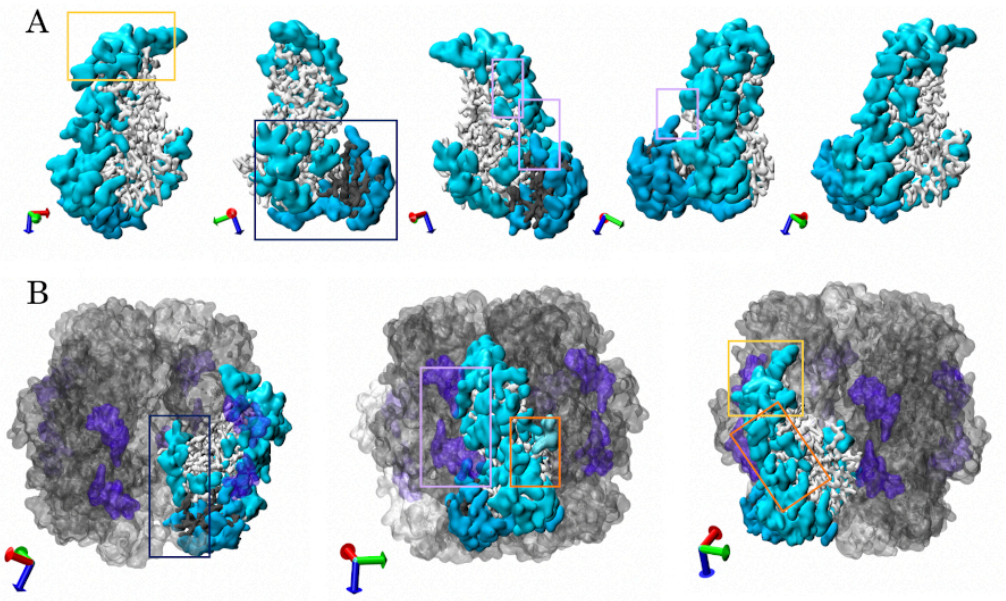


Figure 4.12: Solvent accessible and interaction surface of *Hn*RubisCO in the hexadecameric structure. (A) Image representing the solvent accessible surface area for the large (turquoise) and small (blue) subunits of *Hn*RubisCO when in the hexadecameric structure. The interaction surface with other subunits is depicted in white for the large subunit and in grey for the small subunit. (B) Hexadecameric structure of *Hn*RubisCO with CbbLS dimer solvent accessible surface for the large subunit in turquoise and for the small subunit in blue. The surface of the other subunits of the structure are in grey while the CsoS2 peptides are in in purple.

4.2.3 Functional complementation of *C. necator* ΔΔ*cbbLS* mutant by *CnMut*RubisCO in autotrophic conditions

Expression and correct folding of the *CnMut*RubisCO represents a major concern for the formation of the catalytically active holoenzyme in *C. necator* H16. For this reason, to confirm the *in vivo* activity of RubisCO after insertion of the mutations in the *cbbL* gene, a *C. necator* mutant strain, lacking both copies of the RubisCO-encoding genes (ΔΔ*cbbLS)*, therefore unable to fix CO_2, was used.

To assess the ability of *CnMut*RubisCO to complement the ΔΔ*cbbLS* mutant, a synthetic *cbbL* gene carrying the mutations required in the large subunit of *Cn*RubisCO was ordered from GeneScript® and cloned in the pMTL71301 modular vector, upstream the small subunit of *Cn*RubisCO under the control of the constitutive promoter P_{phaC}. The construct was then transformed in *C. necator* ΔΔ*cbbLS* strain and three independent biological clones of the resulting strain were grown overnight either in heterotrophic conditions (0.4% SG MM) or in organo-autotrophic conditions (80 mM Na-formate in presence of 10% CO_2). Subsequently, they were washed in PBS before inoculation in the serum bottles. The adaptation experiment was carried out in autotrophic conditions (MM, 60% H_2: 20% CO_2: 20% air v/v/v) in serum bottles for 16 days. In this experiment, *C. necator* H16 wild type and its isogenic ΔΔ*cbbLS* mutant were used as the positive and negative control, respectively. All the biological clones carrying *CnMut*RubisCO were able to restore the growth of *C. necator* ΔΔ*cbbLS* strain after two days in autotrophic conditions (Figure 4.13). Samples grown overnight in heterotrophic conditions showed a shorter lag phase compared to the other samples. This trend was also visible in the positive control *C. necator* H16. The delay in growth in the samples form overnights in organo-autotrophic conditions could be due to the lower energy and reduction power offered by the Na-formate media conditions. To be sure that no extra mutations

occurred during the adaptation period, *cbbLS* genes from *CnMut*RubisCO samples were sequenced at the end of the experiment. The results confirmed that the plasmid-born RubisCO sequence wasn't affected during the experiment.

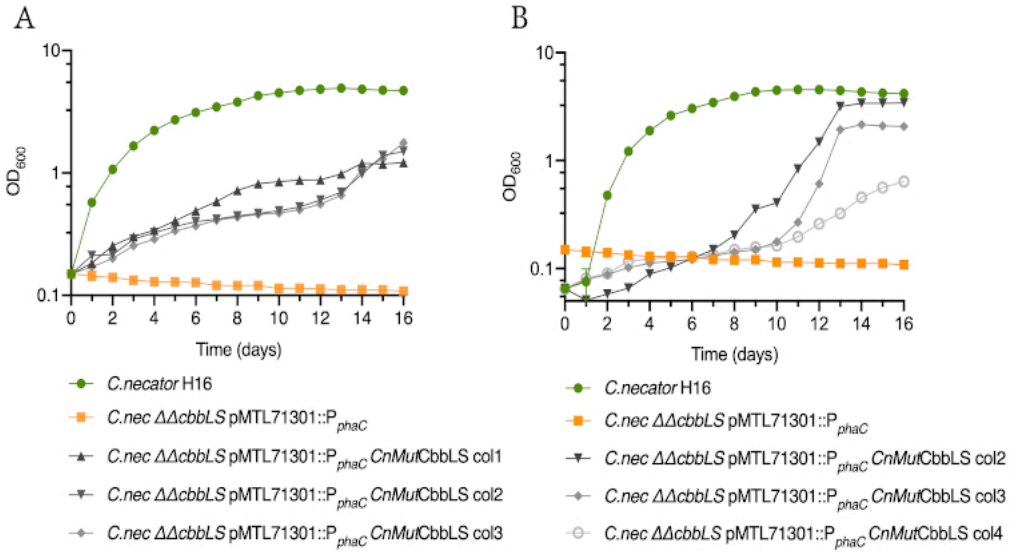


Figure 4.13: Adaptation experiment of *C. necator* ΔΔ*cbbLS* mutant strain carrying *CnMut*RubisCO in autotrophic conditions. Complementation of *C. necator* ΔΔ*cbbLS* deletion strain with constitutively expressed *CnMut*RubisCO genes (P_{phaC} *CnMut*CbbLS), under autotrophic conditions (MM, 60% H_2: 20% CO_2: 20% air v/v/v) after overnight in (A) 0.4% SG MM or (B) in 80mM Na-formate in presence of 10% CO_2.

The adapted strains were then grown again under the same experimental conditions in serum bottles for 9 days. Also in this case, the strains carrying the *CnMut*RubisCO showed complementation of the mutant *C. necator* ΔΔ*cbbLS* strain (Figure 4.14), with a doubling time of 57 h compared to the 55 h of the *C. necator* H16. The strains after adaptation showed also a μ_{max} of 0.0122 h^{-1} very similar to the wild type 0.0127 h^{-1} value, with no statistical difference at the onset of the stationary phase (7 days; data not shown). A genomic analysis of the strain was carried out at the end of the experiment to determine if any genomic mutations occurred that might be responsible for the phenotype observed. The NGS data didn't show any permutation in bases or rearrangements in the genome that might change the performance of the strain in autotrophic conditions. This might suggest an epigenetic adaptation to the culturing conditions.

Altogether, this data strongly suggests that the *CnMut*RubisCO on the plasmid was responsible for the growth complementation and its catalytic activity was not affected by the mutations in the flexible loops of the large subunit.

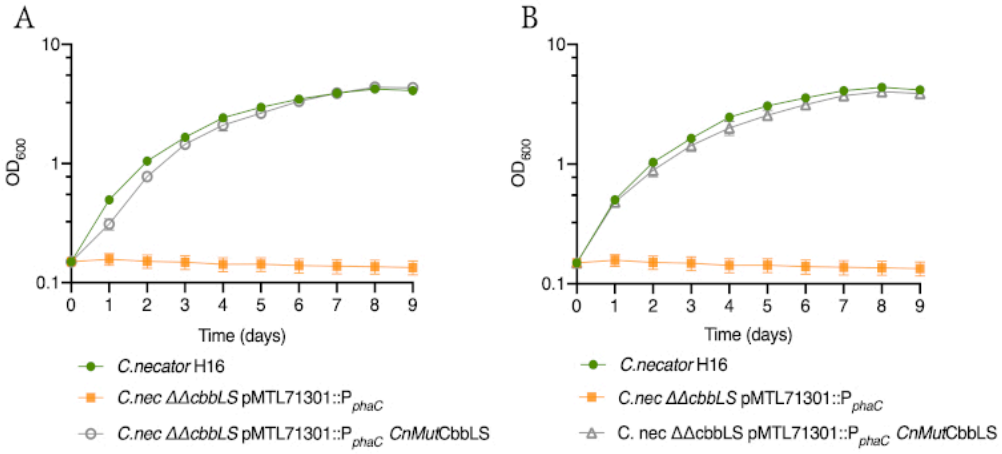


Figure 4.14: Complementation experiment of *CnMut*RubisCO in *C. necator* Δ*cbbLS* mutant strain. Complementation of *C. necator* Δ*cbbLS* deletion strain with constitutively expressed *CnMut*RubisCO genes (P_{phaC} *CnMut*CbbLS), under autotrophic conditions (MM, 60% H_2: 20% CO_2: 20% air v/v/v) original overnight condition in (A) 0.4% SG MM or (B) 80mM Na-formate in presence of 10% CO_2.

4.3 Discussion

This chapter aim was to explore the relevance of the RubisCO electrostatic potential as a driving force for the recruitment of the holoenzyme inside of the carboxysome shells. After the initiation of our study a recent paper by Oltrogge *et al.* 2020, brought to light the interaction sites between the two proteins down to the amino acidic residues, emphasising the importance of a multivalent interaction between the RubisCO holoenzyme and the repeated peptide motifs in the N-terminal region of CsoS2. Here, it was confirmed that the contact interface was predominantly composed of electrostatic interactions. Furthermore, the authors suggested the involvement of several processes occurring simultaneously during the assembly process, like multivalency of interaction, surface avidity of the nascent shell proteins and liquid-liquid phase separation.

The investigation was started by analysing the surface charge of two RubisCO, one naturally localising into the carboxysomes (*Hn*RubisCO) and one found in the cytosol (*Cn*RubisCO). From the AM1 semi-empirical calculation in the gas phase, it was clear that the distribution of the surface charge on the RubisCO protein is very different between the two species. While the *Hn*RubisCO has mainly negatively charged surface at the top and on the backbone of the protein, the *Cn*RubisCO has an overall more neutral and negative surface charge (Figure 4.1). To test this hypothesis, it was decided to modify *Cn*RubisCO sequence to resemble the electrostatic potential of *Hn*RubisCO. The mutations were carefully chosen to avoid loops involved in both catalytic activity and dimer contact regions. The resulting *CnMut*RubisCO structure were subjected to AM1 semi-empirical calculation in the gas phase and its electrostatic potential showed a striking resemblance with the *Hn*RubisCO (Figure 4.3).

The RMSD and RMSF analysis of the MD trajectories suggested that at 20 mM the structure of *Hn*RubisCO was less stable than in the

other conditions (Figure 4.4 and 4.5). This behaviour could be possibly attributed to the dimer's tendency to bind another protein, be it another RubisCO dimer to form the catalytically active hexadecameric structure or to partner up with the recruiting peptides from CsoS2 (Singh *et al.*, 2016; Bhat *et al.*, 2017; Liu *et al.*, 2018). This ionic strength was also shown in a new study (Oltrogge *et al.*, 2020) to facilitate the recruitment between the *Hn*RubsiCO and the N-terminal of the CsoS2. In the same way, the high RMSD values and decreased stability of *Cn*RubisCO at 150 mM (representing the cytosolic ionic strength) could be referred to its cytosolic location in its native host and the conditions found within (Figure 4.4). Suggesting its binding to catalytic partners and RubisCO activases at 150 mM (Singh *et al.*, 2016; Bhat *et al.*, 2017; Bracher *et al.*, 2017).

Unfortunately, during RMSD and RMSF analysis, *CnMut*RubisCO did not exhibit decreased stability in any of the conditions analysed (Figure 4.4 and 4.5), which would have allowed us to evaluate and compare its behaviour with the reference proteins (*Cn*RubisCO and *Hn*RubisCO). The increased stability observed for the *CnMut*RubisCO, was attributed to the mutations on the *Cn*RubisCO structure into Leu76 and Asp79, which might have led to a restricted movement of the loop observed from the RMSF analysis at both 20 mM and 150 mM (Figure 4.5). To corroborate this hypothesis, the visualisation of the loop movement across the MD simulation confirmed the restricted mobility of the loop for *CnMut*RubisCO at both 20 mM and 150 mM (Figure 4.6). This result could raise concerns about its functionality, given the importance of the region flexibility in the catalytic activity of the holoenzyme (Andersson, 2008; Bhat *et al.*, 2017; Bracher *et al.*, 2017). Despite its different behaviour *in silico* the activity of the *CnMut*RubisCO remained intact as confirmed during the experiments *in vivo* (Figure 4.15).

The highly mobile loop observed in all the structures during the RMSF analysis was located near residues linked to substrate binding

during the catalytic activity of RubisCO. This is where the movement of the N-terminal (residues 56 to 63) of one large subunit is required to interact with the C-terminal of the second large subunit in the other dimer, which is essential in the catalytic activity of the holoenzyme (Andersson, 2008; Bracher *et al.*, 2017). The three-dimensional environment of the hexadecameric structure was recreated, taking into account the interaction surface between catalytic dimers and highlighting the interface areas between catalytic dimers (Figure 4.7). As we can observe in the image, the loop location would allow ample movement to occur. Interestingly, the flexibility and movement of the loop were visible even if the complex between catalytic dimers was not subjected to MD simulations. These results were fascinating given the relevance of the flexibility of this loop during the catalytic activity of the holoenzyme.

The investigation of the electrostatic interactions between the solvent and the charged residues of the proteins (*Hn, Cn and CnMut*) highlighted the surface potential changes throughout the MD simulation, as well as, the influence of the residue mutations. When the interactions between the solvent and the protein backbone were initially investigated using RDF analysis, *Hn* and *Cn*RubsiCO showed an opposit interaction behaviour with cations and anions at both 20 and 150 mM (Figure 4.8 and 9). The *Hn*RubisCO showed weak interactions with the Cl^- ions at both 20 and 150 mM compared to *Cn*RubisCO. This difference was especially visible at 20 mM where *Cn*RubisCO interaction peaks for both positively charged residues were three times higher than the *Hn* counterpart. The opposite was observed when Na^+ ions were taken into account. The *Hn*RubisCO demonstrated a very high interaction affinity with Na^+ ions at 20 and 150 mM compared to *Cn*, where the peaks were found to be three times higher than the counterpart. The *CnMut*RubisCO showed a remarkable similarity to *Hn*RubisCO in the interaction intensity with both cations and anions, having only one main difference at 20 mM

with higher interaction between the Lys and Cl^-. These promising results highlighted how the redistribution of the surface electrostatic potential in *CnMut*RubisCO towards an *Hn*RubisCO-like, overall changed the protein's short- and long-range interactions with the ions during the MD simulation in aqueous phase. Short- and long-range interactions are paramount during protein-protein recruitment mechanisms (Zhang *et al.*, 2011; Mihiri Shashikala *et al.*, 2019; Vascon *et al.*, 2020). Since the CsoS2 peptide recognition regions are positively charged (Oltrogge *et al.*, 2020), these changes in behaviour of *CnMut*RubisCO, especially with Na^+ ions, could lead to suitable forces for its assembly within the CB.

Using the IDA analysis, the locations in which cations and anions spent most of the time were investigated throughout the MD simulations. This information allowed to gauge where the interactions between the protein backbone and the ions observed in the RDF analysis were more likely to take place.

The anion density at 20 mM was mainly localised in the central groove of the protein, extending in the case of *Hn*RubisCO slightly diagonally on the x axis of the structure between loop residues 321–324 and 372–375 and for both *Cn* and *CnMut*RubisCO around loop residues 298–303 and 381–383 (Figure 4.10). At 150 mM, for *Hn*RubisCO the ions density did not change its shape dramatically. It stretched upward into the right side of the protein groove towards residues 331–335 (Figure 4.11 A; highlighted with orange square). In contrast, the anion density in *Cn* and *CnMut*RubisCO stretched diagonally towards the left side of the groove enveloping the residues 298–305 and 312–313 that form an α-helix and a loop (Figure 4.11 C; highlighted with orange square). Interestingly, in both cases and at all ion strengths, the Cl^- occupancy was always positioned near the dimer catalytic interface (Figure 4.13; orange square). It is open to speculation that such specific ion distribution for both red-type and green-type RubisCO, could be due to specific positively charged

residues at the location acting as a trap for chloride anions during MD simulation. This behaviour *in vivo* can then attract the negatively charged substrates (CO_2 and RuBP) in the catalytic active site (Andersson, 2008; Bracher *et al.*, 2017).

Interestingly, a remarkable similarity at 20 mM and 150 mM in the distribution of the Na^+ ions around both *Hn*RubisCO and *CnMut*RubisCO was observed, especially at the top (Figure 4.10; yellow square) and at the bottom of the structures (Figure 4.10; dark blue squares). Both protein variants at 20 mM had a sodium density around the residues on the right side of the flexible loop (*Hn*; 22–28, 69–72 and *CnMut*: residues 35–37, 77–83; Figure 4.10; purple squares) highlighted in a recent research paper (Oltrogge *et al.*, 2020) as interacting with the recruiting peptides of CsoS2 protein. For the *CnMut*RubisCO in this location, residues Asp36, Asp79 and Try83 could provide the salt bridge and H-bond interaction necessary to bind the recruiting peptides. In addition, at the same ionic strength another cations density cloud was found to be intense for *Hn* and more localised for *CnMut* around two β-sheets in the large subunit (*Hn*: residues 346–349 and 358–363; *CnMut*: residue 355–357 and 367–370) and around a loop of the small subunit (*Hn*: residues 94–98; *CnMut*: residues: 86–87; Figure 4.10; highlighted with purple squares in the third frame). In this location, interactions between CsoS2 recruiting peptides and *Hn*RubisCO residues of the large subunit Phe346, Asp360 and small subunit Asp94, Tyr96 were observed (Oltrogge *et al.*, 2020). Unfortunately, the *CnMut*RubisCO offers only three out of four (large subunit Tyr356, Asp370 and small subunit Asp86) interaction points for the peptides. The accessibility of the peptide to the solvent exposed residues in both interaction locations in the hexadecameric structure was confirmed (Figure 4.10; purple square).

The electrostatic forces were shown to be a main factor contributing to recruitment of binding partners at large distances, but also

contributing to the correct positioning of the partners at short distance. The precise balance between the specific pair-wise interactions and the overall electrostatic potential of the region, defines the outcome of the interaction (Mihiri Shashikala *et al.*, 2019). For this reason, the high-density cloud around the interaction zone is a good sign for the long-distance attraction, but it could be argued that the missing the cation-π interaction (between CsoS2 residue 10 and CbbS Thr88) could lead to an incorrect positioning of the recruiting peptides in the zone.

To summarise, during all the MD analysis clear differences in protein stability and electrostatic interactions with ions were observed between the two proteins *Hn* and *Cn*RubisCO. While, the same analysis highlighted a closer behaviour of *CnMut*RubisCO to *Hn*RubisCO, rather than the parental *Cn*RubisCO. Furthermore, during our MD simulations, the ID analysis showed again a close resemblance of the Na^+ ion cloud around *CnMut* to the *Hn*RubisCO. Unfortunately, the differences observed in the electrostatic surface potential of the two proteins could play an important role during RubisCO recognition and therefore influence CsoS2 peptide binding. Given the differences between the three proteins we could speculate that this could give interesting results and behaviours to compare during *in vivo* experiments.

To check the *in vivo* catalytic activity of *CnMut*RubisCO, the protein was expressed in a *C. necator* mutant strain, lacking both copies of the RubisCO-encoding genes (ΔΔ*cbbLS)*, which was therefore unable to fix CO_2. During the experiment, the strain carrying *CnMut*RibisCO was able to grow solely in autotrophic conditions and therefore showing that the catalytic activity of the protein was intact despite mutations (Figure 4.14).

However, our data is not conclusive enough to anticipate the *in vivo* incorporation of either *Hn*RubisCO or *CnMut*RubisCO in the α-

carboxysomes in *C. necator* H16. Further investigations are needed to shed light on the involvement of the surface potential in the recruitment mechanism of the CsoS2 peptides.

4.4 Future work

While our current results provide detailed insights into the influence of salt concentration on the stability of three isolated RubisCO variants, further investigation and modelling of the interactions between CsoS2 peptides and the full hexadecameric RubisCO holoenzyme (*Hn*RubisCO and *CnMut*RubisCO), especially when dealing with heterologous mutated RubisCO, would be interesting. However, this type of calculations is computationally demanding and require a significantly larger amount of resources given that the number of atoms involved in the MD simulation of a complete hexadecameric structure could exceed millions of atoms. Moreover, the crystallographic structure solved by Oltrogge and co-workers has unresolved residues, especially in the flexible loop regions.

For obtaining an experimental confirmation of the simulation results, it would be required to create a chimeric *cso*-operon carrying the *CnMut*RubisCO and *Cn*RubisCO as a negative control. In a recently published study (Dai *et al.*, 2018), the assembled hexadecameric RubisCO and the assembly intermediates leading to the carboxysome formation in Syn5 cyanophage-infected cyanobacteria (*Synechococcus sp.*), have been visualised using Cryo-Electron tomography. A similar experimental approach could be used to further investigate the integrity and assembly of the carboxysomes, as well as their possible co-localisation with RubisCO in *C. necator* strains co-expressing the *cso* operon and the three different RubisCO proteins (*Hn*, *Cn* and *CnMut*).

Chapter 5 – Deregulation of the *cbb* operon in *C. necator* H16 for improved carbon fixation rates

The model organism *Cupriavidus necator* H16 is able to produce biodegradable thermoplastics by growing on CO_2 and H_2 as a sole carbon and energy source. CO_2 assimilation in *C. necator* is mediated by the endogenous Calvin-Benson-Bassham (CBB) cycle encoded by two almost identical *cbb* operons. Both are tightly regulated by the LysR-type regulatory protein CbbR located at the 5' end of the chromosomal copy of the *cbb* operon. Unfortunately, the role and identity of positive and negative CbbR effector molecules are still not well understood in *C. necator*.

In this chapter, the deregulation of the chromosomal copy of the *cbb* operon was investigated. To do so, the promoter region was replaced by either a constitutive (P_{trp}) or an inducible promoter (P_{BAD}) or a UoN patented orthologous expression system (*tcdR*/P_{tdcB}; Minton *et al.*, 2013). Moreover, the constitutive expression of both copies of the *cbb* operon was tested *via* metabolic engineering of the CbbR regulator (Dangel and Tabita, 2015a). All engineered strains were tested for their carbon fixation rates under autotrophic conditions and the best performing strains under mixotrophic feeding conditions.

5.1 Introduction

CO_2 assimilation in *C. necator* H16, is mediated by Calvin-Benson-Bassham (CBB) cycle encoded by two almost identical *cbb* operons (Pohlmann *et al.*, 2006; Little *et al.*, 2019). The heavy energy demand and the burden of additional protein synthesis placed on the cell by CO_2 assimilation, requires the *cbb* operon transcription under the control of multiple mechanisms. The master regulator of the operon is a LysR-type transcription regulator along with its effectors, which provide ideal targets in metabolic engineering of the CBB pathway. Transcription of the *cbb* operons rely on the transcriptional regulator protein, CbbR, a LysR-type transcriptional regulator (LTTR) family member (Dangel and Tabita, 2015b). In *C. necator,* phosphoenolpyruvate (PEP) is known to act as a negative effector (Grzeszik *et al.*, 2000), but the identity of potential positive effector molecules is scarcely understood. Recent research also confirmed that RuBP, ATP and NADPH act as co-repressors increasing the binding of the CbbR to the *cbb* promoter (Dangel and Tabita, 2015b, 2015a). One of the strategies for the deregulation of the chromosomal copy of the *cbb* operon aims to relieve it from its natural regulation, which is highly entangled with the central metabolism of the cell. To neutralise the natural internal regulation of the operon, the deletion of the *cbbR* gene and all the intercistronic regions (P_{cbbL} and P_{cbbR}) is required (Figure 5.1). This strategy investigates LysR independent transcription of the *cbb* operon driven by either the constitutive promoter P_{trp}, the arabinose inducible promoter P_{BAD} or the UoN patented orthologous inducible system *tcdR*/P_{tdcB} in *C. necator*.

Our alternative strategy will also combine the engineering of existing positive control elements of *cbb* operon transcription for a constitutive expression of both genomic copies, using an amino acid residue alteration of the LTTR protein (Dangel and Tabita, 2015a). This constitutive activity could be described as activation of gene expression normally under repressing conditions and often without

the presence of effectors. Previous research (Dangel and Tabita, 2015a) showed that a single amino acid substitution can lead to minimal conformational changes resulting in a constitutive transcriptional activity. The CbbR A167V mutation chosen for this study, appeared to confer a relatively high activity for both permissive (chemoautotrophic) and repressive (chemoheterotrophic) conditions. Moreover, the mutation showed 50% greater RubisCO and reporter gene (β-galactosidase) activity compared to CbbR wild-type in chemoautotrophic conditions allowing us to test the strain in both autotrophic and the mixotrophic feeding conditions. The constant activation of the two copies of the *cbb* operon in both conditions could be desirable for the two-phase fermentation process usually applied for larger scale production of PHAs in *C. necator* (Grousseau, Blanchet, *et al.*, 2014; Ghysels *et al.*, 2018). In particular recent studies (Huschner *et al.*, 2015) showed that mixotrophic feeding of *C. necator* would be desirable for cost reduction, especially with a mixture of uneven carbon chain (propionic and valeric acid) and even chain (butyric acid) as additional carbon sources, as the latter (C4) would increase productivity while the former would improve the thermal and mechanical properties of the produced PHAs.

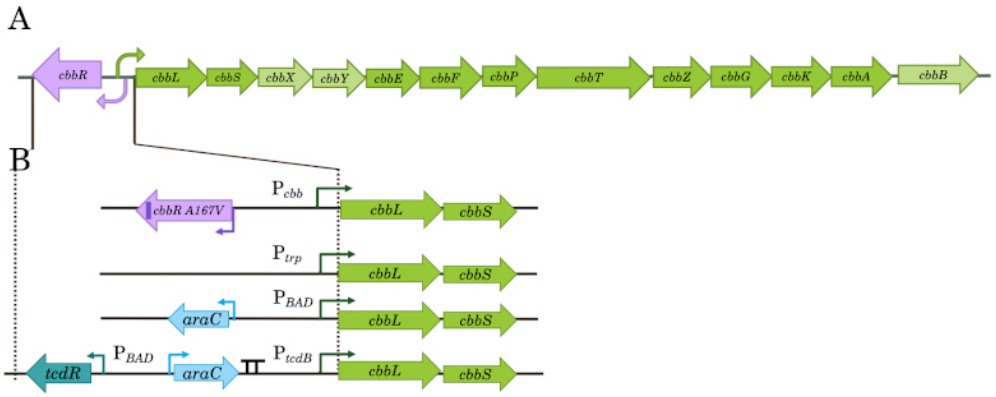


Figure 5.1: Schematic representation of the *cbb* operon and molecular engineering strategy. (A) Schematic view of chromosome II *cbb* operon in *C. necator* H16. Structural *cbb* genes are shown in green and the *cbbR* gene in light violet. (B) The metabolic engineering planned for the *cbb* promoter region on chromosome II. LysR dependent strategy replacing *cbbR* gene with a constitutively active version *cbbR*A167V (mutation Ala167Val). LysR independent strategy replacing the intercistronic region and the regulator *cbbR* with either the constitutive promoter P_{trp} or the inducible promoter P_{BAD} or the *tcdR*/P_{tcdB} UoN patented orthologous inducible system.

The overall aim of this study is to test both LysR independent and LysR dependent constitutive transcription of the *cbb* operons to assess if these modifications prove to be advantageous for enhancing the efficiency of CO_2 fixation of the CBB cycle.

5.2 Results

5.2.1. Construction of a reporter system for evaluating an inducible, orthogonal expression cassette in *C. necator* H16

The *tcdR*/P_{tcdB} orthogonal expression system (WO2013/144647) contains the promoter region of the *tcdB* gene and regulator *tcdR* from *Clostridiodes difficile*. The *tcdR* gene was identified with other genes (*botR, tetR* and *uviA*) in the *Clostridium* family to encode for alternative sigma factors natively used for the expression of toxins in pathogenic organisms such as *C. difficile, C. botulinum, C. tetani*, and *C. perfringens*. They differ enough in structure and function from other sigma factors to be assigned together in group V within the σ^{70}-family (Mani and Dupuy, 2001; Dupuy and Matamouros, 2006). In C. *difficile*, *tcdR* controls the expression of two main toxin genes *tcdB* and *tcdA via* the respective promoters P_{tcdB} and P_{tcdA}. Interestingly, the P_{tcdB} promoter has a really tight regulation given the nature of the product and doesn't show activity in absence of the alternative sigma factor. For this reason, the *tcdB* promoter can be used for vector carrying purposes in *E. coli* or other recipient organisms, without burdening their metabolism until the expression of the alternative sigma factor (*tcdR*) is enabled by induction. Previous studies have also shown that expression of this cassette upon induction resulted in very high expression of genes in the heterologous host *C. acetobutylicum* (Wilding-Steele, 2017) .

This technology was tested in *C. necator* by constructing an arabinose inducible system in which the *tcdR* gene was under the control of the arabinose inducible promoter P_{BAD}, while the P_{tcdB} promoter was cloned upstream the targeted genes. As the arabinose inducible *tcdR*/P_{tcdB} orthogonal expression system was previously untested, a reporter plasmid carrying the well characterised arabinose inducible promoter (Fukui *et al.*, 2011; Alagesan, Hanko, *et al.*, 2018; Johnson *et al.*, 2018) upstream of the *rfp* reporter gene (red fluorescent protein;

pBBR1-2USER-LYC1 plasmid; CRG culture collection) was used as template for *rfp* sequence amplification. Three control plasmids were also built, each lacking a specific element of the expression cassette to evaluate the influence of each element in the overall performance of the system. All the constructs were cloned into the modular vector pMTL71301 and the resulting plasmids transformed in *C. necator* H16. The system along with the appropriate controls were also tested in *E. coli*.

5.2.1.1 Investigating expression levels of the inducible, orthogonal system in *E. coli* and *C. necator* H16

5.2.1.1.1 Evaluation of the orthogonal expression system in E. coli:

To evaluate the correct assembly of the cassette and the level of the expression in *E. coli* the culture samples, carrying the *tcdR*/P_{tcdB} system and the P_{BAD} promoter were grown overnight in LB media. The next day the cultures were spun down, washed and the starting OD_{600} adjusted at 0.2 in the multiplate reader (Infinite M1000 Pro). After incubation, the samples were induced in exponential phase (OD_{600} 0.3–0.5) with 10 mM of L-arabinose. The same procedure was applied to the cultures carrying the control plasmids (pMTL71301::P_{BAD}//P_{tdcB}*rfp*; pMTL71301::P_{BAD}*tdcR*//*rfp*; pMTL71301::P_{BAD}*tdcR* P_{tdcB}//, with the double forward slash indicating the lacking element of the expression cassette). With the double intent to assess the background autofluorescence and have an internal reference for the strain growth in this condition, a culture was added in a separate well containing *E. coli* strain transformed with an empty vector (pMTL71301::P_{BAD}). The analysis of OD_{600} data showed no significant difference in growth rate and maximum growth between the strains carrying either the empty vector, the arabinose inducible promoter or the orthogonal expression cassette. This indicates that no metabolic impairment was suffered by *E. coli* expressing the alternative sigma factor TcdR. And, the fluorescence intensity recorded for the *tcdR*/P_{tcdB} system was 20 times lower than

the intensity recorded for the P_{BAD} promoter (Figure 5.2). Overall, a tight regulation of orthogonal expression cassette given by the alternative sigma factor was observed, resulting though in low transcription rates and expression of the reporter gene. No fluorescence was recorded for any of the control samples and an unexpected slightly faster entrance in exponential phase was observed while reaching the same growth plateau if compared to both the inducible promoter and the orthogonal cassette (Figure 5.3).

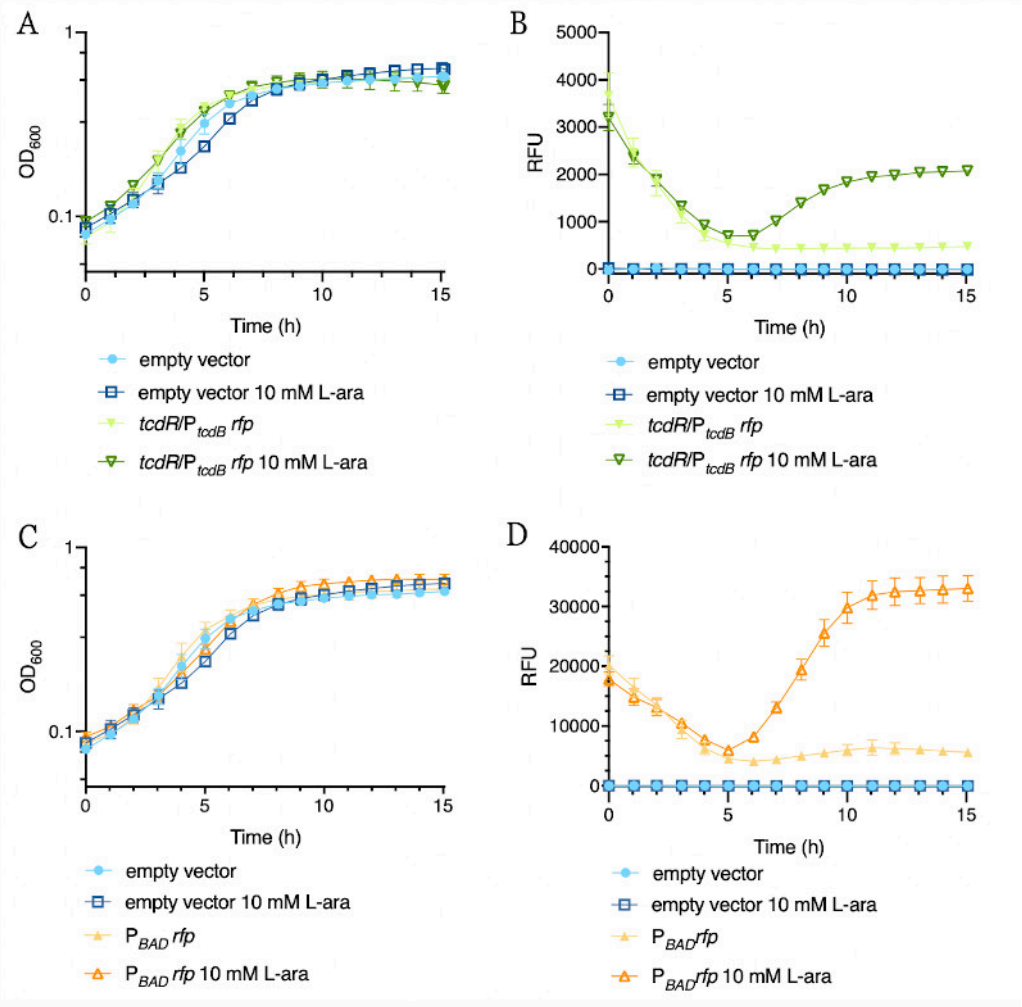


Figure 5.2: Growth and fluorescence measurements of *E. coli* carrying either the arabinose inducible promoter or the orthogonal expression system in LB media. Growth (A and C) and relative fluorescence measurements (B and D) in LB media of *E.coli* carrying either an empty vector (pMTL71301::P_{BAD}), a plasmid containing the arabinose inducible promoter upstream the *rfp* gene (pMTL71301::P_{BAD}*rfp*) or the arabinose inducible orthogonal expression cassette upstream the *rfp* gene (pMTL71301::P_{BAD}*tcdR* P_{tcdB}*rfp*) supplemented at final concentration of 10 mM with L-arabinose or no inducer. Error bars represent standard deviations of three biological replicates. RFU: Relative Fluorescence Units.

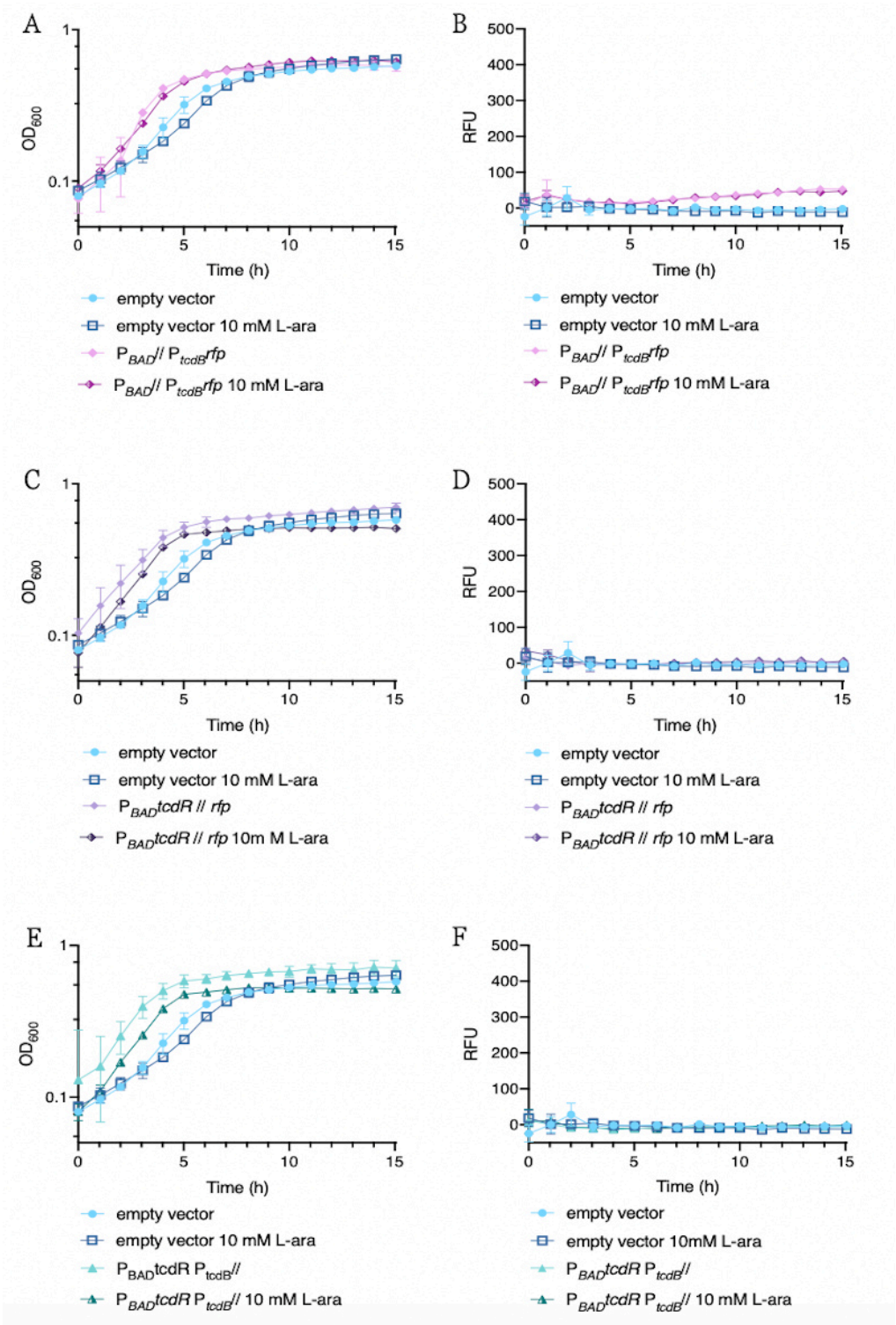


Figure 5.3: Growth and fluorescence measurements of *E. coli* carrying control plasmids in LB media. Growth (A, C and E) and relative fluorescence measurements (B, D and F) in LB media of *E.coli* carrying either an empty vector (pMTL71301::P_{BAD}), a control plasmid without *tcdR* (pMTL71301::P_{BAD}//P_{tcdB}*rfp*), a control plasmid without P_{tcdB} (pMTL71301::P_{BAD}*tcdR*//*rfp*) or a control plasmid without *rfp* (pMTL71301::P_{BAD}*tcdR* P_{tcdB} //) supplemented at final concentration of 10 mM with L-arabinose or no inducer. Error bars represent standard deviations of three biological replicates. RFU: Relative Fluorescence Units.

5.2.1.1.2 Evaluation of the orthogonal expression system strength in C. necator H16

The new orthogonal expression system was tested next in our model organism, *C. necator* H16. As mentioned before, since the arabinose inducible *tcdR*/P_{tcdB} orthogonal expression cassette was not evaluated in *C. necator* H16 before, an internal inducible expression reference (P_{BAD}*rfp*) was used to comparatively assess the level of expression of the cassette. The same experimental conditions described previously were used for *C. necator* H16 cultures, carrying either the internal reference or the *tcdR*/P_{tcdB} system. The samples were grown overnight in LB media. The next day the cultures were spun down, washed and then diluted to an initial OD_{600} of 0.2 in the multiplate reader (Infinite M1000 Pro). After incubation, the samples were induced in exponential phase (OD_{600} 0.3 – 0.5) with 10 mM of L-arabinose. The control vectors (pMTL71301::P_{BAD}//P_{tdcB}*rfp*; pMTL71301::P_{BAD}*tdcR*//*rfp*; pMTL71301::*P_{BAD}tdcR* P_{tdcB}//) were subjected to the same procedure and moved onto the same 96-wells MTP for comparison. As in previous experiments, a culture containing *C. necator* strain transformed with an empty vector (pMTL71301::P_{BAD}) was added in triplicates as negative control.

In *C. necator* a significant difference in lag phase (up to 5 h) and growth rate was observed between the cultures carrying the empty plasmid and the cultures expressing the orthogonal expression system, in both induced and no inducer conditions (Figure 5.4 A). While, no difference was recorded in growth rate and maximum growth for the strains carrying the arabinose inducible promoter in both conditions (10 mM L-arabinose and no-inducer; Figure 5.4 C). Possibly it indicates either a metabolic burden slowing down the replication rate of the organism or a particular sensitivity of *C. necator* to the expression of the alternative sigma factor. Interestingly, the disappearance of the growth impairment was observed in the *C. necator* cultures lacking either *tcdR* or *rfp* genes

from the construct, with no difference between induced and uninduced samples. The reappearance of the growth impairment was seen after the P_{tcdB} promoter removal from the cassette, with a lag phase for the culture of about 2 h. It appears that the absence of the P_{tcdB} promoting the accumulation of the alternative sigma factor in the cell, with no clear target to bind to, could lead to a toxic effect in the microorganism. A similar situation has been previously noticed in our lab in *C. difficile* and *C. acetobutylicum* cultures, when a strong promoter was used for the TcdR expression. Overall these data suggest a possible combined effect where the intrinsic leakiness of P_{BAD} promoter causes the accumulation of the alternative sigma factor detrimental for microorganism replication and the recognition of the P_{tcdB} leading to the expression of cytotoxic *rfp* gene (Verkhusha and Kiryl, 2011) caused the growth impairment observed in Figure 5.4 A. To test this hypothesis further experiments were carried out using a P_{BAD} promoter catabolic repressor in the growth media (1% Glucose LB medium).

Unfortunately, low fluorescence measurements were recorded for the orthogonal expression cassette, about 50 times less intense than the inducible promoter P_{BAD} 5 hr after induction (Figure 5.4 B and D). The fluorescence intensity was almost the same value as the auto-fluorescence produced by the cultures carrying the empty plasmid, with no difference between induced and uninduced samples. No fluorescence was recorded from the cultures carrying the control plasmids (Figure 5.5 B,D and F), while the fluorescence intensity recorded for the arabinose inducible promoter (Figure 5.4 D) was consistent with previous published work carried out in our lab (Alagesan, Hanko, *et al.*, 2018). Unfortunately, the data suggested that the orthogonal system is not functional in our model organism, *C. necator* H16. To be certain that no mutation of the vector occurred during the transformation procedure in *C. necator* further

experiments involving extraction, sequencing and re-transformation of the vector in *E. coli* were carried out (further details below).

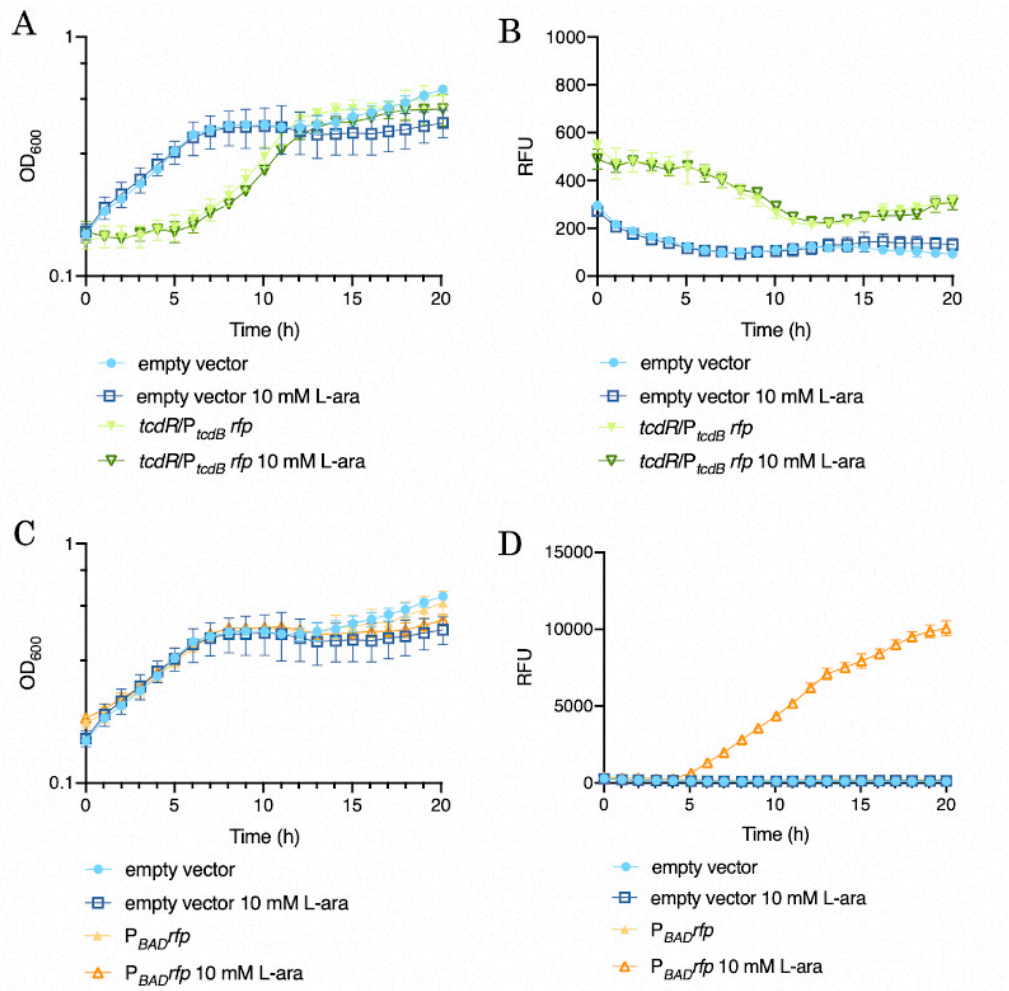


Figure 5.4: Growth and fluorescence measurements of *C. necator* carrying either the arabinose inducible promoter or the orthogonal expression system in LB media. Growth (A and C) and relative fluorescence measurements (B and D) in LB media of *C. necator* H16 carrying either an empty vector (pMTL71301::P_{BAD}), a plasmid containing the arabinose inducible promoter upstream the *rfp* gene (pMTL71301::P_{BAD}rfp) or the arabinose inducible orthogonal expression cassette upstream the *rfp* gene (pMTL71301::P_{BAD}*tcdR* P_{tcdB}*rfp*) supplemented at final concentration of 10 mM with L-arabinose or no inducer. Error bars represent standard deviations of three biological replicates. RFU: Relative Fluorescence Units.

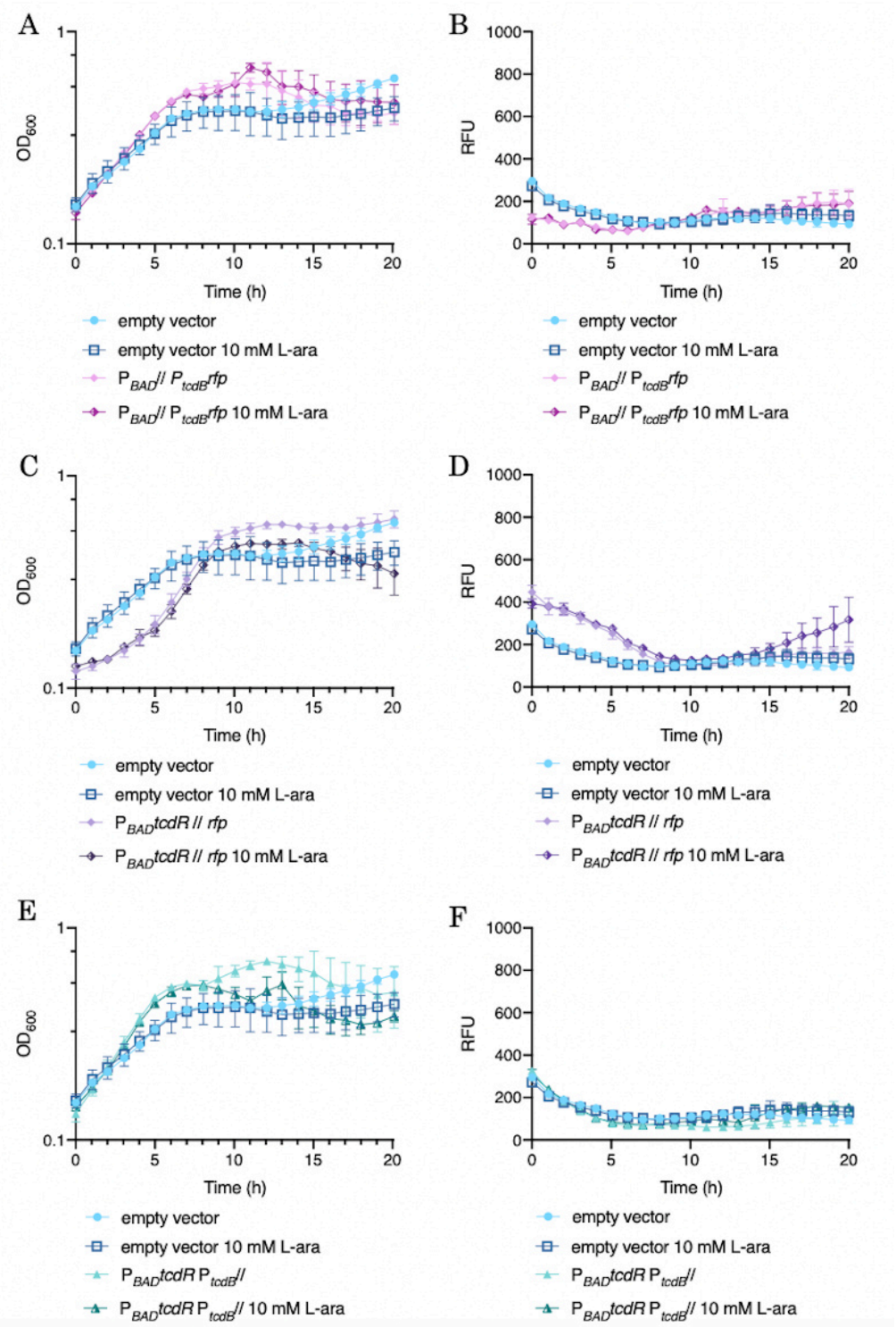


Figure 5.5: Growth and fluorescence measurements of *C. necator* carrying control plasmids in LB media. Growth (A, C and E) and relative fluorescence measurements (B, D and F) in LB media of *C.necator* H16 carrying either an empty vector (pMTL71301::P_{BAD}), a control plasmid without *tcdR* (pMTL71301::P_{BAD}//P_{tcdB}*rfp*), a control plasmid without P_{tcdB} (pMTL71301::P_{BAD} *tcdR*//*rfp*)or a control plasmid without *rfp* (pMTL71301::P_{BAD} *tcdR* P_{tcdB} //) supplemented at final concentration of 10mM with L-arabinose or no inducer. Error bars represent standard deviations of three biological replicates. RFU: Relative Fluorescence Units.

To assess whether the accumulation of TcdR due to P_{BAD} intrinsic leakiness was directly correlated with the growth impairment observed, the experimental set up was kept as described before, with the exception of 1% Glucose LB used as growth media. Interestingly, *C. necator* H16 carries a metabolic deficiency of both glucose uptake and phosphorylation activity due to a missense mutation in *nagE* and a nonsense mutation in *nagR* on Chromosome I (Pohlmann *et al.*, 2006; Orita *et al.*, 2012) . For this reason, glucose was an optimal choice for repressing the background transcription of the *tcdR* gene until induction.

As we can observe in Figure 5.6 and 5.7, in the presence of glucose, no growth impairment was detected in the cultures carrying either the complete orthogonal expression cassette (Figure 5.6 A) or the control plasmid without P_{tcdB} (Figure 5.7 B). Consistently with what was observed before, the fluorescence intensity recorded was close to the auto-fluorescence signal for both cultures carrying either the complete *tcdR*/P_{tcdB} cassette (Figure 5.6 B) or the control samples (Figure 5.7 B, D and F). The signal for the inducible promoter was consistent with the previous runs in *C. necator* (Figure 5.6 D).

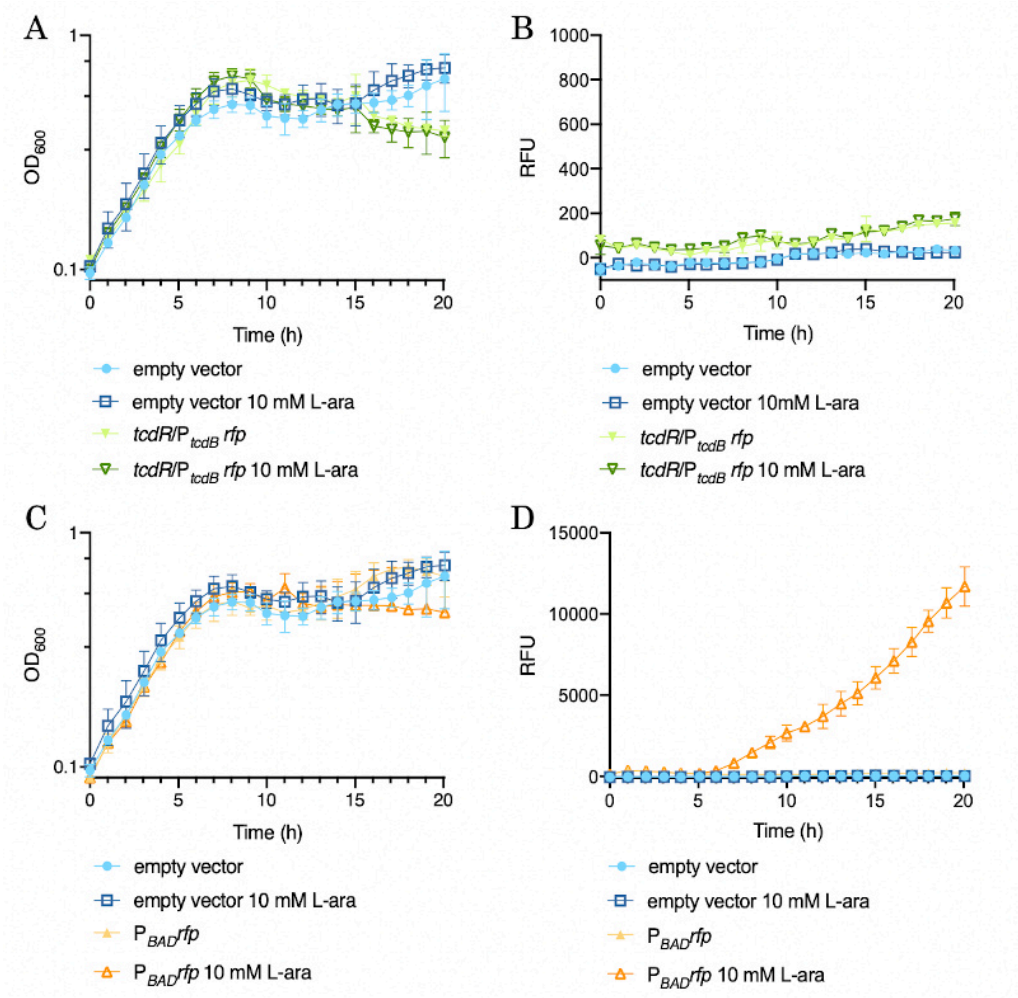


Figure 5.6: Growth and fluorescence measurements of *C. necator* carrying either the arabinose inducible promoter or the orthogonal expression system in LB 1% Glucose media. Growth (A and C) and relative fluorescence measurements (B and D) in LB 1% Glucose media of *C.necator* H16 carrying either an empty vector (pMTL71301::P_{BAD}), a plasmid containing the arabinose inducible promoter upstream the *rfp* gene (pMTL71301::P_{BAD}*rfp*) or the arabinose inducible orthogonal expression cassette upstream the *rfp* gene (pMTL71301::P_{BAD}*tcdR* P_{tcdB}rfp*)* supplemented at final concentration of 10 mM with L-arabinose or no inducer. Error bars represent standard deviations of three biological replicates. RFU: Relative Fluorescence Units.

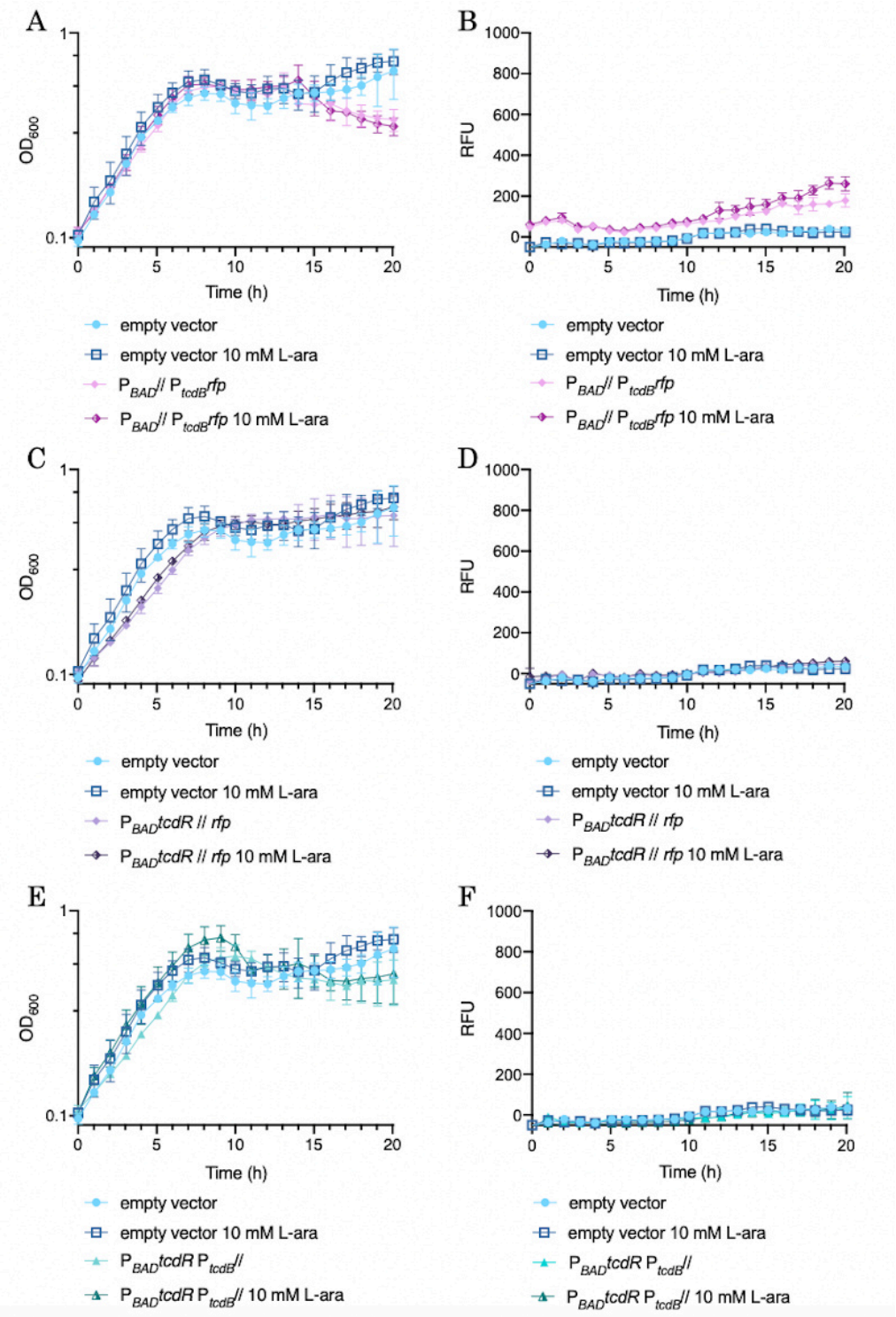


Figure 5.7: Growth and fluorescence measurements of *C. necator* carrying control plasmids in LB 1% Glucose media. Growth (A, C and E) and relative fluorescence measurements (B, D and F) in LB 1% Glucose media of *C.necator* H16 carrying either an empty vector (pMTL71301::P_{BAD}), a control plasmid without *tcdR* (pMTL71301::P_{BAD}//P_{tcdB}*rfp*), a control plasmid without P_{tcdB} (pMTL71301::P_{BAD}*tcdR*//*rfp*)or a control plasmid without *rfp* (pMTL71301:: P_{BAD}*tcdR* P_{tcdB}//) supplemented at final concentration of 10 mM with L-arabinose or no inducer. Error bars represent standard deviations of three biological replicates. RFU: Relative Fluorescence Units.

To rule out the possibility of a mutation event in the vector influencing the fluorescence assay measurements in *C. necator* strain, the plasmids carrying the *tcdR*/P$_{tcdB}$ cassette were extracted from the analytical triplicates of the 96-well plate, they were sent for sequencing and independently transformed back into *E. coli*. The experimental set up for the rescue plasmid fluorescence assay in *E. coli* was kept as previously described in section 5.2.1.1.1. In this instance, we used only *E. coli* cultures with *P$_{BAD}$ rfp* as an internal fluorescence reference and with the empty plasmid for growth and auto-fluorescence reference.

As we can observe in Figure 5.8, the fluorescence measurements recorded from the *E. coli* culture harbouring the *tcdR*/P$_{tcdB}$ system in triplicate showed that the system was ultimately functional in this strain. Therefore, no mutations occurred during the transformation protocol in *C. necator* H16. Also, the sequencing results (data not shown) confirmed the insert carrying the *tcdR*/P$_{tcdB}$ system didn't suffer any SNP or insertion/deletion event.

To our surprise, as seen in Figure 5.4 B, *tcdR*/P$_{tcdB}$ orthogonal system does not seem to be functional in *C. necator* H16, despite working well in *E. coli* or any other *Clostridia* background, therefore the arabinose inducible orthogonal expression system cannot be used for engineering of *C. necator* H16 strain.

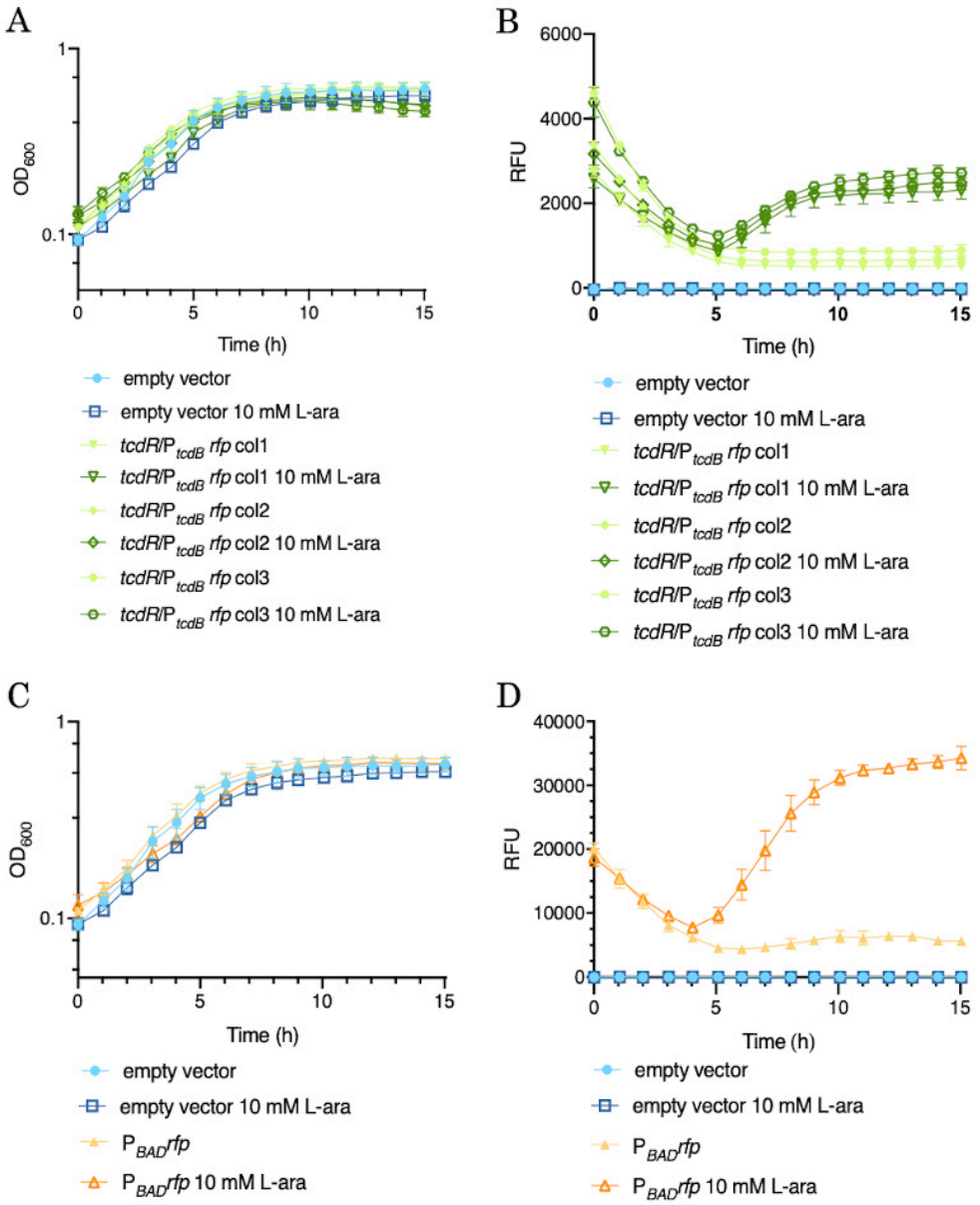


Figure 5.8: Growth and fluorescence measurements for the rescue plasmid assay of the orthogonal expression system in *E. coli*. Growth (A and C) and relative fluorescence measurements (B and D) in LB media of *E.coli* carrying either an empty vector (pMTL71301::P_{BAD}), a plasmid containing the arabinose inducible promoter upstream the *rfp* gene (pMTL71301:: P_{BAD}*rfp*) or the arabinose inducible orthogonal expression cassette upstream the *rfp* gene (pMTL71301::P_{BAD} *tcdR* P_{tcdB} *rfp*) supplemented at final concentration of 10 mM with L-arabinose or no inducer. Error bars represent standard deviations of three analytical replicates. RFU: Relative Fluorescence Units

5.2.2 Biological effect of constitutive or inducible expression of the *cbb* operons in *C. necator* mutants

The constitutive transcription of the *cbb* operons in both LysR independent and LysR dependent manner was investigated to assess if these changes would enable increased CO_2 fixation efficiencies in *C. necator* H16. Deregulation of the chromosomal copy of the *cbb* operon aims to relieve it from its natural regulation highly entangled with the central metabolism of the cell (Dangel and Tabita, 2015b; Gruber *et al.*, 2017). The dual promoter present in the intercistronic region on the chromosome II between P_{cbbL} and P_{cbbR} is well known for its complex regulation due to effectors acting on the transcriptional regulator CbbR (Dangel and Tabita, 2015b), most of which remain still unknown. Several attempts were made in the past to understand the complex relations between the effectors and CbbR (Dangel and Tabita, 2015b; Gruber *et al.*, 2017), but no thorough understanding was reached of all the effectors regulating this mechanism in *C. necator*.

In the LysR independent strategy, to relieve the *cbb* operons from the entanglement with the central metabolism and its effectors, a series of *C. necator* allelic mutant strains were created. The intercistronic region between *cbbL* and *cbbR* on chromosome II was removed along with the *cbbR* gene and replaced with either a medium strength constitutive promoter (P_{trp}) (Johnson *et al.*, 2018) from *E. coli* or an inducible promoter P_{BAD} (Fukui *et al.*, 2011; Bi *et al.*, 2013; Johnson *et al.*, 2018). The choice of a sustained and constant expression of only one copy of the *cbb* operon with P_{trp} was made to compare the growth fitness of the allelic mutant against the wild type, in hope that the constant transcription of the CBB pathway would be beneficial to the microorganism. Meanwhile, the P_{BAD} allelic mutant was created to further investigate the correlation between different levels of expression of only one copy of *cbb* operon and its growth fitness. Allowing to explore a variety of conditions that would potentially lead

towards optimum expression levels. Insertion of a strong constitutive promoter, P_{j5} derived from the bacteriophage T5 (Gruber *et al.*, 2014; Johnson *et al.*, 2018), was attempted multiple times with no success. After an extensive screening activity for the integration of the P_{j5} promoter, we noticed that all the colonies after the second homologous recombination reverted to wild type. The conjugation protocol was then modified to test the possibility of either an early or late second homologous recombination event. Dilutions 10^{-4} to 10^{-6} were plated out 24 and 48 h after inoculation in LSLB sucrose 15%, but no difference was observed. Based on the published data (Shimizu *et al.*, 2015; Alagesan, Minton, *et al.*, 2018) on *C. necator* H16, during heterotrophic growth conditions, only a partial de-repression of the *cbb* operons allows recycling of CO_2 emitted through sugar degradation. Possibly, the metabolic burden due to the high expression of CBB pathway induced by the strong promoter could be detrimental to the microorganism's overall growth, making the isolation of this mutant very difficult.

For a LysR dependent continuous activation of both *cbb* operons, we decided to use an amino acid substitution in the transcriptional regulator CbbR (Ala167Val), which was shown to lead to a constitutively active CbbR form in both autotrophic and heterotrophic conditions (Dangel and Tabita, 2015a).

Therefore, P_{trp}, P_{BAD} promoters and the mutated CbbRA167V inserts were cloned in a suicide vector (pMTL60641 series, Muhammad Ehsaan unpublished), which allows the integration of the desired fragments by homologous recombination. The resulting plasmids were transformed in *E. coli* S17 and then conjugated with in *C. necator* H16. Three individual, double cross-over colonies were selected for each construct, confirmed by sequencing and their growth was compared to that of the wild type *C. necator* H16 strain. The newly created *C. necator* strains containing the deregulated *cbb* operons were then tested for their growth in organo-autotrophic

conditions in the microbioreactor (BioLector Pro®), in autotrophic and mixotrophic conditions in the bioreactors (DASGIP® Parallel Bioreactors System).

5.2.2.1 Evaluation of mutant strains with constitutive expression of the *cbb* operon under organo-autotrophic conditions in microbioreactor (BioLector Pro®)

In these experiments, three biological replicates of each allelic mutant strain carrying either P_{trp} or CbbbR A167V were grown overnight in 0.4% SG minimal medium, washed in PBS and loaded into an MTP compatible with the BioLector Pro® machine at OD_{600} of about 0.15. Strains were grown for 50 h, in minimal media supplemented with 80mM Na-formate with 10% CO_2, as source of energy and carbon. As mentioned above, the wild type strain of *C. necator* was used as control. The growth of the allelic mutants carrying the constitutive promoter P_{trp}, unfortunately, was remarkably slower than the wild type strain under these conditions, with a doubling time of 32 h compared to the 12 h of the wild type and a μ_{max} of 0.021 h^{-1} compared to the wild type 0.057 h^{-1}, with no significant difference between biological replicates. The allelic mutant cultures had a lower growth plateau, with a significant difference ($p<0.05$) in biomass accumulation at the onset of the stationary phase when compared to the wild type *C. necator* strain (Figure 5.9).

The mutants carrying the constitutively active CbbR performed better than the P_{trp} ones under these growth conditions, reaching a similar biomass accumulation levels to that of the wild type strain at the onset of the stationary phase (Figure 5.10). For the allelic mutant the doubling time was slightly longer (17.4 h) compared to the wild type (12 h). The same trend was seen in the μ_{max} of the former (0.040 h^{-1}) compared against the wild-type strain (0.057h^{-1}). The colony two showed a noticeable higher biomass accumulation than the other two biological replicates despite having similar growth rate and doubling time (Figure 5.10). For this reason, a genomic analysis of colony one

and two was carried out. It revealed no difference in the genome of the two strains.

Based on the overall set of data we can conclude that the constant expression of both copies of the *cbb* operon is required to match the growth fitness of the wild type strain. The results might also suggest that the CbbR regulation is essential to maintain *C. necator* metabolism balance for biomass production.

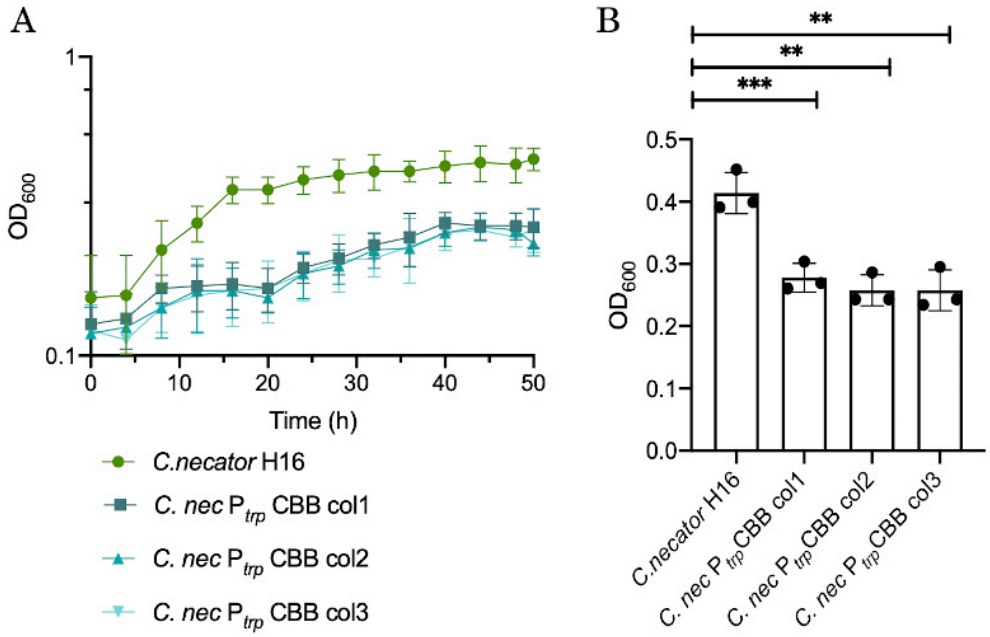


Figure 5.9: Growth of *C. necator* P_{trp} CBB allelic mutant under organo-autotrophic conditions after overnight incubation in 0.4% SG MM. (A) Growth curve of *C. necator* H16 (in green) and allelic mutants of *C. necator* H16: *C. necator* P_{trp} CBB col 1 to 3 (different shades of blue representing the three biological replicates) under organo-autotrophic conditions (Na-formate 80mM in presence of 10% CO_2) after overnight incubation in 0.4% SG MM. (B) One-way ANOVA analysis of the biomass produced by *C. necator* H16 and allelic mutants *C. necator* P_{trp} CBB col 1 to 3 after 30 h of growth under organo-autotrophic conditions. Error bars represent standard deviations of three analytical replicates. *** = $p<0.005$; ** = $p<0.05$.

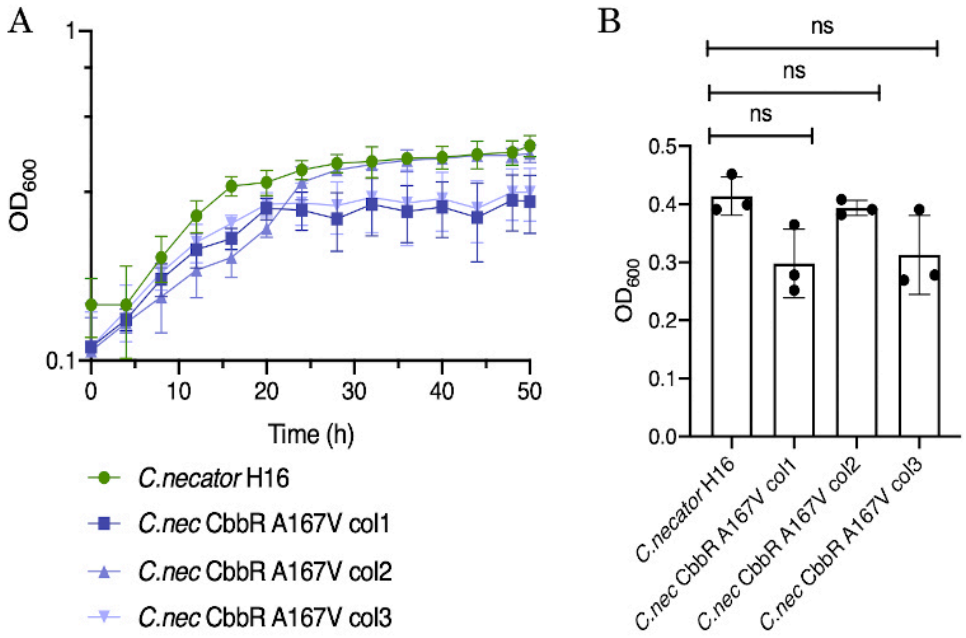


Figure 5.10 Growth of *C. necator* CbbRA167V allelic mutant under organo-autotrophic conditions after overnight incubation in 0.4% SG MM. (A) Growth curve of *C. necator* H16 (in green) and allelic mutants of *C. necator* H16: *C. necator* CbbR A167V col 1 to 3 (different shades of lilac representing the three biological replicates) under organo-autotrophic conditions (Na-formate 80mM in presence of 10% CO_2) after overnight incubation in 0.4% SG MM. (B) One-way ANOVA analysis of the biomass produced by *C. necator* H16 and allelic mutants *C. necator* CbbR A167V col 1 to 3 after 32 h of growth under organo-autotrophic conditions. Error bars represent standard deviations of three analytical replicates. ns = not significant difference.

A similar experiment was carried out, exchanging the overnight media with Na-formate 80mM MM in presence of 10% CO_2 and the resulting cultures were then grown for 36 h to test if a previous adaptation to the organo-autotrophic medium could lead to better overall performance of the strains. The wild type strain in this experiment didn't perform as expected. The strain reached higher OD_{600} during overnight incubation compared against the mutant strains and for this reason it might have already reached stationary phase. Additionally, its higher dilution during inoculum in the MTP could have hindered its performance during the BioLector run. While, both allelic mutants maintained similar values of μ_{max} and T_d if compared to the previous experiment (P_{trp} μ_{max} = 0.019 h^{-1} and T_d = 35 h; CbbR A167V μ_{max} = 0.038 h^{-1} and T_d = 18 h). The biomass accumulation for the CbbR A167V mutant at the onset of stationary phase was slightly lower than recorded in the previous experiment,

and this led to a significant difference between two of the biological replicates and the wild type (Figure 5.12). While, the biomass accumulation for the P_{trp} mutant was almost identical to the one observed in the previous experiment (Figure 5.11).

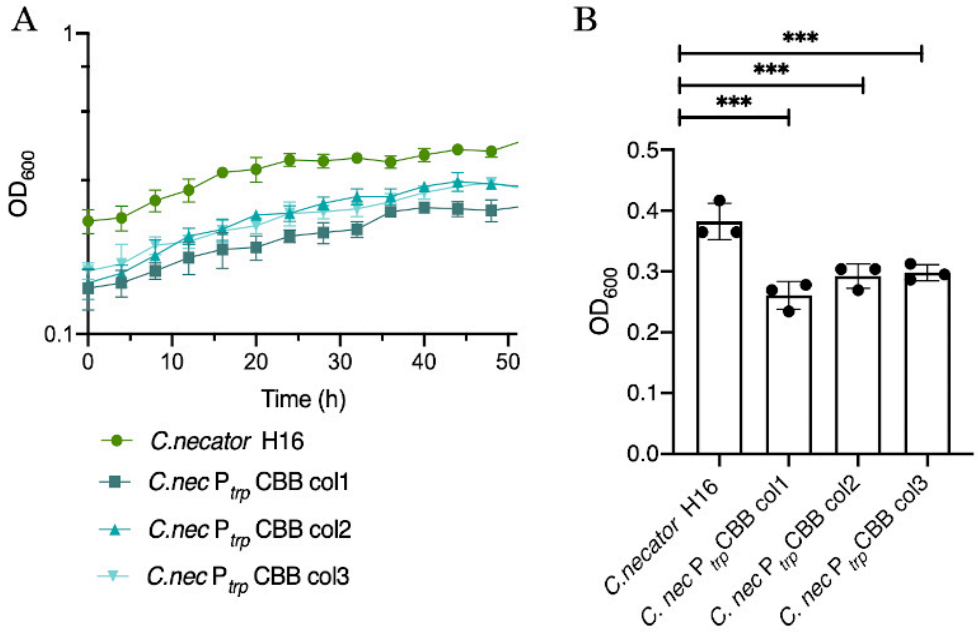


Figure 5.11: Growth of *C. necator* P_{trp} CBB allelic mutant under organo-autotrophic conditions after overnight incubation in the same conditions. (A) Growth curve of *C. necator* H16 (in green) and allelic mutants of *C. necator* H16: *C. necator* P_{trp} CBB col 1 to 3 (different shades of blue representing the three biological replicates) under organo-autotrophic conditions (80mM Na-formate MM in presence of 10% CO_2) after overnight incubation in the same conditions. (B) One-way ANOVA analysis of the biomass produced by *C. necator* H16 and allelic mutants *C. necator* P_{trp} CBB col 1 to 3 after 40 h of growth under organo-autotrophic conditions. Error bars represent standard deviations of three analytical replicates. *** = $p<0.005$.

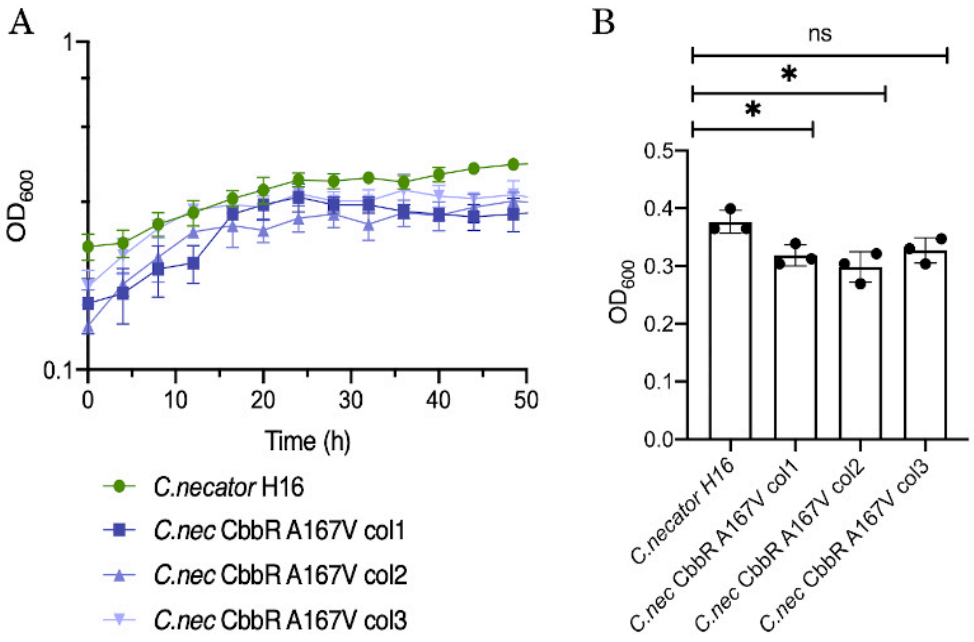


Figure 5.12: Growth of *C. necator* CbbRA167V allelic mutant under organo-autotrophic conditions after overnight incubation in the same conditions. (A) Growth curve of *C. necator* H16 (in green) and allelic mutants of *C. necator* H16: *C. necator* CbbR A167V col 1 to 3 (different shades of lilac representing the three biological replicates) under organo-autotrophic conditions (80mM Na-formate MM in presence of 10% CO_2) after overnight incubation in the same conditions. (B) One-way ANOVA analysis of the biomass produced by *C .necator* H16 and allelic mutants *C. necator* CbbR A167V col 1-3 after 32 h of growth under organo-autotrophic conditions. Error bars represent standard deviations of three analytical replicates. * = $p<0.5$; ns = not significant difference.

Overall these results confirm the consistency and reproducibility of the performance of the allelic mutant P_{trp} and CbbR A167V strains in organo-autotrophic conditions, where the overnight conditions were shown to only slightly impact the behaviour during the fermentation in the microbioreactor. Unfortunately, the *C. necator* CbbR A167V mutant didn't show better fitness (higher μ_{max}) than the wild type strain under organo-autotrophic conditions, as suggested in the literature (Dangel and Tabita, 2015a). Despite this, the strain might perform better in autotrophic and mixotrophic conditions. Therefore, it was tested under these conditions in larger fermentation experiments.

5.2.2.2 Evaluation of mutant strains with inducible expression of the *cbb* operon under organo-autotrophic conditions in microbioreactor (BioLector Pro®)

In this experiment, the allelic mutant carrying the P_{BAD} promoter and the growth reference *C. necator* H16 were grown overnight in 0.4% SG MM, washed in PBS and moved into an MTP compatible with the BioLector Pro® machine at OD_{600} of about 0.15. They were induced with L-arabinose at 0, 2.5, 5 and 10 mM at the beginning of the run and grown for 65 h in 80mM Na-formate MM with 10% CO_2. The growth control behaved as expected with a short lag phase (4 h) and μ_{max}, T_d values (μ_{max} = 0.064 h^{-1} and T_d = 11 h) very similar to the ones recorded in previous experiments, suggesting a consistent experimental setup and environmental conditions. Unexpectedly, we observed a long lag phase of about 30 h for *C. necator* P_{BAD} mutant strain under all the induction levels. Furthermore, the biomass accumulation observed for the allelic mutant under all induction levels tested, barely doubled throughout the experiment (Figure 5.13) and the maximum growth rates recorded were not significantly different between induction levels. The mutant strain with no inducer was not able to grow and there was no biomass accumulation throughout the experiment Figure 5.13). A genomic analysis was performed on the mutant srtain to confirm that the microorganism growth was not hindered by a genomic mutation. No difference was found between the *C. necator* P_{BAD} allelic mutant and the wild-type strain *C. necator* H16, except for the replacement of the P_{BAD} promoter in the chromosomal region.

Overall, these results suggest that the arabinose induction does not seem to be appropriate for controlling the expression of the *cbb* operon and investigate the correlation between different levels of expression of one copy of *cbb* operon and its growth fitness. Furthermore, the fact that the allelic mutant cultures, independent from the induction levels, had a lower growth plateau compared to the wild type, consolidates the idea that the expression of only one copy of the *cbb*

operon is not sufficient to match the wild type fitness, in these conditions. This further indicates the involvement of the CbbR regulation might be essential to maintain *C. necator* metabolism balance for biomass production.

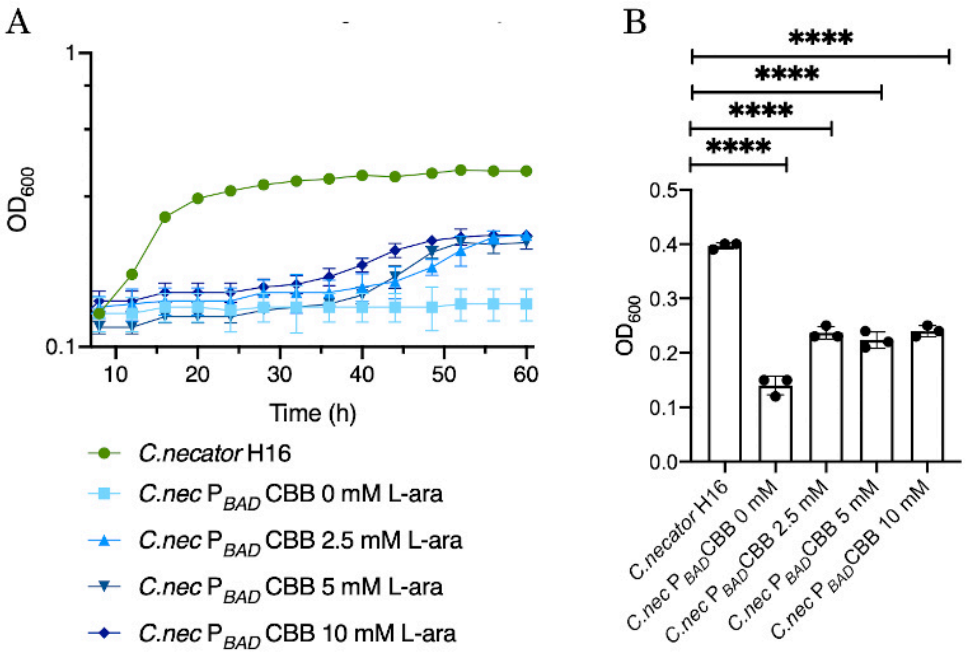


Figure 5.13: Growth of *C. necator* P_{BAD} CBB allelic mutant under organo-autotrophic conditions after overnight incubation in 0.4% SG MM. (A) Growth curve of *C. necator* H16 (in green) and allelic mutants of *C. necator* H16: *C. necator* P_{BAD} col 1 to 3 (different shades of blue representing levels of induction 0, 2.5, 5 and 10 mM L-arabinose) under under organo-autotrophic conditions (Na-formate 80mM in presence of 10% CO_2) after overnight incubation in 0.4% SG MM. (B) One-way ANOVA analysis of the biomass produced by *C. necator* H16 and allelic mutants *C. necator* P_{BAD}CBB col 1-3 after 55 h of growth under organo-autotrophic conditions. Error bars represent standard deviations of three analytical replicates. **** = $p<0.0005$.

5.2.2.3 Fermentation of the *C. necator* CbbR A167V mutant strain under autotrophic conditions (DASGIP® Parallel Bioreactor System)

The aim of this experiment was to investigate the growth fitness and the CO_2 consumption of the best performing mutant strain in organo-autotrophic conditions (*C. necator* CbbR A167V col 1 and col 2) and compare it in duplicate against the wild type strain, under autotrophic conditions. The fermentation was carried out using the bioreactors (DASGIP® Parallel Bioreactor System) with a working volume of 0.75 L. The starter cultures were grown in LB gentamicin (10µg/mL) medium and sub-cultured once in 100 mL total volume in order to reach adequate cell density for inoculation of the bioreactor at OD_{600} of 0.2. A continuous supply of 45 l/h gas mix (78% H_2, 3% CO_2 and 19% air) were provided by mass flow controller to maintain autotrophic growth. In this autotrophic condition we observed higher values for the maximum growth rate of the *C. necator* mutant strains (col 1 $\mu_{max} = 0.071\ h^{-1}$ and $T_d = 9$ h; col 2 $\mu_{max} = 0.101\ h^{-1}$ and $T_d = 6.7$ h) compared against the wild type strain ($\mu_{max} = 0.055\ h^{-1}$ and $T_d = 12$ h). These values compare well with the *in vitro* enzymatic activity calculated in the literature (Dangel and Tabita, 2015a), where a 2-fold increase in expression levels of the reference gene and therefore of its activity was recorded in autotrophic conditions for the CbbR mutation A167V. In our study, we found an *in vivo* decrease of the doubling time of 1.8-fold for *C. necator* CbbR A167V col 2 and 1.3-fold for col 1, compared against the wild type. The same trend was followed by the maximum growth rate with an increase in the values. This result suggests a higher activation of the *cbb* operons in this condition, resulting in a faster CO_2 consumption and growth.

The allelic mutants started to consume higher amounts of CO_2 earlier (around 20 h) compared to the wild type strain (30 h) (Figure 5.14). Moreover, the CO_2 consumption for the allelic mutants CbbR A167V peaked at 35 h (0.6 mol) for col 1 and at 40 h for col 2 (0.5 mol). While, for the wild type at 42 h. The *C. necator* H16 culture showed a peak

in CO_2 consumption around 42 h fermentation (0.5 mol). The oxygen and hydrogen consumption followed the same trend with a peak around 0.5 mol of O_2 and 1.8 mol of H_2 consumed for *C. necator* H16 and an average of 0.45 mol of O_2 and 1.5 mol of H_2 consumption for *C. necator* mutants (col 1 and col 2). All the cultures reached similar biomass accumulation in stationary phase ($p<0.05$; one-way ANOVA data not shown), which was reached after 50 h for both duplicate of the allelic mutant and around 65 h for the wild type strain from the start of the fermentation.

The overall performance of the mutant strains seems to follow the predicted trend, allowing a higher expression of the both *cbb* operons due to the constant activation of the transcriptional regulator CbbR.

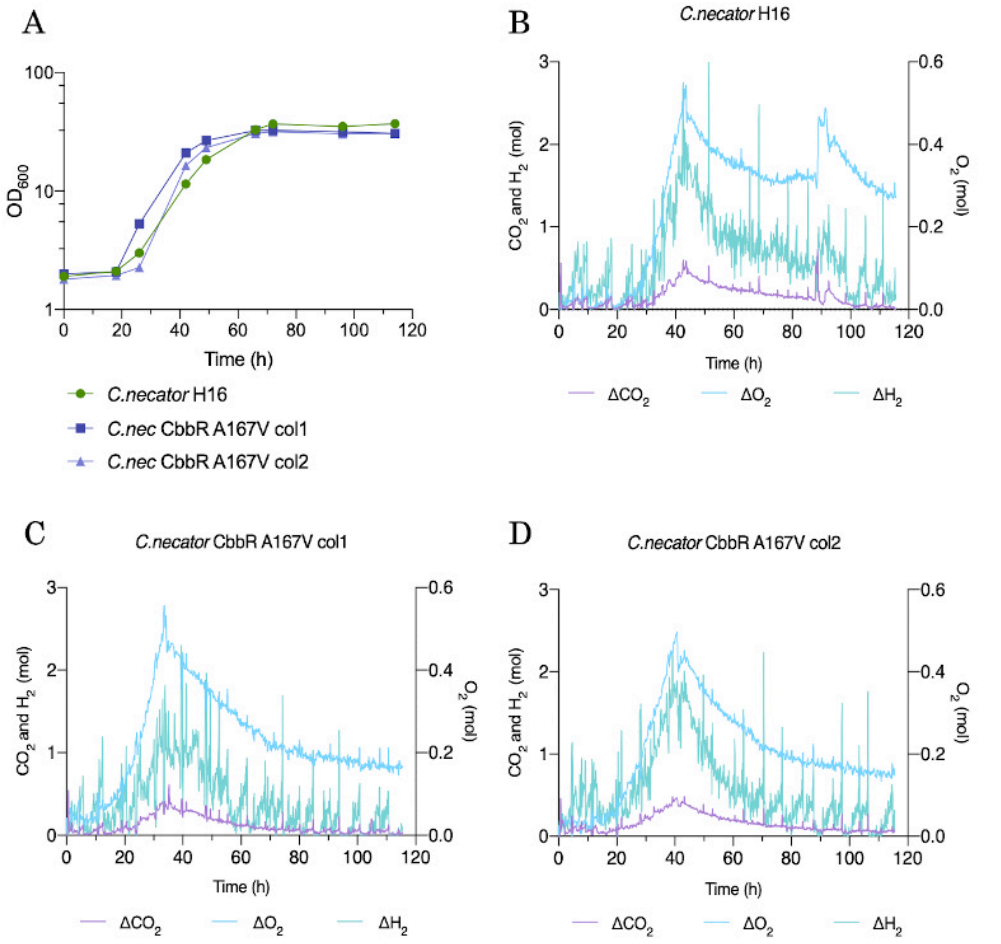


Figure 5.14: Growth and gas consumption profile of *C. necator* CbbR A167V under autotrophic conditions. (A) Growth curve of *C. necator* H16 (in green) and allelic mutants *C. necator* CbbR A167V col 1 and 2 (different shades of lilac representing the two biological replicates) in autotrophic conditions. (B, C and D) Gas profile of H_2, CO_2 and O_2 consumption (mol) for *C. necator* H16, allelic mutants *C. necator* CbbR A167V col 1 and col 2.

5.2.2.4 Fermentation of the *C. necator* CbbR A167V mutant strain under mixotrophic conditions (DASGIP® Parallel Bioreactor System)

Recent literature (Huschner *et al.*, 2015) underlined the importance of cost reduction of biopolymer production, such as PHAs, using cheap and renewable carbon sources, for example volatile fatty acids (VFAs) from anaerobic digestion (AD), to make the production process economically viable on industrial scale. Moreover, the ability to utilise VFAs of even and uneven chain length would introduce desirable characteristics in the product improving thermal and mechanical properties (Wang *et al.*, 2013; Huschner *et al.*, 2015). Unfortunately, *C. necator* ability to grow on VFAs and CO_2 couldn't be previously exploited at its optimum, due to the downregulation of the *cbb* operons activity in presence of organic compounds (Dangel and Tabita, 2015b). The possibility to use both substrates (CO_2 and VFAs) at the same time opens new avenues for reducing costs while implementing PHAs properties exploiting unused side products streams from the first stage AD fermentation process.

The aim of this first experiment was to assess PHAs production and growth for *C. necator* CbbR167 under mixotrophic conditions. The fermentation was performed in the bioreactors (DASGIP® Parallel Bioreactor System) with working volume of 0.75 L. The starter cultures of *C. necator* CbbR A167V (col 2) was grown in LB gentamicin (10µg/mL) medium and sub-cultured once in 100 mL total volume in order to reach adequate cell density for inoculation of the bioreactor at OD_{600} of 0.2. The culture was grown under mixotrophic conditions using a mixture of 10mM of VFAs from the time of inoculation. A long lag phase of about (90 h) was observed in the fermentation along with low biomass accumulation reaching an OD_{600} of 5, possibly due to the intrinsic toxicity of the VFAs at medium – high concentrations (Figure 5.15). Previous studies assessed the tolerance of *C. necator* being between 5 to 10 mM total amount of VFAs (Bhatia *et al.*, 2019). Unfortunately, the majority of C2 to C5 consumption occurred

between 93 and 114 h, where we didn't collect data points, partially hindering our understanding of co-consumption and substrate preferences for the VFAs mix components. Despite that, we recorded a sharp decrease in formic acid to 6mM between 0 and 42 h with a further major decline below 1 mM between 42 and 93 h. At the same time, the butyric acid concentration slowly declined to 6mM between 0 and 93 h, suggesting a possible substrate preference pattern (Formic acid > Butyric acid). While, all the other VFAs remained virtually untouched until 93 hr and then plummeted towards 0mM after that, with the exception of acetic acid (around 1 mM) and hexanoic acid (around 2.5 mM). This result is consistent with previous work carried out in our laboratory (Jawed and Irorere, unpublished) and present in the literature (Yan *et al.*, 2003), where the acetic acid produced in the decarboxylation of propionate and *via* β-oxidation rose and then remained un-consumed until the medium chain length VFAs were completely metabolised. The percentage of PHAs in the biomass was found to be around 63-68% during 93 h to 144 h of the cultivation, while the composition changed towards an equal amount of 3HB and 3HV monomers at 113 h (Figure 5.15 B).

Overall, the results showed the ability of *C.necator* CbbR A167V to successfully grow under mixotrophic conditions and to produce co-

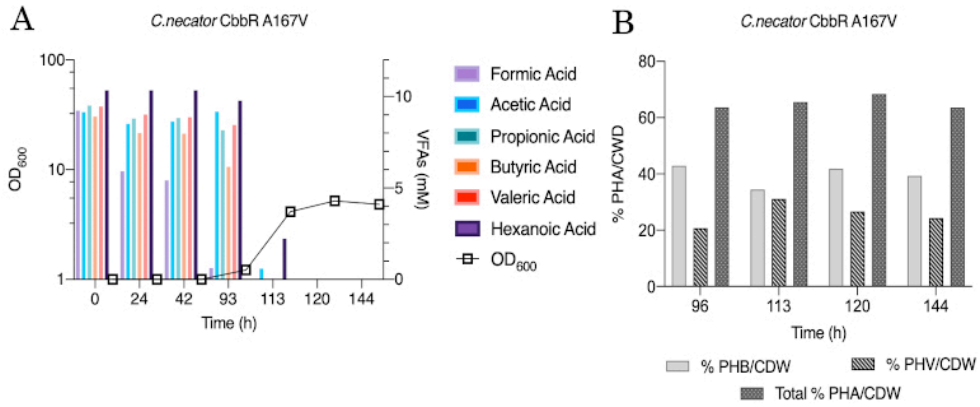


Figure 5.15: Growth, VFAs consumption and PHAs composition of *C. necator* CbbR A167V under aerobic mixotrophic conditions. (A) Growth curve of *C.necator* insertional mutant (*C. necator* CbbR A167V col 2) under aerobic mixotrophic conditions (10 mM VFAs). Second axis representing VFAs concentration (mM) throughout the experiment; (B) Percentage of PHAs and PHAs composition in the biomass from 96 h till stationary phase 144 h (Data collected by Kamran Jawed).

polymers of P(3HB-co-3HV) at almost 50% of 3-HV and 3-HB fraction, more suited to wider range of applications compared to PHB (Wang *et al.*, 2013). To reduce the lag phase and to possibly evaluate co-consumption ability of the mutant strain against the wild-type *C. necator* further experiments were executed.

The scope of this experiment was to compare the co-consumption abilities of the mutant strain, *C. necator* CbbR A167V col2, against wild type *C. necator* H16 strain under mixotrophic conditions. The fermentation was carried out in the bioreactors (DASGIP® Parallel Bioreactor System) with working volume of 0.75 L. The starter cultures of *C. necator* H16 and *C. necator* CbbR A167V were grown in LB gentamicin (10μg/mL) medium and sub-cultured once in 100 mL total volume in order to reach adequate cell density for inoculation of the bioreactor at OD_{600} of 0.2. The cultures were grown in autotrophic conditions for 24 h to build up and adequate biomass density to reduce lag phase and then injected with 5mM VFAs mix to reduce the toxicity of the substrates. During the autotrophic phase of the fermentation we had a rapid growth of the wild type strain in the first 24 h compared to the mutant strain as can be confirmed by the OD_{600} and the consumption for CO_2 and H_2 (Figure 5.16 and 5.17). Moreover, the biomass accumulation at the onset of the stationary phase (60 h) was different between the two reactors ($p<0.05$; One-way ANOVA; data not shown), possibly due to an operator error where a higher concentration of VFAs (about 6 mM) was injected in the wild type strain bioreactor compared to the 5 mM VFAs recorded for the mutant strain. In presence of all organic acids, the first carbon source utilised was formic acid, while the last one was acetic acid for both *C. necator* H16 wild type and mutant strain. This was more visible in the *C. necator* wild type strain were the consumption rate of the organic acids was slower (24 h) compared to mutant strain (18 h), allowing us to collect more time points. This result is consistent with previous work carried out in our laboratory (Jawed and Irorere, unpublished),

where the consumption of acetic acid was observed after the depletion of propionic acid. Furthermore, we observed in *C. necator* H16 consumed formic acid first, followed by the co-consumption of propionic acid with hexanoic acid and butyric acid with valeric acid. While for the mutant strain, it was harder to discern a co-consumption pattern given the lack of data between 36 and 42 h when the organic acids were completely exhausted. The substrate preference observed in our experiment for the *C. necator* H16 do not completely match previous findings (Bhatia *et al.*, 2019). The discrepancy is possibly due to the different VFAs mix and conditions (heterotrophic in *Bhatia et* al. 2019) used in these two studies and the higher concentration of VFAs 5 g/L in *Bhatia et* al. 2019 compared to our experiment 2.5 g/L. Additionally, we recorded an increase in acetic acid in the mutant strain bioreactor at 30 h, possibly due to the decarboxylation of propionate and the metabolization of hexanoic acid *via* β-oxidation. This production of acetate could be beneficial to the microorganism because of the side production of ATP during the reaction from acetyl-CoA to acetate. A similar trend was previously found in the literature for *C. necator* H16 strain (Yan *et al.*, 2002, 2003).

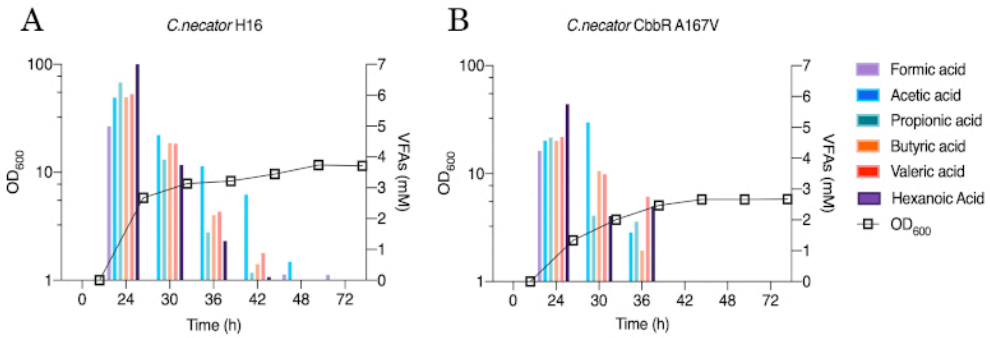


Figure 5.16: Growth and VFAs consumption of *C. necator* H16 and *C. necator* CbbR A167V during two-phase fermentation. Growth curve of (A) *C. necator* H16 and (B) allelic mutant *C. necator* CbbR A167V col 2 under aerobic mixotrophic conditions (5 mM VFAs). Second axis representing VFAs concentration (mM) throughout the experiment.

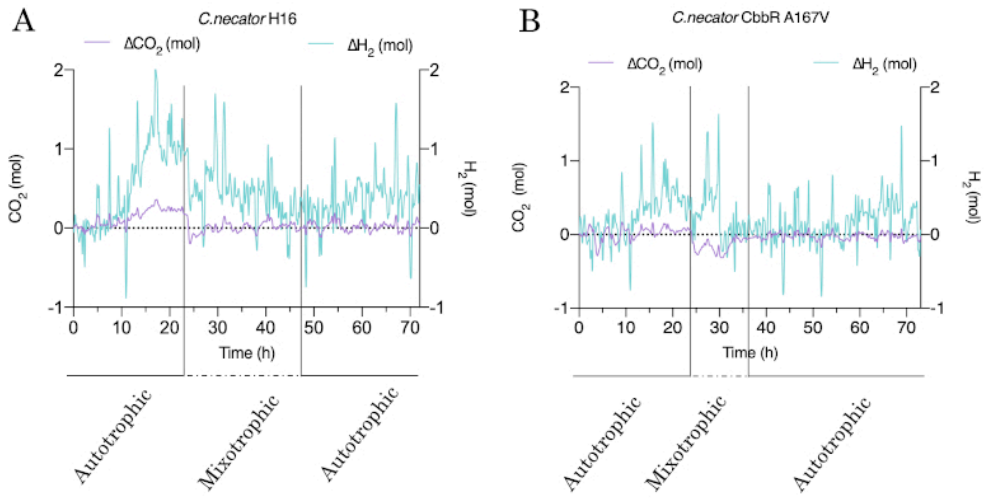


Figure 5.17: Gas consumption profile of *C. necator* H16 and *C. necator* CbbR A167V during two-phase fermentation. Gas profile of H_2, CO_2 and O_2 consumption (mol) for (A) *C. necator* H16 and (B) allelic mutants *C. necator* CbbR A167V col 2 grown under autotrophic and mixotrophic conditions

Ultimately, we could observe a faster uptake of the organic acids for the mutant strain compared to the wild type, but unfortunately, given the high flow rate of the CO_2 in the head space of the fermenter we were not able to confirm the co-consumption ability of the mutant strain.

5.3 Discussion

The main focus of this chapter was to evaluate how the deregulation of the *cbb* operons influence CO_2 assimilation and utilisation efficiency in our chassis organism *C. necator* H16. Firstly, we evaluated the possibility of using an orthologous expression system patented by the University of Nottingham (Minton *et al.*, 2013) to control the expression of the chromosomal copy of the *cbb* operon. The system originally developed for *C. difficile* was adapted in our study for *C. necator,* putting the alternative sigma factor TcdR under the expression of the arabinose inducible promoter (P_{BAD}). Subsequently, the P_{tcdB} promoter would regulate the expression of the targeted genes in an arabinose dependent manner. The functionality of the cassette was evaluated using the expression of a reporter gene (*rfp*) and compared against the activity of the P_{BAD} promoter. The study of orthologous expression cassette (*tcdR*/P_{tcdB}) was attempted at first in *E. coli*. This was performed only at one induction level (10mM L-arabinose) to ensure the correct assembly of the expression vector and control vectors (Figure 5.2 and 5.3). After confirming that the *tcdR*/P_{tcdB} system could be expressed in *E. coli*, the vector encoding for the cassette and the controls were also introduced in *C. necator* H16. Unfortunately, the *tcdR*/P_{tcdB} system wasn't functional in the chassis organism (Figure 5.4 and 5.5) and therefore excluded from future *cbb* operon engineering experiments.

The *tcdR*/P_{tcdB} system overall showed inconsistent results in gram-negative bacteria. It is a possibility that the recognition of the alternative sigma factor *tcdR* was not optimal in either *E. coli* or *C. necator* gram negative strains, given that it originated from a *C. difficile* gram-positive strain.

The deregulation of the *cbb* operons in both LysR independent (P_{j5} or P_{trp}) and LysR dependent (CbbR mutant A167V) constitutive manner was tested to assess if it would prove to be advantageous for the

efficiency of CO_2 fixation in *C. necator* H16. The screening was firstly carried out in organo-autotrophic conditions, where the P_{trp} CBB mutant strain showed less than half of the maximum growth rate ($0.021h^{-1}$) and half of biomass accumulation compared to the wild type strain. The CbbR A167V mutant strains exhibited a behaviour more similar to the wild type, reaching a μ_{max} 0.040 h^{-1} (0.057 h^{-1} for the wild type) and similar biomass accumulation (Figure 5.9 and 5.10). Slower growth rate of the P_{trp} CBB engineered strain might suggest that the constitutive expression of only one copy of the *cbb* operon it isn't enough to match the fitness of the wild type (Srinivasan *et al.*, 2003).

The use of an inducible promoter, P_{BAD}, leading only the chromosomal copy of the *cbb* operon was used to investigate the correlation between different levels of induction (2.5, 5 and 10 mM L-arabinose) of one copy of *cbb* operon and the mutant growth fitness (Figure 5.13). Similar to what we have seen in the case of constitutive expression of the *cbb* operon, the biomass accumulation was half of the value of the control strain, with no difference between increasing arabinose induction levels. The similarity with the P_{trp} mutant regarding the maximum biomass accumulation, despite the difference in μ_{max}, gave further indication of the importance of the activity of both copies of the *cbb* operon to achieve similar wild type growth. We could speculate that in order to maintain a metabolic balance for biomass accumulation, there is an interplay between the dosage effect of the expression of the two operons regulated by the CbbR and its interaction *via* effectors which indicate energy levels and redox state of the cell. This has been previously suggested in the literature in *R. sphaeroides* and *C. necator* (Dubbs *et al.*, 2000; Dangel and Tabita, 2015b; Gruber *et al.*, 2017).

In our study of fermentation in autotrophic conditions, we found an *in vivo* decrease of the doubling time of 1.8-fold for *C. necator* CbbR A167V col 2 and 1.3-fold for col 1, compared against the wild type.

Similarly, the trend was followed by the maximum growth rate with an increase in the values. Both duplicates of the allelic mutant reached the stationary phase at 50 h, while around 65 h were necessary for the wild type strain (Figure 5.14). This result is supported by an earlier consumption of CO_2 of the *C. necator* CbbR A167V strains. An *in vitro* study (Dangel and Tabita, 2015a) showed that the activity of their reporter gene under CbbR mutant A167V control doubled in autotrophic conditions, suggesting doubled expression of the reference gene. Therefore, our result likely indicates a higher activation of the *cbb* operons in this condition, resulting in a higher CO_2 consumption and faster growth.

The ability of the mutant *C. necator* CbbR A167V to produce co-polymers was explored in aerobic mixotrophic conditions using a mix of CO_2, H_2 and VFAs (10 mM C1-C6). The fermentation led to the production of high value copolymers with 50% 3HB/3HV ratio after about 115 h of fermentation (Figure 5.15). As mentioned in the literature (Wang *et al.*, 2013), a higher % of 3HV monomer increases the material ductility and flexibility, giving desirable mechanical and thermal properties. Although, the strain exhibited a long lag phase of 90 h, possibly due to the high concentration of VFAs, at the highest end of *C. necator* tolerability range 5 g/L (Bhatia *et al.*, 2019). Further studies in mixotrophic conditions with lower concentration of VFAs were carried out with the mutant strain comparing its co-consumption ability against the wild type. During this study we were able to observe a faster consumption of VFAs for the mutant strain (Figure 5.16). Unfortunately, the high flow rate of the CO_2 in the head space of the fermenter made the confirmation of the co-consumption ability of the mutant strain very difficult.

Overall, the data in our possession shed some light on the functionality and influence of the mutant CbbR and deletion of transcriptional regulator with the intercistronic promoter region on the expression and activation of the *cbb* operons in *C. necator* H16.

Indeed, even though we were able to observe faster growth of the *C. necator* CbbR A167V mutant strains compared to the wild type in autotrophic conditions, it was not possible to unequivocally demonstrate its co-consumption abilities in aerobic mixotrophic conditions. Further studies will be needed to evaluate and elucidate this option.

5.4 Future work

To further investigate the ability of the *C. necator* CbbR A167V co-consumption ability in aerobic mixotrophic conditions a closed system, as serum bottles, could be used for the purpose. In this condition we could take advantage of a more constant environment to calculate the mol of CO_2 consumed by the mutant strain against the wild type during mixotrophic feeding. Allowing us to understand if indeed in the mutant strain the CO_2 uptake does not stop in the presence of organic compounds.

In the light of recent studies (Jawed and Irorere, unpublished) carried on in our lab, it would also very interesting to further investigate the production abilities of the *C. necator* CbbR A167V strain compared against the wild type in mixotrophic conditions for optimisation of the copolymer P(HB-co-HV) production, using different concentrations of VFAs and percentage of relative organic compounds.

Further work could be carried out in order to achieve complete mixotrophy of the strains, for example replacing both MBH and SH hydrogenases promoters with constitutive ones or alternatively mutating the highly conserved Asp55 in the *hoxA* gene as suggested in Lenz *et al.*(2002) to avoid the catabolite repression operated by the organic acids in the media.

An alternative strategy could be also applied, taking inspiration from the plant realm, introducing non-native pathways to recycle the products of RubisCO oxygenation, 2-PGA, more efficiently. In several

studies photorespiration bypasses were mentioned for crop improvement (Bar-Even, 2018; Weber and Bar-Even, 2019), proving their relevance during field studies where higher biomass accumulation and photosynthetic rates were achieved successfully in tobacco (South *et al.*, 2019). In our model organism, the glycolate oxidase operon (*glcD1*, *glcE* and *glcF*) could be replaced with *Chlamydomonas reinhardtii* glycolate dehydrogenase (CrGDH) to avoid H_2O_2 production when the glycolate is converted to glyoxylate.

Chapter 6 – General discussion

6.1 General discussion and future directions

In this study I evaluated the use of two main strategies to improve carbon fixation in our chassis organism, *C. necator* H16. In the first strategy, it was tested whether α-carboxysomes from *H. neapolitanus* could be heterologously expressed and function in *C. necator* H16 (Chapter 3). The successful recovery and visualization of the carboxysomal components in the membrane fractions of the whole-cell protein samples in both the control strain, *E. coli*, and *C. necator* suggested their correct assembly (Figure 3.1 to 3.4). To corroborate the results, the four major protein bands present in the *C. necator* sample, were identified as carboxysomal components by MALDI-TOF-MS protein identification analysis. Altogether, these results strongly suggest that the *H. naepolitanus cso* operon was successfully expressed in *C. necator* H16. Furthermore, the correct structural assembly of the α-carboxysome shell proteins in both *E. coli* and *C. necator* was evaluated using TEM imaging. A higher image definition was achieved in *C. necator* with the strain constitutively expressing the *cso* operon, where the transversal sections of the strain's microscopy slices allowed the identification of the CB structures in their closed conformation (Figure 3.7 C). To the best of our knowledge, this is the first report in the literature to demonstrate the heterologous expression and assembly of α-carboxysomes in this bacterial species. These results, together with the previous literature (Bonacci *et al.*, 2012; Cai *et al.*, 2016; Baumgart *et al.*, 2017) suggests the flexibility of the carboxysome as a bioengineering tool for heterologous expression in a variety of bacterial strains.

Seeking to use this CCM to implement carbon fixation in *C. necator*, I investigated the *in vivo* functionality of the α-carboxysomes in our RubisCO null-mutant *C. necator* strain under both organo-autotrophic and autotrophic conditions (Figure 3.10 and 3.11).

Unfortunately, the inability of the strain to grow in these conditions showed that, despite their expression and assembly in a closed conformation, CBs were ultimately not functional in *C. necator* H16. As recently suggested (Turmo *et al.*, 2017; Desmarais *et al.*, 2019), our result appears to provide additional evidence in support of the existence of a range of ancillary proteins in the genomic surroundings of the *cso* operon, necessary for their function. One of the key components of the carboxysome *superloci* is the RubisCO activases, able to remove the inhibitory substrates and recycle active RubisCO inside of the carboxysome shell (Sutter *et al.*, 2015; Tsai *et al.*, 2015). The expression of the CB-encoding genes in combination with these RubisCO activases restored the ability of *C. necator* ΔΔ*cbbLS* to grow under organo-autotrophic conditions (Figure 3.18), suggesting the *in vivo* functionality of the α-carboxysomes in the heterologous host *C. necator* H16. However, the lack of a co-localisation experiment and distinct carboxysomal expression bands in the membrane fraction of the *C. necator* RubisCO null mutant strains (under both heterotrophic and organo-autotrophic conditions), leaves room to speculate that this growth complementation may have been the result of RubisCO expression in the cytosol of *C. necator* cells, independent from the CBs. It must be mentioned that the TEM imaging data seem to suggest the presence of heterologously expressed CB structures in both *C. necator* PHB-producing and PHB-negative (Δ*phaCAB*) strains harboring the *cso* operon in combination with the RubisCO activases genes (Figure 3.15 and 3.16). Although despite our best efforts, the images' resolution was not high enough to assess the structural integrity and the architecture of the CB structures.

To overcome this limitation, it would be interesting to take advantage of recent developments in the technological field of microscopy, following the steps of Dai *et al.* (2018), and use Cryo-Electron tomography to further investigate the integrity of the carboxysomal micro-structures, as well as their possible co-localisation with

RubisCO, in *C. necator* strains co-expressing carboxysomes and the RubisCO activases. Another strategy to rule out the activity of the RubisCO in the cytoplasmic compartment, would be to remove the *csoSCA* gene from the *cso* operon. In the event that the resulting strain can no longer restore growth on CO_2, this would suggest that the RubisCO is solely active within the CB shell.

At the end of our experimental work, a paper (Flamholz *et al.*, 2020) confirmed that the expression of 20 genes from *H. naepolitanus* was essential to enabled α-CBs functionality in *E. coli*. In their research, the expression of the canonical *cso* operon was coupled with the RubisCO associated proteins (acRAF and CsoCbbOQ) and inorganic carbon uptake operon (*dab*AB1). This and previous work (Desmarais *et al.*, 2019), confirm that the RubisCO activases are essential for the functionality of the CB in an heterologous host. I showed that in *C. necator Hn*RubisCO folding and activity was independent of acRAF, and therefore not included in the heterologously expressed operon. It is possible that the inclusion and expression of the inorganic carbon transporter (*dab*AB1) could increase the C_i uptake efficiency, enabling *C. necator* strain to grow at atmospheric CO_2 concentration.

Until now, in structural computational studies, attention has been focused on the hydrophobic and charged interactions for the stability of the protein core (Vascon *et al.*, 2020). While, the role of the surface charge and its relevance during molecular and macromolecular assembly was overlooked. The impact of electrostatic potential has been widely studied in protein stability, catalytic activity and protein binding to molecules (eg., RNA or DNA) (Tsai *et al.*, 2016; Zhou and Pang, 2018), but only recently in protein-protein interactions (Mihiri Shashikala *et al.*, 2019; Vascon *et al.*, 2020). To further promote innovation, it was decided to explore the relevance of the RubisCO electrostatic potential as a driving force for the recruitment of the holoenzyme inside of the α-carboxysome shells (Chapter 4). An *in-*

silico C. necator mutant RubisCO (*CnMut*RubisCO) resembling the *H. neapolitanus* RubisCO (*Hn*RubisCO) was created, with a more negative electrostatic potential at the top and back of the protein structure. Analysing the MD trajectories, it was observed the fluctuation of the whole protein backbone against the reference structure (RMSD) and against the average residue position (RMSF) of the *CnMut*RubisCO compared to the parental RubisCO (*Cn*RubisCO) and to the *Hn*RubisCO at different ionic strengths (RMSD Figure 4.4 and RMSF Figure 4.5). The RMSD analysis suggested that *Hn*RubisCO at 20 mM and *Cn*RubisCO at 150 mM, were less stable than in other conditions possibly indicating their tendency to bind another protein (Singh *et al.*, 2016), such as other RubisCO dimers or recruiting peptides from CsoS2 (Bracher *et al.*, 2017; Oltrogge *et al.*, 2020). As confirmed by the RMSF analysis, the movement was mostly localised in the flexible loop region (*Hn*: residues 57–71; *Cn*: residues 68-80) of both proteins, located near to key residues for the catalytic activity of RubisCO (Andersson, 2008; Bhat *et al.*, 2017). In contrast, the *CnMut*RubisCO showed an increased stability at both 20 mM and 150 mM, compared to both *Cn*RubisCO and *Hn*RubisCO. This result was corroborated by the restricted movements observed in the flexible loop (residues 70–83), during RMSF and visual analysis of the trajectories (Figure 4.6). We hypothesised that the movement restriction could be caused by the mutations (Arg76Leu; Ala79Asp) of the protein structure and negatively affect the holoenzyme activity. However, despite the fact that the in-depth implications of this structural stiffness on the functionality are still unknown, it did not affect the *in vivo* catalytic activity of the *CnMut*RubisCO (Figure 4.15).

Moreover, when the interaction between the solvent and the protein backbone were investigated using the RDF analysis of the modelled MD trajectories, I showed that *Hn* and *Cn*RubisCO oppositely interact with cations and anions in solution (Figure 4.8 and 4.9). The

*Hn*RubisCO exhibited a high interaction affinity with Na^+ ions at 20 and 150 mM compared to *Cn*, where the peaks were found to be three times higher. While, the *CnMut*RubisCO demonstrated a remarkable similarity to *Hn*RubisCO in the interaction intensity with both cations and anions. The interaction with the ions observed for the *CnMut*RubisCO during RDF analysis, is very promising regarding our attempt to modify the surface potential of the *Cn*RubisCO towards behaving more like *Hn*RubisCO.

Another outcome of this work is the striking similarity, during the ID analysis, in the distribution of the Na^+ ions around both *Hn*RubisCO and *CnMut*RubisCO at both 20 mM and 150 mM, especially at the top and at the bottom of the structures (Figure 4.10). Where, in both protein variants at 20 mM and 150 mM, it was observed cation density clouds around the residues shown to interact with the recruiting peptides of the CsoS2 protein (Oltrogge *et al.*, 2020) (*Hn*: CbbL 22–28, 69–72, 346–349, 358–363 and CbbS 94–98; *CnMut*: CbbL 35–37, 77–83, 355–357, 367–370 and CbbS 86–87). For the *CnMut*RubisCO, I identified in one of the interaction zones, residues Asp36, Asp79 and possibly Try83, that could provide the salt bridge and H-bond interaction necessary to bind the recruiting peptides. Unfortunately, in the second interaction location the *CnMut*RubisCO offers only three out of four conserved residues considered necessary for interaction with the recruiting peptides (CbbL Tyr356, Asp370 and CbbS Asp86). To further validate the importance of said peptide-protein electrostatic interactions (Oltrogge *et al.*, 2020), the truncation of the CsoS2 N-terminal entirely prohibits the CBs formation, during *in vivo* experiments (Flamholz *et al.*, 2020).

These promising results in RDF and IDA analysis highlighted how the redistribution of the surface electrostatic potential of *CnMut*RubisCO, overall changed the protein's short- and long-range interactions with the ions during the MD simulation. Short- and long-range interactions were shown to be a key factor during protein-

protein recruitment mechanisms (Zhang *et al.*, 2011; Mihiri Shashikala *et al.*, 2019; Vascon *et al.*, 2020). Since the CsoS2 peptide recruitment regions are positively charged (Oltrogge *et al.*, 2020), these changes in behaviour of *CnMut*RubisCO, especially with Na^+ ions, could lead to suitable forces for the recruitment within the CB.

Overall despite the promising results in Chapter 4, our data are not conclusive enough to demonstrate that an encapsulation of *CnMut*RubisCO in the α-carboxysomes would be possible. Therefore, an interesting continuation of the work would be the creation of a two chimeric *cso*-operons carrying either the *CnMut*RubisCO or *Cn*RubisCO in *C. necator* H16. Where, a similar experimental approach to Dai *et al.* (2018) using Cryo-Electron tomography could be used to further investigate the carboxysome assembly mechanism *in vivo* and the possible co-localisation of the chimeric *CnMut*RubisCO with the shell proteins. Since a recent study (Oltrogge *et al.*, 2020) highlighted the interaction residues between the recruiting peptides and the backbone of *Hn*RubisCO, it would be interesting to further modify the *CnMut*RubisCO structure to match the missing residue interactions (*Cn*: Met82Tyr, Try356Phe and Thr88Tyr). It would be expected that these additional modifications would further change the electrostatic potential of the protein structure, possibly expanding the cations density cloud during MD simulations around the conserved residues involved the RubisCO recruitment and hopefully increasing the chances of *CnMut*RubisCO *in vivo* encapsulation.

In the second strategy to improve carbon fixation, I evaluated the influence on CO_2 assimilation by the deregulation of the *cbb* operons in a LysR independent and LysR dependent manner, in *C. necator* H16. We assessed the use of an orthologous expression system patented by the University of Nottingham (Minton *et al.*, 2013), to control the expression of the chromosomal copy of the *cbb* operon in a tight inducible manner. Unfortunately, after confirming the correct

function of the *tcdR*/P*tcdB* system in *E. coli* (Figure 5.2 and 5.3), the vectors encoding the cassette and the controls were also introduced in *C. necator* H16, where it was found to be not functional (Figure 5.4 and 5.5). Thus, the *tcdR*/P*tcdB* system was excluded from future *cbb* operon engineering experiments.

The expression of the chromosomal *cbb* operon in a constitutive (P*trp*) and inducible (P*BAD*) LysR independent manner under organo-autotrophic conditions led to halved values of biomass accumulation compared to the control strain, with no difference between increasing arabinose induction levels (Figure 5.9 and 5.13, respectively). The similarity between the two mutant strains regarding the maximum biomass accumulation, despite the difference in μ_{max}, highlighted the relevance of the activity of both copies of the *cbb* operon to achieve optimal growth, similar to the wild type strain. Interestingly, we could speculate that in order to maintain a metabolic balance for biomass accumulation, there is an interplay between the dosage effect of the expression of the two operons regulated by the CbbR and its interaction *via* effectors with the central metabolism (Dubbs *et al.*, 2000; Dangel and Tabita, 2015b; Gruber *et al.*, 2017).

I also observed that the expression of both *cbb* operon copies in a constitutive LysR dependent manner (CbbR A167V) led to a maximum biomass accumulation similar to the wild type strain under both organo-autotrophic conditions and aerobic autotrophic fermentation (Figure 5.10 and 5.14, respectively). Furthermore, during autotrophic fermentation a decrease of the doubling time (1.5-fold) was recorded for *C. necator* CbbR A167V and faster reach of stationary phase, compared to the wild type strain. As previously reported for the CbbR A167V mutation in an *in vitro* study (Dangel and Tabita, 2015a), this result likely indicates a higher activation of the *cbb* operons *in vivo* under autotrophic conditions, which results in an earlier CO_2 uptake and faster entrance into exponential growth.

Further fermentations were carried out for the CbbR A167V mutant strain under aerobic mixotrophic conditions. The ability of the strain to produce higher value co-polymers (3HB/3HV) was successfully tested under mixotrophic conditions, where the 50% 3HB/3HV ratio was reached at 115 h (Figure 5.15), after a lag phase of 90 h possibly due to the high VFA (10 mM) concentrations at the higher end of the tolerability range for *C. necator* (Bhatia *et al.*, 2019). While, the co-consumption ability of the CbbR A167V mutant strain unfortunately appears less clear. Because of the high flow rate (45 l/h) of the CO_2 in the head space of the fermenter it was not possible to distinguish differences in CO_2 consumption between the mutant and the wild type control strain (Figure 5.17).

In order to overcome this experimental limitation, the co-consumption experiment could be carried out in a closed system, such as serum bottles, where the mol of CO_2 consumed under mixotrophic feeding could be calculated, in a more constant environment. Moreover, to allow the strain to reach complete mixotrophy, further genomic modifications would be required to avoid the catabolite repression of the *hox* and *hyp* regulon, by organic acids in the media.

6.2 Concluding remarks

To conclude, this PhD project set out to investigate potential strategies to improve CO_2 fixation in *C. necator* H16. The two aims of this study were: 1) attempt heterologous expression of α-carboxysomes in *C. necator* 2) bypass the natural regulation of the CBB pathway in order to assess the benefit of placing the *cbb* operons under LysR independent and LysR dependent constitutive transcription.

I have shown that it is possible to heterologously express the *cso* operon in *C. necator*, however the functionality of the α-carboxysomes could not be confirmed. The relevance of the accessory proteins, RubisCO activases, was further highlighted in this thesis, providing

a valuable proof of concept for additional work in our host strain. Furthermore, after the modification of the surface charge of the *Cn*RubisCO *in-silico* to resemble *Hn*RubisCO, MD simulations were used to observe changes in the protein interactions with ions in aqueous phase. This initial work provides a foundation to better understand the forces involved in RubisCO encapsulation during the carboxysome self-assembly process.

I also discovered that in *C. necator* the expression of both copies of the *cbb* operon is advisable to match the maximum biomass accumulation of the wild type under autotrophic conditions. I hypothesise an interplay between the dosage effect of the expression of the two operons regulated by the CbbR and its interaction *via* effectors with the central carbon metabolism. It was also observed that a single amino acid mutation in the CbbR, was able to induce a reduced lag phase and faster entrance into exponential phase, in the mutant strain, possibly due to the constitutive expression of the both *cbb* operons and absence of catabolite feedback.

All of these findings improve our understanding of the complex mechanisms regulating carbon assimilation and energy balance in *C. necator* H16 and generates a valuable proof of concept to further improve carbon fixation.

References

Abascal, F. *et al.* (2005) 'ProtTest: selection of best-fit models of protein evolution', *Bioinformatics*, 21(9), pp. 2104–2105. doi: 10.1093/bioinformatics/bti263.

Alagesan, S., Hanko, E. K. R., Malys, N., Ehsaan, M., Winzer, K. and Minton, N. P. (2018) 'Functional genetic elements for controlling gene expression in Cupriavidus necator H16', *Applied and Environmental Microbiology*, 84(19), pp. 1–17. doi: 10.1128/AEM.00878-18.

Alagesan, S., Minton, N. P. and Malys, N. (2018) 'C-assisted metabolic flux analysis to investigate heterotrophic and mixotrophic metabolism in Cupriavidus necator H16', *Metabolomics*, 14(9), pp. 2–10.

Alberty, A. R. (2003) 'Thermodynamics of Biochemical Reactions', in *Thermodynamics of Biochemical Reactionsz*. John Wiley & Sons, Ltd. doi: 10.1002/0471332607.fmatter.

Anandakrishnan, R., Aguilar, B. and Onufriev, A. V (2012) 'H+ + 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations', *Nucleic Acids Research*, 40, pp. W537–W541. doi: 10.1093/nar/gks375.

Andersson, I. (2008) 'Catalysis and regulation in Rubisco', *Journal of Experimental Botany*, 59(7), pp. 1555–1568. doi: 10.1093/jxb/ern091.

Arrhenius, S. (1896) 'On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground', *Philosophical Magazine and Journal of Science Series*, 5, pp. 237–276.

Ashida, H., Saito, Y., Kojima, C., Kobayashi, K., Ogasawara, N. and Yokota, A. (2003) 'A Functional Link Between RuBisCO-like Protein of Bacillus and Photosynthetic RuBisCO', *Science*, 302(5643), pp. 286–290. doi: 10.1126/science.1086997.

Ashida, H., Danchin, A. and Yokota, A. (2005) 'Was photosynthetic RuBisCO recruited by acquisitive evolution from RuBisCO-like proteins involved in sulfur metabolism?', *Research in Microbiology*, 156, pp. 611–618. doi: 10.1016/j.resmic.2005.01.014.

Axen, S. D., Erbilgin, O. and Kerfeld, C. A. (2014) 'A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method', *PLoS Computational Biology*, 10(10). doi: 10.1371/journal.pcbi.1003898.

Badger, M. R. and Bek, E. J. (2008a) 'Multiple Rubisco forms in proteobacteria: Their functional significance in relation to CO2 acquisition by the CBB cycle', *Journal of Experimental Botany*, 59(7), pp. 1525–1541. doi: 10.1093/jxb/erm297.

Badger, M. R. and Bek, E. J. (2008b) 'Multiple Rubisco forms in proteobacteria: Their functional significance in relation to CO2 acquisition by the CBB cycle', *Journal of Experimental Botany*, 59(7), pp. 1525–1541. doi: 10.1093/jxb/erm297.

Badger, M. R., Hanson, D. and Price, G. D. (2002) 'Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria', *Functional Plant Biology*, 29(29), p. 161. doi: 10.1071/PP01213.

Badger, M. R. and Price, G. D. (1992) 'The CO2 concentrating mechanism in cyanobactiria and microalgae', *Physiologia Plantarum*, 84(4), pp. 606–615. doi: 10.1111/j.1399-3054.1992.tb04711.x.

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

Bar-Even, A. (2018) 'Daring metabolic designs for enhanced plant carbon fixation', *Plant Science*. doi: 10.1016/j.plantsci.2017.12.007.

Bassham, J. A., Benson, A. A., Kay, L. D., Harris, A. Z., Wilson, A. T. and Calvin, M. (1954) 'The Path of Carbon in Photosynthesis. XXI. The Cyclic Regeneration of Carbon Dioxide Acceptor', *Journal of the American Chemical Society*, 76(7), pp. 1760–1770. doi: 10.1021/ja01636a012.

Baumgart, M., Huber, I., Abdollahzadeh, I., Gensch, T. and Frunzke, J. (2017) 'Heterologous expression of the Halothiobacillus neapolitanus carboxysomal gene cluster in Corynebacterium glutamicum', *Journal of Biotechnology*, (December 2016), pp. 1–10. doi: 10.1016/j.jbiotec.2017.03.019.

Berg, I. A., Kockelkorn, D., Buckel, W. and Fuchs, G. (2007) 'A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea', *Science*, 318(5857), pp. 1782–1786. doi: 10.1126/science.1149976.

Berg, I. A. (2011) 'Ecological aspects of the distribution of different autotrophic CO2 fixation pathways', *Applied and Environmental Microbiology*, 77(6), pp. 1925–1936. doi: 10.1128/AEM.02473-10.

Berner, R. A., Beerling, D. J., Dudley, R., Robinson, J. M. and Wildman, R. A. (2003) 'Phanerozoic atmospheric oxygen', *Annu. Rev. Earth Planet. Sci*, 31, pp. 105–34. doi: 10.1146/annurev.earth.31.100901.141329.

Berner, R. A. and Kothavala, Z. (2001) 'GEOCARB III: A revised model of atmospheric CO2 over Phanerozoic time', *American Journal of Science*, 301(2), pp. 182–204. doi: 10.2475/ajs.301.2.182.

Bhat, J. Y., Thieulin-Pardo, G., Hartl, F. U. and Hayer-Hartl, M. (2017) 'Rubisco activases: AAA+ chaperones adapted to enzyme repair', *Frontiers in Molecular Biosciences*, 4(APR), pp. 1–10. doi: 10.3389/fmolb.2017.00020.

Bhatia, S. K., Gurav, R., Choi, T. R., Jung, H. R., Yang, S. Y., Song, H. S., Jeon, J. M., Kim, J. S., Lee, Y. K. and Yang, Y. H. (2019)

'Poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) production from engineered Ralstonia eutropha using synthetic and anaerobically digested food waste derived volatile fatty acids', *International Journal of Biological Macromolecules*, 133, pp. 1–10. doi: 10.1016/j.ijbiomac.2019.04.083.

Bi, C., Su, P., Müller, J., Yeh, Y.-C., Chhabra, S. R., Beller, H. R., Singer, S. W. and Hillson, N. J. (2013) 'Development of a broad-host synthetic biology toolbox for ralstonia eutropha and its application to engineering hydrocarbon biofuel production', *Microbial Cell Factories*, 12(1), p. 107. doi: 10.1186/1475-2859-12-107.

Biebl, H. and Pfennig, N. (1981) 'Isolation of Members of the Family Rhodospirillaceae', in *The Prokaryotes*. Springer Berlin Heidelberg, pp. 267–273. doi: 10.1007/978-3-662-13187-9_14.

Bonacci, W., Teng, P. K., Afonso, B., Niederholtmeyer, H., Grob, P. and Silver, P. a (2012) 'Modularity of a carbon- fixing protein organelle', *Pnas*, 109(2), pp. 478–483. doi: 10.1073/pnas.1108557109.

Bowien, B. and Kusian, B. (2002) 'Genetics and control of CO2 assimilation in the chemoautotroph Ralstonia eutropha', *Archives of Microbiology*, 178(2), pp. 85–93. doi: 10.1007/s00203-002-0441-3.

Boyle, N. R. and Morgan, J. A. (2011) 'Computation of metabolic fluxes and efficiencies for biological carbon dioxide fixation', *Metabolic Engineering*, 13, pp. 150–158. doi: 10.1016/j.ymben.2011.01.005.

Bracher, A., Sharma, A., Starling-Windhof, A., Hartl, F. U. and Hayer-Hartl, M. (2015) 'Degradation of potent Rubisco inhibitor by selective sugar phosphatase', 1, p. 14002.

Bracher, A., Whitney, S. M., Hartl, F. U. and Hayer-Hartl, M. (2017) 'Biogenesis and Metabolic Maintenance of Rubisco', *Annual Review of Plant Biology*, 68(December 2016), pp. 29–60. doi: 10.1146/annurev-arplant-043015-111633.

Brigham, C J, Zhila Shishatskaya, N. E., Volova, B. T., Sinskey, A J, Brigham, Christopher J, Zhila, N., Shishatskaya, E., Volova, T. G. and Sinskey, Anthony J (2012) 'Manipulation of Ralstonia eutropha Carbon Storage Pathways to Produce Useful Bio-Based Products', *Subcellular Biochemistry*, 64. doi: 10.1007/978-94-007-5055-5_17.

Cai, F., Heinhorst, S., Shively, J. M. and Cannon, G. C. (2008) 'Transcript analysis of the Halothiobacillus neapolitanus cso operon', *Arch Microbiol*, 189, pp. 141–150. doi: 10.1007/s00203-007-0305-y.

Cai, F., Menon, B. B., Cannon, G. C., Curry, K. J., Shively, J. M. and Heinhorst, S. (2009) 'The Pentameric Vertex Proteins Are Necessary for the Icosahedral Carboxysome Shell to Function as a CO2 Leakage Barrier', *PLoS ONE*. Edited by N. Ahmed, 4(10), p. e7521. doi: 10.1371/journal.pone.0007521.

Cai, F., Dou, Z., Bernstein, S. L., Leverenz, R., Williams, E. B., Heinhorst, S., Shively, J., Cannon, G. C., Kerfeld, C. A., Meeks, J. C. and Haselkorn, R. (2015) 'Advances in Understanding Carboxysome Assembly in Prochlorococcus and Synechococcus Implicate CsoS2 as a Critical Component', *Life*, 5(2), pp. 1141–1171. doi: 10.3390/life5021141.

Cai, F., Sutter, M., Bernstein, S. L., Kinney, J. N. and Kerfeld, C. A. (2015) 'Engineering bacterial microcompartment shells: Chimeric shell proteins and chimeric carboxysome shells', *ACS Synthetic Biology*, 4(4), pp. 444–453. doi: 10.1021/sb500226j.

Cai, F., Bernstein, S. L., Wilson, S. C. and Kerfeld, C. A. (2016) 'Production and Characterization of Synthetic Carboxysome Shells with Incorporated Luminal Proteins.', *Plant physiology*, 170(3), pp. 1868–77. doi: 10.1104/pp.15.01822.

Cai, F., Kerfeld, C. A. and Sandh, G. (2012) 'Bioinformatic Identification and Structural Characterization of a New Carboxysome Shell Protein', in *Functional Genomics and Evolution*

of Photosynthetic Systems . Springer, Dordrecht, pp. 345–356. doi: 10.1007/978-94-007-1533-2_14.

Cameron, J. C., Wilson, S. C., Bernstein, S. L. and Kerfeld, C. A. (2013) 'Biogenesis of a bacterial organelle: The carboxysome assembly pathway', *Cell*, 155(5), pp. 1131–1140. doi: 10.1016/j.cell.2013.10.044.

Campbell, R. E., Tour, O., Palmer, A. E., Steinbach, P. A., Baird, G. S., Zacharias, D. A. and Tsien, R. Y. (2002) 'A monomeric red fluorescent protein', *Pnas*, 99(12), pp. 7877–7882. doi: 10.1073/pnas.082243699.

Cannon, G. C., Baker, S. H., Soyer, F., Johnson, D. R., Bradburne, C. E., Mehlman, J. L., Davies, P. S., Jiang, Q. L., Heinhorst, S. and Shively, J. M. (2003) 'Organization of carboxysome genes in the thiobacilli', *Current Microbiology*, 46(2), pp. 115–119. doi: 10.1007/s00284-002-3825-3.

Cannon, G. C., Heinhorst, S. and Kerfeld, C. A. (2010) 'Carboxysomal carbonic anhydrases: Structure and role in microbial CO2 fixation', *BBA - Proteins and Proteomics*, 1804, pp. 382–392. doi: 10.1016/j.bbapap.2009.09.026.

Cegelski, L. and Schaefer, J. (2006) 'NMR determination of photorespiration in intact leaves using in vivo 13CO2 labeling', *Journal of Magnetic Resonance*, 178(1), pp. 1–10. doi: 10.1016/j.jmr.2005.10.010.

Chen, D. H., Madan, D., Weaver, J., Lin, Z., Schröder, G. F., Chiu, W. and Rye, H. S. (2013) 'Visualizing GroEL/ES in the act of encapsulating a folding protein', *Cell*, 153(6), p. 1354. doi: 10.1016/j.cell.2013.04.052.

Chen, J. S., Colón, B., Dusel, B., Ziesack, M., Way, J. C. and Torella, J. P. (2015) 'Production of fatty acids in Ralstonia eutropha H16 by engineering ß-oxidation and carbon storage.', *PeerJ*, 3, p. e1468. doi:

10.7717/peerj.1468.

Claassens, N. J., Bordanaba-Florit, G., Cotton, C. A. R., De Maria, A., Finger-Bou, M., Friedeheim, L., Giner-Laguarda, N., Munar-Palmer, M., Newell, W., Scarinci, G., Verbunt, J., de Vries, S. T., Yilmaz, S. and Bar-Even, A. (2020) 'Replacing the Calvin cycle with the reductive glycine pathway in Cupriavidus necator', *Metabolic Engineering*, 62, pp. 30–41. doi: 10.1016/j.ymben.2020.08.004.

Codd, G. A. and Kuenen, J. G. (1987) 'Physiology and Biochemistry of Autotrophic Bacteria', in van Verseveld, H. W. and Duine, J. A. (eds) *Microbial Growth on C1 Compounds*. Springer, Dordrecht. doi: https://doi.org/10.1007/978-94-009-3539-6_2.

Cot, S. S.-W., So, A. K.-C. and Espie, G. S. (2008) 'A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria.', *Journal of bacteriology*, 190(3), pp. 936–45. doi: 10.1128/JB.01283-07.

Cotton, C. A., Edlich-Muth, C. and Bar-Even, A. (2018) 'Reinforcing carbon fixation: CO2 reduction replacing and supporting carboxylation', *Current Opinion in Biotechnology*, 49, pp. 49–56. doi: 10.1016/j.copbio.2017.07.014.

Cramm, R. (2009) 'Genomic view of energy metabolism in Ralstonia eutropha H16', *Journal of Molecular Microbiology and Biotechnology*, pp. 38–52. doi: 10.1159/000142893.

Dai, W., Chen, M., Myers, C., Ludtke, S. J., Pettitt, B. M., King, J. A., Schmid, M. F. and Chiu, W. (2018) 'Visualizing Individual RuBisCO and Its Assembly into Carboxysomes in Marine Cyanobacteria by Cryo-Electron Tomography', *Journal of Molecular Biology*, 430(21), pp. 4156–4167. doi: 10.1016/j.jmb.2018.08.013.

Dangel, A. W. and Tabita, F. R. (2009) 'Protein-protein interactions between CbbR and RegA (PrrA), transcriptional regulators of the *cbb* operons of *Rhodobacter sphaeroides*', *Molecular Microbiology*,

71(3), pp. 717–729. doi: 10.1111/j.1365-2958.2008.06558.x.

Dangel, A. W. and Tabita, F. R. (2015a) 'Amino acid substitutions in the transcriptional regulator CbbR lead to constitutively active CbbR proteins that elevate expression of the cbb CO2 fixation operons in Ralstonia eutropha (Cupriavidus necator) and identify regions of CbbR necessary for gene', *Microbiology*, 161(9), pp. 1816–1829. doi: 10.1099/mic.0.000131.

Dangel, A. W. and Tabita, F. R. (2015b) 'CbbR, the Master Regulator for Microbial Carbon Dioxide Fixation', *Journal of Bacteriology*, 197, pp. 3488–3498. doi: 10.1128/JB.00442-15.

Desmarais, J. J., Flamholz, A. I., Blikstad, C., Dugan, E. J., Laughlin, T. G., Oltrogge, L. M., Chen, A. W., Wetmore, K., Diamond, S., Wang, J. Y. and Savage, D. F. (2019) 'DABs are inorganic carbon pumps found throughout prokaryotic phyla', *Nature Microbiology*, 4(12), pp. 2204–2215. doi: 10.1038/s41564-019-0520-8.

Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. and Stewart, J. J. P. (1985) 'AM1: A New General Purpose Quantum Mechanical Molecular Model1', *Journal of the American Chemical Society*, 107(13), pp. 3902–3909. doi: 10.1021/ja00299a024.

Dobrinski, K. P., Longo, D. L. and Scott, K. M. (2005) 'The carbon-concentrating mechanism of the hydrothermal vent chemolithoautotroph Thiomicrospira crunogena.', *Journal of bacteriology*, 187(16), pp. 5761–6. doi: 10.1128/JB.187.16.5761-5766.2005.

Dou, Z., Heinhorst, S., Williams, E. B., Murin, C. D., Shively, J. M. and Cannon, G. C. (2008) 'CO2 fixation kinetics of Halothiobacillus neapolitanus mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO2', *Journal of Biological Chemistry*, 283(16), pp. 10377–10384. doi:

10.1074/jbc.M709285200.

Drew, G. and Niklowitz, W. (1956) 'Cytology of Cyanophycea. II. Centroplasm and granular inclusions of Phormidium uncinatum.', *Archiv fur Mikrobiologie*, 24(2), pp. 147–62.

Du, H., Liu, Z., Jennings, R. and Qian, X. (2016) 'The effects of salt ions on the dynamics and thermodynamics of lysozyme unfolding', *Separation Science and Technology*. doi: 10.1080/01496395.2016.1229336.

Dubbs, J. M., Bird, T. H., Bauer, C. E. and Robert Tabita, F. (2000) 'Interaction of CbbR and RegA*Transcription Regulators with the Rhodobacter sphaeroides cbb I Promoter-Operator Region*', *The Journal of Biological Chemistry*, 275(25), pp. 19224–19230. doi: 10.1074/jbc.M002125200.

Dupuy, B. and Matamouros, S. (2006) 'Regulation of toxin and bacteriocin synthesis in Clostridium species by a new subgroup of RNA polymerase σ-factors', *Research in Microbiology*, 157(3), pp. 201–205. doi: 10.1016/j.resmic.2005.11.004.

Edernhofer, O., Pichs-Madruga, R., Sokona, Y., Minx, J. C., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlömer, S., Von Stechow, C. and Zwickel, T. (2014) *Climate Change 2014 Mitigation of Climate Change Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Evans, M. C., Buchanan, B. B. and Arnon, D. I. (1966) 'A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium.', *Pnas*, 55(4), pp. 928–34.

Farago, O. (2019) 'Langevin thermostat for robust configurational and kinetic sampling', *Physica A: Statistical Mechanics and its Applications*, 534, p. 122210. doi: 10.1016/j.physa.2019.122210.

Finn, M. W. and Tabita, F. R. (2004) 'Modified Pathway To Synthesize Ribulose 1,5-Bisphosphate in Methanogenic Archaea', *Journal of Bacteriology*, 186(19), pp. 6360–6366. doi: 10.1128/JB.186.19.6360-6366.2004.

Flamholz, A., Dugan, E., Blikstad, C., Gleizer, S., Ben-Nissan, R., Amram, S., Antonovsky, N., Ravishankar, S., Noor, E., Bar-Even, A., Milo, R. and Savage, D. (2020) 'Functional reconstitution of a bacterial CO2 concentrating mechanism in E. coli', *eLife*, 9, p. e59882. doi: 10.7554/eLife.59882.

Freeman-Rosenzweig, E. S., Xu, B., Kuhn Cuellar, L., Martinez-Sanchez, A., Schaffer, M., Strauss, M., Cartwright, H. N., Ronceray, P., Plitzko, J. M., Förster, F., Wingreen, N. S., Engel, B. D., Mackinder, L. C. M. and Jonikas, M. C. (2017) 'The Eukaryotic CO2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization', *Cell*, 171(1), pp. 148-162.e19. doi: 10.1016/j.cell.2017.08.008.

Fuhrmann, S., Ferner, M., Jeffke, T., Henne, A., Gottschalk, G. and Meyer, O. (2003) 'Complete nucleotide sequence of the circular megaplasmid pHCG3 of Oligotropha carboxidovorans: function in the chemolithoautotrophic utilization of CO, H2 and CO2.', *Gene*, 322, pp. 67–75.

Fukui, T., Ohsawa, K., Mifune, J., Orita, I. and Nakamura, S. (2011) 'Evaluation of promoters for gene expression in polyhydroxyalkanoate- producing Cupriavidus necator H16', *Applied Microbiology and Biotechnology*, 89(5), pp. 1527–1536. doi: 10.1007/s00253-011-3100-2.

Gai, C. S., Lu, J., Brigham, C. J., Bernardi, A. C. and Sinskey, A. J. (2014) 'Insights into bacterial CO2 metabolism revealed by the characterization of four carbonic anhydrases in Ralstonia eutropha H16.', *AMB Express*, 4(1), p. 2. doi: 10.1186/2191-0855-4-2.

Gay, P., Le Cqq, D., Steinmetz, M., Berkelman, T. and Kadol, A. C. I. (1985) 'Positive Selection Procedure for Entrapment of Insertion Sequence Elements in Gram-Negative Bacteria', *Journal of Bacteriology*, 164(2), pp. 918–921.

Ghysels, S., Mozumder, M. S. I., De Wever, H., Volcke, E. I. P. and Garcia-Gonzalez, L. (2018) 'Targeted poly(3-hydroxybutyrate-co-3-hydroxyvalerate) bioplastic production from carbon dioxide', *Bioresource Technology*, 249(August 2017), pp. 858–868. doi: 10.1016/j.biortech.2017.10.081.

Gibson, J. L. and Tabita, F. R. (1996) 'The molecular regulation of the reductive pentose phosphate pathway in Proteobacteria and Cyanobacteria.', *Archives of microbiology*, 166(3), pp. 141–50.

Gibson, J. L. and Tabita, F. R. (1997) 'Analysis of the cbbXYZ operon in Rhodobacter sphaeroides.', *Journal of bacteriology*, 179(3), pp. 663–9. doi: 10.1128/JB.179.3.663-669.1997.

Gonzalez-Esquer, C. R., Newnham, S. E. and Kerfeld, C. A. (2016) 'Bacterial microcompartments as metabolic modules for plant synthetic biology', *Plant Journal*, 87(1), pp. 66–75. doi: 10.1111/tpj.13166.

Gonzalez De La Cruz, J., Machens, F., Messerschmidt, K. and Bar-Even, A. (2019) 'Core Catalysis of the Reductive Glycine Pathway Demonstrated in Yeast', *Cite This: ACS Synth. Biol*, 8, pp. 911–917. doi: 10.1021/acssynbio.8b00464.

Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S. and Onufriev, A. (2005) 'H++: a server for estimating pK a s and adding missing hydrogens to macromolecules', *Nucleic Acids Research*, 33, pp. W368–W371. doi: 10.1093/nar/gki464.

Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S. and Walker, R. C. (2012) 'Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born', *Journal of*

Chemical Theory and Computation, 8(5), pp. 1542–1555. doi: 10.1021/ct200909j.

Le Grand, S., Götz, A. W. and Walker, R. C. (2013) 'SPFP: Speed without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations', *Computer Physics Communications*, 184(2), pp. 374–380. doi: 10.1016/j.cpc.2012.09.022.

Grousseau, E., Lu, J., Gorret, N., Guillouet, S. E. and Sinskey, A. J. (2014) 'Isopropanol production with engineered Cupriavidus necator as bioproduction platform', *Applied Microbiology and Biotechnology*, 98(9), pp. 4277–4290. doi: 10.1007/s00253-014-5591-0.

Grousseau, E., Blanchet, E., Déléris, S., Albuquerque, M. G. E., Paul, E. and Uribelarrea, J. L. (2014) 'Phosphorus limitation strategy to increase propionic acid flux towards 3-hydroxyvaleric acid monomers in Cupriavidus necator', *Bioresource Technology*, 153, pp. 206–215. doi: 10.1016/j.biortech.2013.11.072.

Gruber, S., Hagen, J., Schwab, H. and Koefinger, P. (2014) 'Versatile and stable vectors for efficient gene expression in Ralstonia eutropha H16', *Journal of Biotechnology*, 192(3), pp. 410–418. doi: 10.1016/j.jbiotec.2014.09.023.

Gruber, S., Schwab, H. and Heidinger, P. (2017) 'CbbR and RegA regulate cbb operon transcription in Ralstonia eutropha H16', *Journal of Biotechnology*, 257, pp. 78–86. doi: 10.1016/j.jbiotec.2017.07.005.

Grzeszik, C., Jeffke, T., Schäferjohann, J., Kusian, B. and Bowien, B. (2000) 'Phosphoenolpyruvate is a signal metabolite in transcriptional control of the cbb CO2 fixation operons in Ralstonia eutropha.', *Journal of molecular microbiology and biotechnology*, 2(3), pp. 311–20.

Hanko, E. K. R., Minton, N. P. and Malys, N. (2017)

'Characterisation of a 3-hydroxypropionic acid-inducible system from Pseudomonas putida for orthogonal gene expression control in Escherichia coli and Cupriavidus necator', *Scientific Reports*, 7(1), p. 1724. doi: 10.1038/s41598-017-01850-w.

Hansen, S., Vollan, V. B., Hough, E. and Andersen, K. (1999) 'The crystal structure of rubisco from Alcaligenes eutrophus reveals a novel central eight-stranded ß-barrel formed by ß-strands from four subunits', *Journal of Molecular Biology*, 288(4), pp. 609–621. doi: 10.1006/jmbi.1999.2701.

Hayer-Hartl, M. (2017) 'From chaperonins to Rubisco assembly and metabolic repair', *Protein Science*, 26(12), pp. 2324–2333. doi: 10.1002/pro.3309.

Heinhorst, S., Cannon, G. C. and Shively, J. M. (2006) 'Carboxysomes and Carboxysome-like Inclusions', in. Springer, Berlin, Heidelberg, pp. 141–165. doi: 10.1007/7171_023.

Hennemann, M. and Clark, T. (2014) 'EMPIRE: A highly parallel semiempirical molecular orbital program: 1: Self-consistent field calculations', *Journal of Molecular Modeling*, 20(7). doi: 10.1007/s00894-014-2331-4.

Herter, S., Fuchs, G., Bacher, A. and Eisenreich, W. (2002) 'A Bicyclic Autotrophic CO2 Fixation Pathway in Chloroflexus aurantiacus', *Journal of Biological Chemistry*, 277(23), pp. 20277–20283. doi: 10.1074/jbc.M201030200.

Holo, H. (1989) 'Chloroflexus aurantiacus secretes 3-hydroxypropionate, a possible intermediate in the assimilation of CO2 and acetate', *Archives of Microbiology*, 151(3), pp. 252–256. doi: 10.1007/BF00413138.

Huber, H., Gallenberger, M., Jahn, U., Eylert, E., Berg, I. A., Kockelkorn, D., Eisenreich, W. and Fuchs, G. (2008) 'A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation

cycle in the hyperthermophilic Archaeum Ignicoccus hospitalis.', *Proceedings of the National Academy of Sciences of the United States of America*, 105(22), pp. 7851–6. doi: 10.1073/pnas.0801043105.

Huber, H., Gallenbergger, M., Jahn, U., Eylert, E., Berg, I. A., Kockelkorn, D., Eisenreich, W. and Fuchs, G. (2008) 'A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum Ignicoccus hospitalis', *PNAS*, 105(22), pp. 7851–7856.

Hügler, M., Wirsen, C. O., Fuchs, G., Taylor, C. D. and Sievert, S. M. (2005) 'Evidence for autotrophic CO2 fixation via the reductive tricarboxylic acid cycle by members of the epsilon subdivision of proteobacteria.', *Journal of bacteriology*, 187(9), pp. 3020–7. doi: 10.1128/JB.187.9.3020-3027.2005.

Humphrey, W., Dalke, A. and Schulten, K. (1996) 'VMD: Visual molecular dynamics', *Journal of Molecular Graphics*, 14(1), pp. 33–38. doi: 10.1016/0263-7855(96)00018-5.

Huschner, F., Grousseau, E., Brigham, C. J., Plassmeier, J., Popovic, M., Rha, C. and Sinskey, A. J. (2015) 'Development of a feeding strategy for high cell and PHA density fed-batch fermentation of Ralstonia eutropha H16 from organic acids and their salts', *Process Biochemistry*, 50(2), pp. 165–172. doi: 10.1016/j.procbio.2014.12.004.

Iancu, C. V., Morris, D. M., Dou, Z., Heinhorst, S., Cannon, G. C. and Jensen, G. J. (2010) 'Organization, Structure, and Assembly of α-Carboxysomes Determined by Electron Cryotomography of Intact Cells', *Journal of Molecular Biology*, 396(1), pp. 105–117. doi: 10.1016/j.jmb.2009.11.019.

Ivanovsky, R. N., Fal, Y. I., Berg, I. A., Ugolkova, N. V., Krasilnikova, E. N., Keppen, O. I., Zakharchuc, L. M. and Zyakun, A. M. (1999) 'Evidence for the presence of the reductive pentose phosphate cycle in a filamentous anoxygenic photosynthetic

bacterium, Oscillochloris trichoides strain DG-6', *Microbiology*, 145(7), pp. 1743–1748. doi: 10.1099/13500872-145-7-1743.

Jeffke, T., Gropp, N., Kaiser, C., Kusian, B., Bowien, B. and Grzeszik, C. (1999) 'Mutational Analysis of the cbb Operon (CO 2 Assimilation) Promoter of Ralstonia eutropha Mutational Analysis of the cbb Operon (CO 2 Assimilation) Promoter of Ralstonia eutropha', *Journal of Bacteriology*, 184(14), pp. 4374–4380.

Johnson, A. O., Gonzalez-Villanueva, M., Tee, K. L. and Wong, T. S. (2018) 'An engineered constitutive promoter set with broad activity range for Cupriavidus necator H16', *ACS Synthetic Biology*. doi: 10.1021/acssynbio.8b00136.

Jorda, J., Lopez, D., Wheatley, N. M. and Yeates, T. O. (2013) 'Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria', *Protein Science*, 22(2), pp. 179–195. doi: 10.1002/pro.2196.

Jordan, D. B. and Chollet, R. (1983) 'Inhibition of ribulose bisphosphate carboxylase by substrate ribulose 1,5-bisphosphate', *Journal of Biological Chemistry*, 258(22), pp. 13752–13758.

Jung, H. R., Jeon, J. M., Yi, D. H., Song, H. S., Yang, S. Y., Choi, T. R., Bhatia, S. K., Yoon, J. J., Kim, Y. G., Brigham, C. J. and Yang, Y. H. (2019) 'Poly(3-hydroxybutyrate-co-3-hydroxyvalerate-co-3-hydroxyhexanoate) terpolymer production from volatile fatty acids using engineered Ralstonia eutropha', *International Journal of Biological Macromolecules*, 138, pp. 370–378. doi: 10.1016/j.ijbiomac.2019.07.091.

Kaneko, Y., Danev, R., Nagayama, K. and Nakamoto, H. (2006) 'Intact carboxysomes in a cyanobacterial cell visualized by hilbert differential contrast transmission electron microscopy.', *Journal of bacteriology*, 188(2), pp. 805–8. doi: 10.1128/JB.188.2.805-808.2006.

Kaplan, A. and Reinhold, L. (1999) 'CO2 concentrating mechanisms

in photosynthetic microorganisms', *Annual Review of Plant Physiology and Plant Molecular Biology*, 50(1), pp. 539–570. doi: 10.1146/annurev.arplant.50.1.539.

Kerfeld, C. A., Sawaya, M. R., Pashkov, I., G., C., E., W., K., T. and T.O, Y. (2004) *The structure of Halothiobacillus neapolitanus RuBisCO*, *RSCB Protein Data Bank*. Available at: https://www.rcsb.org/structure/1SVD (Accessed: 3 July 2020).

Kerfeld, C. A., Sawaya, M. R., Tanaka, S., Nguyen, C. V., Phillips, M., Beeby, M. and Yeates, T. O. (2005) 'Protein Structures Forming the Shell of Primitive Bacterial Organelles', *Science*, 309(5736).

Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. and Sutter, M. (2018) 'Bacterial microcompartments', *Nature Review Microbiology*, (16), pp. 277–290. doi: https://doi.org/10.1038/nrmicro.2018.10.

Kerfeld, C. A. and Erbilgin, O. (2015) 'Bacterial microcompartments and the modular construction of microbial metabolism', *Trends in Microbiology*, 23(1), pp. 22–34. doi: 10.1016/j.tim.2014.10.003.

Kerfeld, C. A., Heinhorst, S. and Cannon, G. C. (2010) 'Bacterial Microcompartments', *Annual Review of Microbiology*, 64(1), pp. 391–408. doi: 10.1146/annurev.micro.112408.134211.

Kerfeld, C. A. and Melnicki, M. R. (2016) 'Assembly, function and evolution of cyanobacterial carboxysomes', *Current Opinion in Plant Biology*, 31, pp. 66–75. doi: 10.1016/j.pbi.2016.03.009.

Kinney, J. N., Salmeen, A., Cai, F. and Kerfeld, C. A. (2012) 'Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly.', *The Journal of biological chemistry*, 287(21), pp. 17729–36. doi: 10.1074/jbc.M112.355305.

Kinney, J. N., Axen, S. D. and Kerfeld, C. A. (2011) 'Comparative analysis of carboxysome shell proteins.', *Photosynthesis research*, 109(1–3), pp. 21–32. doi: 10.1007/s11120-011-9624-6.

Klatt, C. G., Bryant, D. A. and Ward, D. M. (2007) 'Comparative genomics provides evidence for the 3-hydroxypropionate autotrophic pathway in filamentous anoxygenic phototrophic bacteria and in hot spring microbial mats', *Environmental Microbiology*, 9(8), pp. 2067–2078. doi: 10.1111/j.1462-2920.2007.01323.x.

Klein, M. G., Zwart, P., Bagby, S. C., Cai, F., Chisholm, S. W., Heinhorst, S., Cannon, G. C. and Kerfeld, C. A. (2009) 'Identification and Structural Analysis of a Novel Carboxysome Shell Protein with Implications for Metabolite Transport', *Journal of Molecular Biology*, 392, pp. 319–333. doi: 10.1016/j.jmb.2009.03.056.

Kusian, B., Bednarski, R., Husemann, M. and Bowien, B. (1995) 'Characterization of the duplicate ribulose-1,5-bisphosphate carboxylase genes and cbb promoters of Alcaligenes eutrophus.', *Journal of bacteriology*, 177(15), pp. 4442–50.

Kusian, B. and Bowien, B. (1997) 'Organization and regulation of cbb CO2 assimilation genes in autotrophic bacteria', *FEMS Microbiology Reviews*, 21(2), pp. 135–155. doi: 10.1016/S0168-6445(97)00054-5.

Kusian, B., Sültemeyer, D. and Bowien, B. (2002) 'Carbonic Anhydrase Is Essential for Growth of Ralstonia eutropha at Ambient CO2 Concentrations', *Journal of Bacteriology*, 184(18), pp. 5018–5026. doi: 10.1128/JB.184.18.5018-5026.2002.

Lauterbach, L. and Lenz, O. (2019) 'How to make the reducing power of H2 available for in vivo biosyntheses and biotransformations', *Current Opinion in Chemical Biology*, 49, pp. 91–96. doi: 10.1016/j.cbpa.2018.11.020.

Leadbeater, L., Siebert, K., Schobert, P. and Bowien, B. (1982) 'Relationship between activities and protein levels of ribulosebisphosphate carboxylase and phosphoribulokinase in Alcaligenes eutrophus', *FEMS Microbiology Letters*, 14(4), pp. 263–

266. doi: 10.1111/j.1574-6968.1982.tb00010.x.

Lenz, O., Schwartz, E., Dernedde, J., Eitinger, M. and Friedrich, B. (1994) 'The Alcaligenes eutrophus H16 hoxX gene participates in hydrogenase regulation.', *Journal of bacteriology*, 176(14), pp. 4385–93. doi: 10.1128/JB.176.14.4385-4393.1994.

Lenz, O., Bernhard, M., Buhrke, T., Schwartz, E. and Friedrich, B. (2002) 'The hydrogen-sensing apparatus in Ralstonia eutropha.', *Journal of molecular microbiology and biotechnology*, 4(3), pp. 255–62.

Li, H., Opgenorth, P. H., Wernick, D. G., Rogers, S., Wu, T.-Y., Higashide, W., Malati, P., Huo, Y.-X., Cho, K. M. and Liao, J. C. (2012) 'Integrated Electromicrobial Conversion of CO2 to Higher Alcohols', *Science*, 335(6076), pp. 1596–1596. doi: 10.1126/science.1217643.

Lin, Y., Pan, D., Li, J., Zhang, L. and Shao, X. (2017) 'Application of Berendsen barostat in dissipative particle dynamics for nonequilibrium dynamic simulation', *Journal of Chemical Physics*, 146(12). doi: 10.1063/1.4978807.

Lindsey, R. (2020) *Climate Change: Atmospheric Carbon Dioxide*, *NOAA Climate.gov*. Available at: https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide (Accessed: 14 March 2021).

Little, G. T., Ehsaan, M., Arenas-López, C., Jawed, K., Winzer, K., Kovacs, K. and Minton, N. P. (2019) 'Complete Genome Sequence of Cupriavidus necator H16 (DSM 428)', *American Society for microbiology*, 8(37), pp. 814–819. doi: 10.1128/MRA.00814-19.

Liu, Y., He, X., Lim, W., Mueller, J., Lawrie, J., Kramer, L., Guo, J. and Niu, W. (2018) 'Deciphering molecular details in the assembly of alpha-type carboxysome', *Scientific Reports*, 8(1), pp. 1–10. doi: 10.1038/s41598-018-33074-x.

Lo, C.-W., Wu, H.-S. and Wei, Y.-H. (2009) 'Optimizing acidic methanolysis of poly(3-hydroxyalkanoates) in gas chromatography analysis', *Journal of Chemical Engineering*, (4), pp. 487–494. doi: 10.1002/apj.

Long, B. M., Badger, M. R., Whitney, S. M. and Price, G. D. (2007) 'Analysis of Carboxysomes from Synechococcus PCC7942 Reveals Multiple Rubisco Complexes with Carboxysomal Proteins CcmM and CcaA *', *The Journal of Biochemical Chemistry*, 282(40), pp. 29323–29335. doi: 10.1074/jbc.M703896200.

Long, B. M., Tucker, L., Badger, M. R. and Dean Price, G. (2010) 'Functional cyanobacterial ß-carboxysomes have an absolute requirement for both long and short forms of the CcmM protein', *Plant Physiology*, 153(1), pp. 285–293. doi: 10.1104/pp.110.154948.

Long, B. M., Rae, B. D., Badger, M. R. and Dean Price, G. (2011) 'Over-expression of the ß-carboxysomal CcmM protein in Synechococcus PCC7942 reveals a tight co-regulation of carboxysomal carbonic anhydrase (CcaA) and M58 content', *Photosynthesis Research*, 109(1–3), pp. 33–45. doi: 10.1007/s11120-011-9659-8.

Maddocks, S. E., F Oyston, P. C. and Sarah Maddocks, C. E. (2008) 'Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins', *Microbiology*, 154, pp. 3609–3623. doi: 10.1099/mic.0.2008/022772-0.

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. and Simmerling, C. (2015) 'ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB', *Journal of Chemical Theory and Computation*, 11(8), pp. 3696–3713. doi: 10.1021/acs.jctc.5b00255.

Mani, N. and Dupuy, B. (2001) 'Regulation of toxin synthesis in Clostridium difficile by an alternative RNA polymerase sigma

factor', *Proceedings of the National Academy of Sciences of the United States of America*, 98(10), pp. 5844–5849. doi: 10.1073/pnas.101126598.

Masson-Delmotte, V. *et al.* (2018) 'IPCC, 2018: Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global'. doi: 10.1016/j.oneear.2019.10.025.

Meijer, W. G. (1994) 'The Calvin cycle enzyme phosphoglycerate kinase of Xanthobacter flavus required for autotrophic CO2 fixation is not encoded by the cbb operon.', *Journal of bacteriology*, 176(19), pp. 6120–6.

Menon, B. B., Dou, Z., Heinhorst, S., Shively, J. M. and Cannon, G. C. (2008) 'Halothiobacillus neapolitanus carboxysomes sequester heterologous and chimeric RubisCO species', *PLoS ONE*, 3(10). doi: 10.1371/journal.pone.0003570.

Mihiri Shashikala, H. B., Chakravorty, A. and Alexov, E. (2019) 'Modeling electrostatic force in protein-protein recognition', *Frontiers in Molecular Biosciences*, 6. doi: 10.3389/fmolb.2019.00094.

Minton, N. P., Zhang, Y., Schwarz, K. and Heap, J. (2013) 'Bacterial expression system'. International Patent No. WO 2013/144647.

Miyamoto, S. and Kollman, P. A. (1992) 'Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models', *Journal of Computational Chemistry*, 13(8), pp. 952–962. doi: 10.1002/jcc.540130805.

Mora, C. 1, Driese, S. G. and Colarusso, L. A. (1996) 'Middle to Late Paleozoic Atmospheric CO2 Levels from Soil Carbonate and Organic Matter', *Science*, 271(5252), pp. 1105–1107.

Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F. U., Bracher, A. and Hayer-Hartl, M. (2011) 'Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase', *Nature*, 479, pp. 194–

199. doi: 10.1038/nature10568.

Mueller-Cajar, O. (2017) 'The diverse AAA+ machines that repair inhibited Rubisco active sites', *Frontiers in Molecular Biosciences*, 4(MAY), pp. 1–17. doi: 10.3389/fmolb.2017.00031.

Myers, J., Grothaus, G., Narayanan, S. and Onufriev, A. (2006) 'A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules', *Proteins: Structure, Function and Genetics*, 63(4), pp. 928–938. doi: 10.1002/prot.20922.

Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S. and Savage, D. F. (2020) 'Multivalent interactions between CsoS2 and Rubisco mediate α- carboxysome formation', *Nature Structural & Molecular Biology*, 27(3), pp. 281–287. doi: 10.1038/s41594-020-0387-7.

Orita, I., Iwazawa, R., Nakamura, S. and Fukui, T. (2012) 'Identification of mutation points in Cupriavidus necator NCIMB 11599 and genetic reconstitution of glucose-utilization ability in wild strain H16 for polyhydroxyalkanoate production', *Journal of Bioscience and Bioengineering*, 113(1), pp. 63–69. doi: 10.1016/j.jbiosc.2011.09.014.

Otsuki, T. (2001) 'A study for the biological CO2 fixation and utilization system', *Science of The Total Environment*, 277(1–3), pp. 21–25. doi: 10.1016/S0048-9697(01)00831-2.

Parry, M. A. J., Keys, A. J., Madgwick, P. J., Carmo-Silva, A. E. and Andralojc, P. J. (2008) 'Rubisco regulation: a role for inhibitors', *Journal of Experimental Botany*, 59(7), pp. 1569–1580. doi: 10.1093/jxb/ern084.

Pearce, F. G. (2006) 'Catalytic by-product formation and ligand binding by ribulose bisphosphate carboxylases from different phylogenies', *Biochem. J*, 399, pp. 525–534. doi: 10.1042/BJ20060430.

Peña, K. L., Castel, S. E., de Araujo, C., Espie, G. S. and Kimber, M. S. (2010) 'Structural basis of the oxidative activation of the carboxysomal gamma-carbonic anhydrase, CcmM.', *Proceedings of the National Academy of Sciences of the United States of America*, 107(6), pp. 2455–60. doi: 10.1073/pnas.0910866107.

Pohlmann, A., Fricke, W. F., Reinecke, F., Kusian, B., Liesegang, H., Cramm, R., Eitinger, T., Ewering, C., Pötter, M., Schwartz, E., Strittmatter, A., Voss, I., Gottschalk, G., Steinbüchel, A., Friedrich, B. and Bowien, B. (2006) 'Genome sequence of the bioplastic-producing "Knallgas" bacterium Ralstonia eutropha H16.', *Nature biotechnology*, 24(10), pp. 1257–62. doi: 10.1038/nbt1244.

Portis, A. R. (2003) 'Rubisco activase – Rubisco's catalytic chaperone', *Photosynthesis Research*, 75, pp. 11–27.

Portis, A. R. and Salvucci, M. E. (2002) 'The discovery of Rubisco activase - Yet another story of serendipity', *Photosynthesis Research*. Photosynth Res, pp. 257–264. doi: 10.1023/A:1020423802875.

Price, G. D., Sültemeyer, D., Klughammer, B., Ludwig, M. and Badger, M. R. (1998) 'The functioning of the CO2 concentrating mechanism in several cyanobacterial strains: a review of general physiological characteristics, genes, proteins, and recent advances', *Canadian Journal of Botany*, 76(6), pp. 973–1002. doi: 10.1139/b98-081.

Price, G. D., Badger, M. R., Woodger, F. J. and 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), pp. 1441–1461. doi: 10.1093/jxb/erm112.

Qiancheng Ma (1998) 'Greenhouse Gases: Refining the Role of Carbon Dioxide', *NASA Science Briefs*.

Rae, B. D., Long, B. M., Badger, M. R. and Price, G. D. (2012) 'Structural Determinants of the Outer Shell of B-Carboxysomes in Synechococcus elongatus PCC 7942: Roles for CcmK2, K3-K4, CcmO, and CcmL', *PLoS ONE*, 7(8), p. e43871 1-10. doi: 10.1371/journal.pone.0043871.

Rae, B. D., Long, B. M., Badger, M. R. and 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 : MMBR*, 77(3), pp. 357–79. doi: 10.1128/MMBR.00061-12.

Ragsdale, S. W., Clark, J. E., Ljungdahl, L. G., Lundie, L. L. and Drake, H. L. (1983) 'Properties of purified carbon monoxide dehydrogenase from Clostridium thermoaceticum, a nickel, iron-sulfur protein.', *The Journal of biological chemistry*, 258(4), pp. 2364–9.

Ragsdale, S. W. (2008) 'Enzymology of the Wood-Ljungdahl Pathway of Acetogenesis', *Annals of the New York Academy of Sciences*, 1125(1), pp. 129–136. doi: 10.1196/annals.1419.015.

Ragsdale, S. W. and Pierce, E. (2008) 'Acetogenesis and the Wood–Ljungdahl pathway of CO2 fixation', *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1784(12), pp. 1873–1898. doi: 10.1016/j.bbapap.2008.08.012.

Reinhold, L., Kosloff, R. and Kaplan, A. (1991) 'A model for inorganic carbon fluxes and photosynthesis in cyanobacterial carboxysomes', *Canadian Journal of Botany*, 69, pp. 984–988.

Richardson, K., Steffen, W. and Liverman, D. (2011) *Climate change: Global risks, challenges and decisions*, *Climate Change: Global Risks, Challenges and Decisions*. doi: 10.1017/CBO9780511973444.

Rittmann, B. E. (2008) 'Opportunities for renewable bioenergy using

microorganisms', *Biotechnology and Bioengineering*, 100(2), pp. 203–212. doi: 10.1002/bit.21875.

Roberts, G. W., Fortier, M. O. P., Sturm, B. S. M. and Stagg-Williams, S. M. (2013) 'Promising pathway for algal biofuels through wastewater cultivation and hydrothermal conversion', *Energy and Fuels*, 27(2), pp. 857–867. doi: 10.1021/ef3020603.

Saini, R., Kapoor, R., Kumar, R., Siddiqi, T. and Kumar, A. (2011) 'CO2 utilizing microbes — A comprehensive review', *Biotechnology Advances journal*, 29, pp. 949–960. doi: 10.1016/j.biotechadv.2011.08.009.

Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. and Walker, R. C. (2013) 'Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald', *Journal of Chemical Theory and Computation*, 9(9), pp. 3878–3888. doi: 10.1021/ct400314y.

Sánchez-Andrea, I., Guedes, I. A., Hornung, B., Boeren, S., Lawson, C. E., Sousa, D. Z., Bar-Even, A., Claassens, N. J. and Stams, A. J. M. (2020) 'The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans', *Nature Communications*. doi: 10.1038/s41467-020-18906-7.

Santelli, E., Bankston, L. A., Leppla, S. H. and Liddington, R. C. (2004) 'Crystal structure of a complex between anthrax toxin and its host cell receptor', *Nature*, 430(7002), pp. 905–908. doi: 10.1038/nature02763.

Savir, Y., Noor, E., Milo, R. and Tlusty, T. (2010) 'Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape.', *PNAS*, 107(8), pp. 3475–80. doi: 10.1073/pnas.0911663107.

Sawaya, M. R., Cannon, G. C., Heinhorst, S., Tanaka, S., Williams, E. B., Yeates, T. O. and Kerfeld, C. A. (2006) 'The Structure of

Carbonic Anhydrase from the Carboxysomal Shell Reveals a Distinct Subclass with One Active Site for the Price of Two', *The Journal of Biological Chemistry*, 281, pp. 7546–7555. doi: 10.1074/jbc.M510464200.

Schäferjohann, J., Bednarski, R. and Bowien, B. (1996) 'Regulation of CO2 assimilation in Ralstonia eutropha : premature transcription termination within the cbb operon', *Journal of bacteriology*, 178(23), pp. 6714–6719.

Schmit, J. D., Kariyawasam, N. L., Needham, V. and Smith, P. E. (2018) 'SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation', *Journal of Chemical Theory and Computation*, 14(4), pp. 1823–1827. doi: 10.1021/acs.jctc.7b01254.

Schrödinger, LLC (2015) *The {PyMOL} Molecular Graphics System, Version~1.8*.

Schwartz, E., Henne, A., Cramm, R., Eitinger, T., rbel Friedrich, B. and Gottschalk, G. (2003) 'Complete Nucleotide Sequence of pHG1: A Ralstonia eutropha H16 Megaplasmid Encoding Key Enzymes of H 2 -based Lithoautotrophy and Anaerobiosis', *Journal of Molecular Biology*, 332, pp. 369–383. doi: 10.1016/S0022-2836(03)00894-5.

Schwartz, E., Voigt, B., Zuhlke, D., Pohlmann, A., Lenz, O., Albrecht, D., Schwarze, A., Kohlmann, Y., Krause, C., Hecker, M. and Friedrich, B. (2009) 'A proteomic view of the facultatively chemolithoautotrophic lifestyle of Ralstonia eutropha H16', *Proteomics*, 9(22), pp. 5132–5142. doi: 10.1002/pmic.200900333.

Shimizu, R., Chou, K., Orita, I., Suzuki, Y., Nakamura, S. and Fukui, T. (2013) 'Detection of phase-dependent transcriptomic changes and Rubisco-mediated CO2 fixation into poly (3-hydroxybutyrate) under heterotrophic condition in Ralstonia eutropha H16 based on RNA-seq and gene deletion analyses', *BMC Microbiology*, 13, p. 1. doi: 10.1186/1471-2180-13-169.

Shimizu, R., Dempo, Y., Nakayama, Y., Nakamura, S., Bamba, T., Fukusaki, E. and Fukui, T. (2015) 'New Insight into the Role of the Calvin Cycle: Reutilization of CO2 Emitted through Sugar Degradation.', *Scientific reports*, 5, p. 11617. doi: 10.1038/srep11617.

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

Shively, J M, Bradburne, C. E., Aldrich, H. C., Bobik, T. A., Mehlman, J. L., Jin, S. and Baker, S. H. (1998) 'Sequence homologs of the carboxysomal polypeptide CsoS1 of the thiobacilli are present in cyanobacteria and enteric bacteria that form carboxysomes - polyhedral bodies', *Canadian Journal of Botany*, 76(6), pp. 906–916. doi: 10.1139/b98-088.

Shively, J. M., Decker, G. L. and Greenawalt, J. W. (1970) 'Comparative Ultrastructure of the Thiobacilli', *Journal of Bacteriology*, 101(2), pp. 618–627.

Shively, Jessup M., van Keulen, G. and Meijer, W. G. (1998) 'Something from almost nothing: Carbon Dioxide Fixation in Chemoautotrophs', *Annual Review of Microbiology*, 52(1), pp. 191–230. doi: 10.1146/annurev.micro.52.1.191.

Singh, W., Karabencheva-Christova, T. G., Sparagano, O., Black, G. W., Petrov, P. Y. and Christov, C. Z. (2016) 'Dimerization and ligand binding in tyrosylprotein sulfotransferase-2 are influenced by molecular motions', *RSC Advances*, 6(22), pp. 18542–18548. doi: 10.1039/c6ra01899h.

Slade, H. D., Wood, H. G., Nier, A. 0, Hemingway, A. and Werkman, C. H. (1942) 'Assimilation of heavy carbon dioxide by heterotrophic bacteria', *Biochemical Journal*, 143, pp. 133–145.

So, A. K.-C., Cot, S. S.-W. and Espie, G. S. (2002) 'Characterization of the C-terminal extension of carboxysomal carbonic anhydrase

from Synechocystis sp. PCC6803', *Functional Plant Biology*, 29(3), p. 183. doi: 10.1071/PP01179.

So, A. K. C., Espie, G. S., Williams, E. B., Shively, J. M., Heinhorst, S. and Cannon, G. C. (2004) 'A novel evolutionary lineage of carbonic anhydrase ($ε$ class) is a component of the carboxysome shell', *J Bacteriol*, 186. doi: 10.1128/JB.186.3.623-630.2004.

South, P. F., Cavanagh, A. P., Liu, H. W. and Ort, D. R. (2019) 'Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field', *Science*, 363(6422), p. eaat9077. doi: 10.1126/SCIENCE.AAT9077.

Srinivasan, S., Barnard, G. C. and Gerngross, T. U. (2003) 'Production of recombinant proteins using multiple-copy gene integration in high-cell-density fermentations of Ralstonia eutropha', *Biotechnology and Bioengineering*, 84(1), pp. 114–120. doi: 10.1002/bit.10756.

Sutter, M., Roberts, E. W., Gonzalez, R. C., Bates, C., Dawoud, S., Landry, K., Cannon, G. C., Heinhorst, S. and Kerfeld, C. A. (2015) 'Structural Characterization of a Newly Identified Component of α-Carboxysomes: The AAA+ Domain Protein CsoCbbQ', *Nature Scientific Reports*, 5(1), p. 16243. doi: 10.1038/srep16243.

Tabita, F. R. (1999) 'Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective', *Photosynthesis Research*, 60, pp. 1–28.

Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C. and Yeates, T. O. (2008) 'Atomic-Level Models of the Bacterial Carboxysome Shell', *Science*, 319(5866), pp. 1083–1086.

Tanaka, S., Sawaya, M. R., Phillips, M. and Yeates, T. O. (2009) 'Insights from multiple structures of the shell proteins from the beta-carboxysome.', *Protein science : a publication of the Protein Society*, 18(1), pp. 108–20. doi: 10.1002/pro.14.

Thauer, R. K. (2007) 'A Fifth Pathway of Carbon Fixation', *Science*, 318(5857).

Tivol, W. F., Briegel, A. and Jensen, G. J. (2008) 'An Improved Cryogen for Plunge Freezing', *Microscopy and Microanalysis*, 14(5), pp. 375–379. doi: 10.1017/S1431927608080781.

Torella, J. P., Gagliardi, C. J., Chen, J. S., Kwabena Bediako, D., Colón, B., Way, J. C. and Silver, P. A. (2015) 'Efficient solar-to-fuels production from a hybrid microbial–water-splitting catalyst system', *Pnas*, 112(10), pp. 2337–2342.

Tsai, M.-Y., Zheng, W., Balamurugan, D., Schafer, N. P., Kim, B. L., Cheung, M. S. and Wolynes, P. G. (2016) 'Electrostatics, structure prediction, and the energy landscapes for protein folding and binding', *Protein Science*, 25, pp. 255–269. doi: 10.1002/pro.2751.

Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., Kerfeld, C. A. and Yeates, T. O. (2007) 'Structural Analysis of CsoS1A and the Protein Shell of the Halothiobacillus neapolitanus Carboxysome', *PLOS Biology*, 5(6), pp. 1345–1354.

Tsai, Y. C. C., Lapina, M. C., Bhushan, S. and Mueller-Cajar, O. (2015) 'Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria', *Nature Communications*, 6(May), pp. 1–10. doi: 10.1038/ncomms9883.

Turmo, A., Gonzalez-Esquer, C. R. and Kerfeld, C. A. (2017) 'Carboxysomes: metabolic modules for CO2 fixation', *FEMS microbiology letters*, 364(18), pp. 1–7. doi: 10.1093/femsle/fnx176.

Vascon, F., Gasparotto, M., Giacomello, M., Cendron, L., Bergantino, E., Filippini, F. and Righetto, I. (2020) 'Protein electrostatics: From computational and structural analysis to discovery of functional fingerprints and biotechnological design', *Computational and Structural Biotechnology Journal*. Elsevier B.V., pp. 1774–1789. doi: 10.1016/j.csbj.2020.06.029.

Verkhusha, V. V. and Kiryl, P. D. (2011) 'Guide to Red Fluorescent Proteins and Biosensors for Flow Cytometry', *Methods Cell Biol.*, (102), pp. 431–461. doi: 10.1002/psc.443.

Wang, B., Li, Y., Wu, N. and Lan, C. Q. (2008) 'CO2 bio-mitigation using microalgae', *Applied Microbiology and Biotechnology*, 79(5), pp. 707–718. doi: 10.1007/s00253-008-1518-y.

Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N. D., Hee, W. Y., Long, B. M., Price, G. D., Hartl, F. U. and Hayer-Hartl, M. (2019) 'Rubisco condensate formation by CcmM in β-carboxysome biogenesis', *Nature*, 566(7742), pp. 131–135. doi: 10.1038/s41586-019-0880-5.

Wang, Y., Chen, R., Cai, J. Y., Liu, Z., Zheng, Y., Wang, H., Li, Q. and He, N. (2013) 'Biosynthesis and Thermal Properties of PHBV Produced from Levulinic Acid by Ralstonia eutropha', *PLoS ONE*, 8(4), pp. 4–11. doi: 10.1371/journal.pone.0060318.

Weber, A. P. M. and Bar-Even, A. (2019) 'Update on Improved Carbon Fixation Update: Improving the Efficiency of Photosynthetic Carbon Reactions', *Plant Physiology_*, 179, pp. 803–812. doi: 10.1104/pp.18.01521.

Wheatley, N. M., Sundberg, C. D., Gidaniyan, S. D., Cascio, D. and Yeates, T. O. (2014) 'Structure and identification of a Pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) assembly factor in the alpha-carboxysome', *Journal of Biological Chemistry*, 289(11), pp. 7973–7981. doi: 10.1074/jbc.M113.531236.

Wilding-Steele, T. (2017) *Improving the expression and secretion of GH48 family cellulases in C. acetobutylicum*. University of Nottingha,.

Windhövel, U. and Bowien, B. (1991) 'Identification of cfxR, an activator gene of autotrophic CO2 fixation in Alcaligenes

eutrophus.', *Molecular microbiology*, 5(11), pp. 2695–705.

Wood, H. G. and Werkman, H. (1940) 'The fixation of CO2 by cell suspensions of propionibacterium pentosaceumi', *Biochemical Journal*, (34), pp. 7–14.

Wunder, T., Le, S., Cheng, H., Lai, S.-K., Li, H.-Y. and Mueller-Cajar, O. (2018) 'The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger', *Nature Communications*, 9(5076). doi: 10.1038/s41467-018-07624-w.

Xiong, J.-P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S. L. and Arnaout, M. A. (2002) 'Crystal Structure of the Extracellular Segment of Integrin αVβ3 in Complex with an Arg-Gly-Asp Ligand', *Science*, 296(5565), pp. 151–155. doi: 10.1126/SCIENCE.1069040.

Yan, Q., Du, G. and Chen, J. (2002) *Biosynthesis of Polyhydroxyalkanoates by Ralstonia eutropha with Short-chain-length Organic Acids*, *The Chinese Journal of Process Engineering*. Available at: http://en.cnki.com.cn/Article_en/CJFDTotal-HGYJ200205013.htm.

Yan, Q., Du, G. and Chen, J. (2003) 'Biosynthesis of polyhydroxyalkanoates (PHAs) with continuous feeding of mixed organic acids as carbon sources by Ralstonia eutropha', *Process Biochemistry*, 39(3), pp. 387–391. doi: 10.1016/S0032-9592(03)00127-4.

Yishai, O., Goldbach, L., Tenenboim, H., Lindner, S. N. and Bar-Even, A. (2017) 'Engineered Assimilation of Exogenous and Endogenous Formate in Escherichia coli', *ACS Synthetic Biology*, 6, p. 15. doi: 10.1021/acssynbio.7b00086.

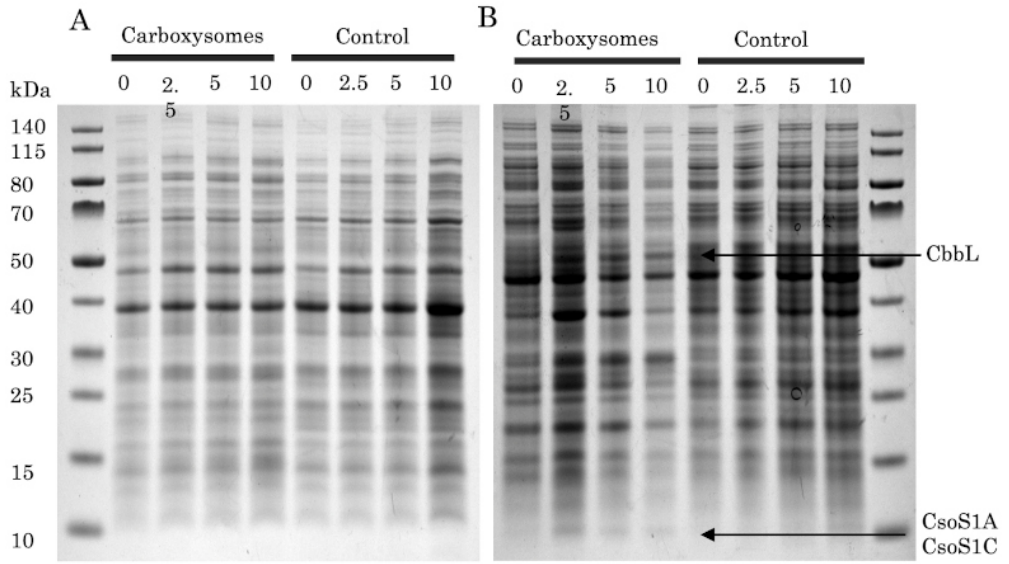
Zarzycki, J., Axen, S. D., Kinney, J. N., Kerfeld, C. A., T, O., EM, A., CA, K., TO, Y., F, K. and C, P. (2012) 'Cyanobacterial-based approaches to improving photosynthesis in plants', *Journal of*

Experimental Botany, 64(3), pp. 787–798. doi: 10.1093/jxb/ers294.

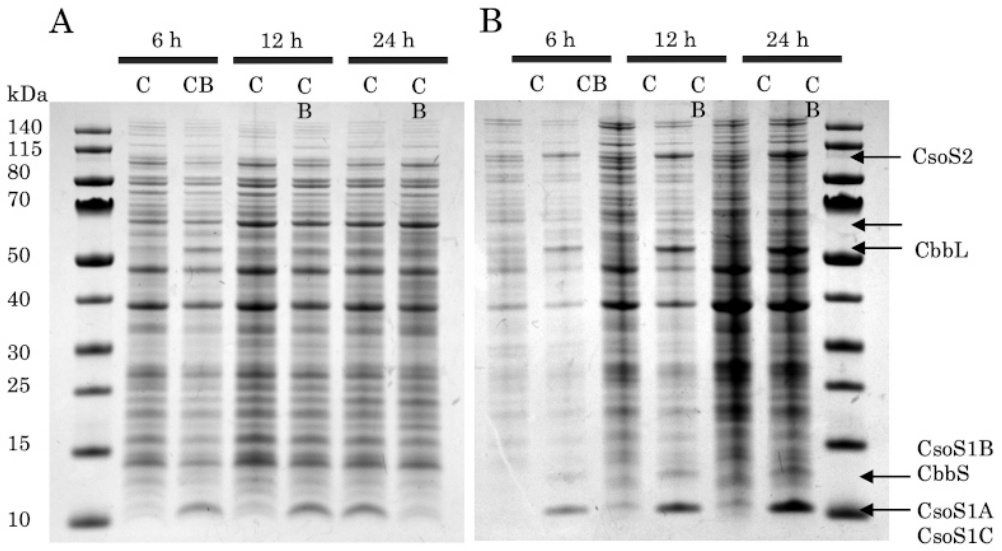
Zhang, Z., Witham, S. and Alexov, E. (2011) 'On the role of electrostatics on protein-protein interactions', *Physical Biology*, 8(3). doi: 10.1088/1478-3975/8/3/035001.

Zhou, H.-X. and Pang, X. (2018) 'Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation Graphical Abstract HHS Public Access', *Chem Rev*, 118(4), pp. 1691–1741. doi: 10.1021/acs.chemrev.7b00305.

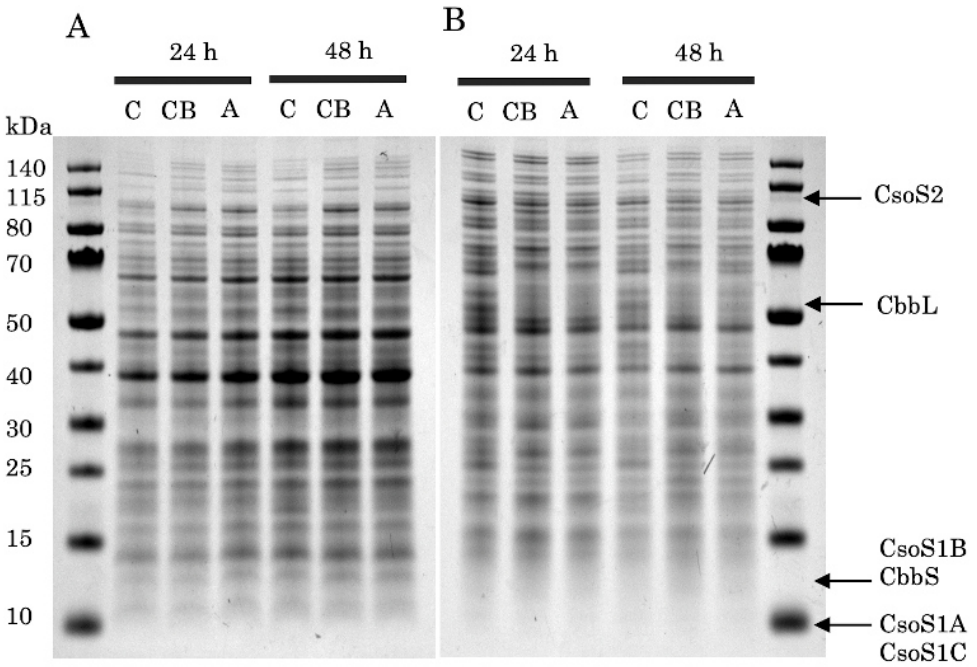
Appendix



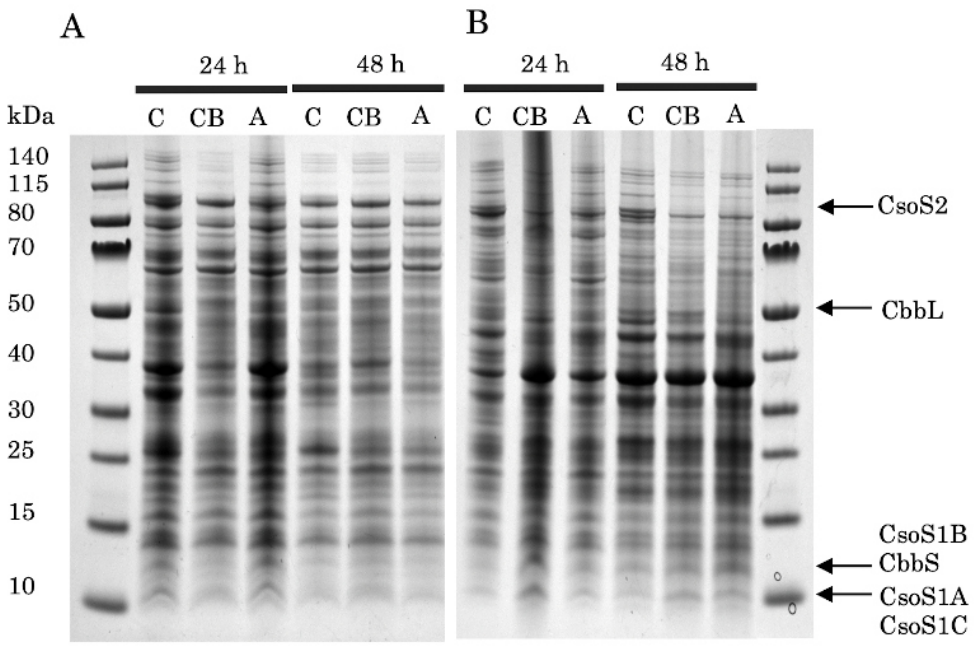
Appendix 1: Expression of *cso* operon under inducible promoter in *C. necator* Δ*phaCAB*. SDS-PAGE of the soluble (A) and membrane (B) protein fractions of the *C. necator* Δ*phaCAB* strain harbouring the plasmid carrying the *H. naepolitanous* α-carboxysomes genes under the control of the P_{BAD} L-arabinose-inducible promoter. The concentrations of L-arabinose used were 2.5, 5 and 10 mM, as indicated at the top of each lane. Cultures were grown in the presence of L-arabinose for 18 h. Carboxysomes = *C. necator* /pMTL71301::P_{BAD} *Hn*CB; Control = *C. necator* /pMTL71301::P_{BAD}.



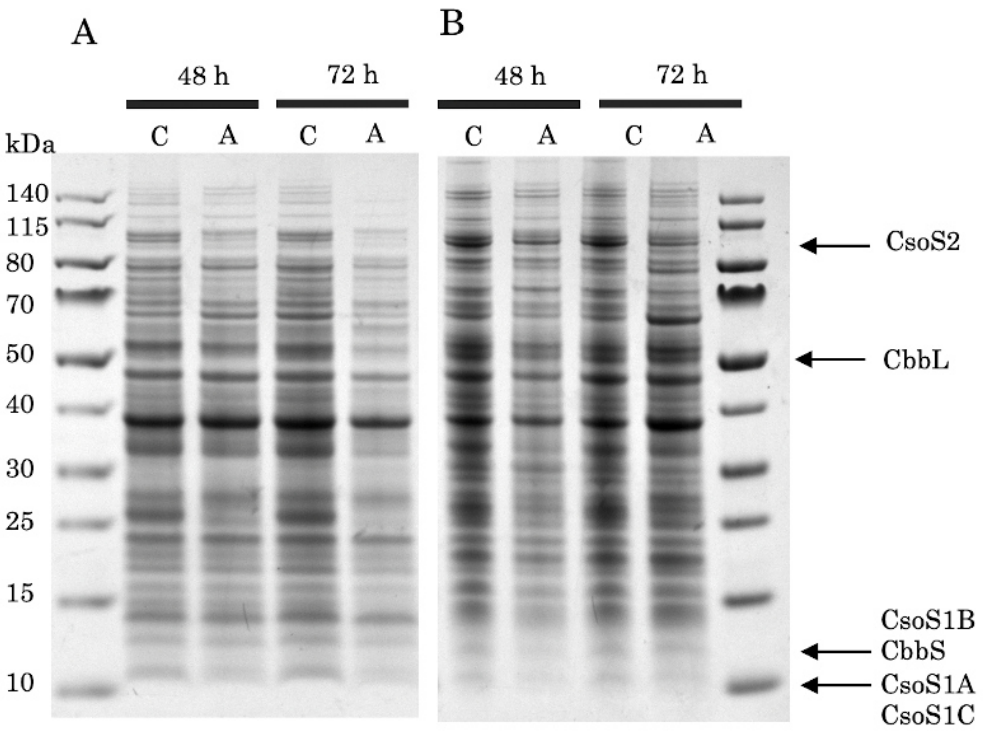
Appendix 2: Expression of *cso* operon under constitutive promoter in *C. necator* Δ*phaCAB*. SDS-PAGE of the soluble (A) and membrane (B) protein fractions of the *C. necator* Δ*phaCAB* strain harbouring the plasmid carrying the *H. naepolitanous* α-carboxysomes genes under the control of the constitutive promoter P_{phaC}. Protein expression was assessed after 6, 12 and 24 h of growth. CB = *C. necator* /pMTL71301::P_{phaC} *Hn*CB; C= control *C. necator* /pMTL71301:: P_{phaC} eYFP.



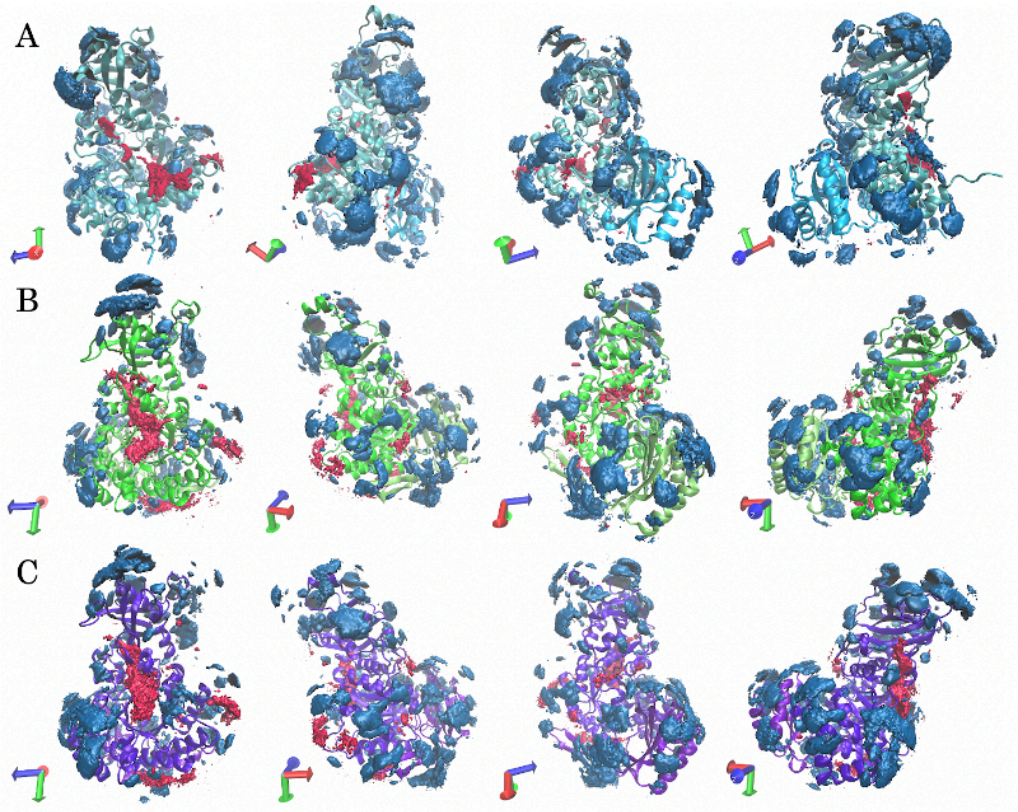
Appendix 3: Expression of *cso* operon with or without RubisCO activases in *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* in LB media. SDS-PAGE gel soluble (A) and membrane fraction (B) of C. necator Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* deletion strain harbouring the plasmid for constitutive heterologous expression of *H. naepolitanous* α-carboxysomes with and without RubisCO activases on stable plasmid grown in LB. Levels of expression after 24 and 48 h. CB = *C.necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* /pMTL71301::P_{phaC} *Hn*CB; A = *C.necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC*/pMTL71301:: P_{phaC} *Hn*CB Act; C= control *C.necator* Δ*phaCAB* Δ*panC*/pMTL71301:: P_{phaC}.



Appendix 4: Expression of *cso* operon with or without RubisCO activases in *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* grown in 0.4% SG MM. SDS-PAGE gel soluble (A) and membrane fraction (B) of *C.necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* deletion strain harbouring the plasmid for constitutive heterologous expression of *H. naepolitanous* α-carboxysomes with and without RubisCO activases on stable plasmid grown in SG MM. Levels of expression after 24 and 48 h. CB = *C.*necator Δ*phaCAB* ΔΔ*cbbLS* Δ*panC*/ pMTL71301::P_{phaC} *Hn*CB; A = *C.necator ΔphaCAB* ΔΔ*cbbLS* Δ*panC*/pMTL71301::P_{phaC} *Hn*CB Act; C= control *C.necator* Δ*phaCAB* Δ*panC*/pMTL71301:: $P_{phaC.}$



Appendix 5: Expression of *cso* operon with or without RubisCO activases in *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* grown in 80mM Na-formate MM. SDS-PAGE gel soluble (A) and membrane fraction (B) of *C. necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC* deletion strain harbouring the plasmid for constitutive heterologous expression of *H. naepolitanous* α-carboxysomes with and without RubisCO activases on stable plasmid grown in Na-Formate with 10% of CO2. Levels of expression after 24 and 48 h. A = *C.necator* Δ*phaCAB* ΔΔ*cbbLS* Δ*panC*/pMTL71301::P_{phaC} *Hn*CB Act; C= control *C.necator* Δ*phaCAB* Δ*panC*/pMTL71301::P_{phaC}.



Appendix 6: Ion Density Analysis of anions and cations during MD simulation of *Hn*, *Cn* and *CnMut*RubisCO at 1000 mM. Image representing ion density distribution of Na^+ (blue) and Cl^- (red) around (A) *Hn*, (B) *Cn* and (C) *CnMut*RubisCO structure following 550 ns molecular dynamics simulation of two independent replica for each condition at 150 mM. Isovalue: 25 e/Å^3.