The development of efficient *hemi-autotrophic* carbon fixation in *Escherichia Coli*

Ruud Hendriks, MSc, *ing.*

February 2023

This thesis is submitted to the University of Nottingham for the degree of Doctor of Philosophy

Conducted in the School of Life Sciences, at the Synthetic Biology Research Centre

Declaration of own work.

Unless otherwise acknowledged, this thesis and any experiments described in it were performed by me from February 2018 to March 2022 within the Synthetic Biology Research Center in the Biodiscovery Institute at the University of Nottingham in the United Kingdom. No part of this material has been submitted previously for a degree or any other qualification at any institute of learning.

Rudolf Martinus Antonius Hendriks, February 2023

To Nikki and Bink

Any work is a product of its time, place and people.

I'm thankful for the cumulative wealth of engineering techniques and vast knowledge available at the present, both greatly assisting and complicating the undertaking presented in this work. I also owe credit to the time I've spent at Hogeschool Zuyd and University of Gothenburg, studying the exciting subjects that have led me to this moment. The direct foundation of this work was published only two years prior to its start, and I hope this work in turn will find its place in a future filled with exciting possibilities. Despite the scientific leaps making it possible, the project was also limited by the inescapable challenges hopefully unique to its time.

This endeavour was made possible by the Synthetic Biology Research Center, maintaining a passionate group of highly specialized researchers and the facilities that support it. The University of Nottingham enabled the project by pursuing bold innovation and discovery in the School of Life Science, and I also owe thanks the United Kingdom's government for having the vision to commit long-term funding to research needed to help biotechnology reach its potential through the BBSRC.

Though I'm responsible for the work performed, I've hardly stood alone and owe thanks to all who have lent their invaluable support in one way or another. Professor Philippe Soucaille has been responsible for conceptualizing and directing the project with his extensive knowledge. Katalin Kovacs offered often-needed practical and moral support, and assembled a like-minded group where ideas and concepts were shared. Dr. Minyeong Yoo made the start to this project, shared her constructs and experience, and took care of a lot of organisation that otherwise would have fallen through the cracks. Furthermore, I also owe great thanks to colleagues Amaury Montarnal, Claudio Tomi Andrino, Marcel te Vrugt and Jonathan Humphreys for their support in- and outside of the lab. But all colleagues in the Synthetic Biology Research Center and many from surrounding lab spaces have contributed to this work in some degree. I also give thanks to my family; particularly my mother: Truus, sister: Inge, and dogs: Fang and Bink for

occasionally supporting this journey, and always giving me a home to come back to. Finally, and perhaps most importantly: thank you, the reader, for your interest in this work; I hope you find value in it.

Finally, I am eternally grateful for Nicola Dijon and her wonderful company; without her, the research presented here would never have made it to its current state.

Abstract

Carbon fixation is a process vital to any life and as by far its most prevalent variant, the Calvin Benson Bassham (CBB) cycle is vital to virtually all known terrestrial life. Mostly occurring in plants, it uses light energy to sequester atmospheric carbon dioxide (CO_2) and convert it into biomass. As the most inefficient natural carboxylation process and source of most biomass documented, even a small increase of its performance could have vast downstream effects. Such a development could assimilate the abundantly available atmospheric CO_2 while generating minimal amounts of waste for any biosynthesized product. The *Escherichia coli* bacterium was previously shown to functionally express the CBB cycle upon the addition of phosphoribulokinase (PRK) and ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO). Further knock-outs severed its energetic metabolism from the carbon metabolism resulted CO_2-dependent biomass accumulation. This carbon fixation is driven by the energy independently generated in the TCA cycle from a supply of pyruvate. This unique, split metabolism was dubbed *hemi-autotrophy*.

The hemi-autotrophic strain of *E. coli* serves as a model organism for the CBB cycle, but lacking any of the difficulties of light-dependent or multicellular organisms. A pyrophosphate-dependent 6-phosphofructokinase (PFP) originating from *Methylococcus capsulatus* Bath was characterised as catalyzing three reactions of the typical CBB cycle. Where PRK completes its catalysis with a dependency on energy-carrier adenosine triphosphate (ATP), PFP was shown to complete this reaction with the less energetic pyrophosphate (PPi) that is partially generated in its FBPase and SBPase-equivalent reactions. Successful integration of this synthetic CBB cycle would conserve 33% of all ATP expended in the native CBB cycle.

The hemi-autrophic E. coli strain's unique culturing requirements proved challenging but methods with increased dependability were established. Transformations without the relief of these conditions remain elusive, requiring pre-cultures in rich media and heterotrophic metabolism. The consecutive sub-culturing of the strain to increase its hampered growth characteristics resulted in mild improvements. Despite observing modest culturing

characteristic and a relatively high chromosomal mutation rate, the strain did not demonstrated an increase in transformation efficiency.

The attempted replacements of the plasmid-encoded *prkA* by *pfp* did not result in hemi-autotrophic growth in any of its constructs, despite modulation of their expression. Troubled by high mutation rates, it remains unknown whether the expression range of the significantly less efficient PFP was sufficient or if the cytoplasmic availability of PPi remained below its functionally required concentration. The putative H^+-pyrophosphatase pump (HPP), natively expressed as the second gene in the *pfp-hpp* operon, remains uncharacterised but its co-expression did not manage to compensate for this deficiency either. Though native *fbp* was successfully knocked-out, the essential inorganic pyrophosphatase gene of *E. coli* remains.

Thorough analysis of the components in the CBB system led to several design improvements and pathway modelling indicates the proposed synthetic CBB cycle is a viable alternative to its natural variant. Thermodynamic feasibility of the synthetic pathway was confirmed and kinetic analysis also predicted it to perform at reduced efficiencies while still indicating culture viability. Growth rates approximating those of the hemi-autotrophic strain were produced in a kinetic model of the central carbon metabolism while incorporating minimal assumptions. Modifying it to support the synthetic CBB cycle suggested its viability at a nominal reduction of growth, while suggesting further directions of research for the system.

Contents

Nomenclature

2PG	2-phosphoglycerate
3PG	3-phosphoglycerate
ADP	adenosine diphosphate
amp	ampicillin
ATP	adenosine triphosphate
BLA	beta-lactamase
bp	basepairs
BPG	1,3-bisphosphoglycerate
CA	carbonic anhydrase
CAM	crassulacean acid metabolism
cAMP	cyclic adenosine monophosphate
CBB	Calvin-Benson-Bassham
CCM	carbon concentrating mechanism
chl	chloramphenicol
CRP	cAMP receptor protein
DHAP	dihydroxyacetone phosphate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
F6P	fructose 6-phosphate
FBA	fructose 1,6-bisphosphate aldolase
FBP	fructose 1,6-bisphosphate
FBPase	fructose 1,6-bisphosphatase
G3P	glyceraldehyde 3-phosphate

GAPDH	glyceraldehyde 3-phosphate dehydrogenase
HPP	H^+-pyrphosphatase pump
indel	insertion/deletion
kan	kanamycin
KO	knock-out
LB	lysogeny broth (a rich culture medium)
NAD^+	nicotinamide adenine dinucleotide, oxidized
NADH	nicotinamide adenine dinuclotide, reduced
$NADP^+$	nicotinamide adenine dinucleotide phosphate, oxidized
NADPH	nicotinamide adenine dinucleotide phosphate, reduced
OD_x	optical density at x nm
ORF	open reading frame
PCR	polymerase chain reaction
PEP	phosphoenolpyruvate
PFK	phosphofructokinase
PFP	phosphofructokinase, pyrophosphate variant
PGK	phosphoglycerate kinase
Pi	orthosphosphate
PPA	inorganic pyrophosphatase
PPi	pyrophosphate
PRK	phosphoribulokinase
R5P	ribose 5-phosphate
RBS	ribosome binding site
RHP	reductive hexulose phosphate

RNA	ribonucleic acid
RPI	ribose 5-phosphate 3-isomerase
Ru5P	ribulose 5-phosphate
RuBisCO	ribulose 1,5-bisphosphatase carboxylase/oxygenase
RuBP	Ribulose 1,5-bisphosphate
RuBP	ribulose 1,5-bisphosphate
S7P	sedoheptulose 7-phosphate
SBP	sedoheptulose 1,7-bisphosphate
SBPase	sedoheptulose 1,7-bisphosphatase
spec	spectinomycin
TCA	tricarboxylic acid
TKT	transketolase
TPI	triose phosphate isomerase
UTR	untranslated region
Xu5P	xylulose 5-phosphate

Chapter 1

Background

"Life can only be understood backwards; but it must be lived forwards."

Søren Kierkegaard

1.1 Introduction

As the primary component of all known life, carbon assimilation is arguably the most vital biochemical process to any organism. Throughout the tree of life several different carbon fixation pathways have evolved, each adapted to the carbon and energy sources available in the organism's respective environments. As only autotrophic organisms can assimilate inorganic carbon, they are responsible for its propagation into the food-chain and are therefore of great relevance to a habitable biosphere [1, 2] and by extension the supply of food [3], chemical [4] and medical [5] resources that the support modern human civilization. Effective use, modification and recombination of these indispensable organisms and their supporting biochemical pathways would prove a reduction in the stresses modern civilization imposes on nature [6], and vice versa [7].

Biosynthesis of organic compounds is generally preferable to organic synthesis as it doesn't require high energy inputs and can be used to recycle waste products with high efficiency [5, 8]. Large variations in habitats, competition, space and time have created organisms able to produce and consume an immense variety of compounds [9]. The advanced techniques and knowledge employed in synthetic biology can be utilized to enhance, recombine and rebalance biological processes to suit the needs of the modern world [4, 10]. Applying these techniques to optimize the use and processing of resources by primary producers can further increase their efficiency and output, and has great potential for both commercial and environmental interests [11] while remedying impending environmental disasters such as the ongoing decline in bio-diversity [12] and atmospheric disruption [13].

The use of carbon in ecosystems is cyclical, with its potential energy changes depending on the utilizing host. Heterotrophic organisms mostly oxidise carbon as they extract the energy condensed in complex carbon compounds to ultimately leave the simplest, least energetic compound as a final product: carbon dioxide (CO_2). As only autotrophs introduce metabolizable carbon into the food-web the entire ecosystem derives its biomass and, with it, metabolic energy from them. The use of more energetic carbon-based compounds does occur but generally not autotrophically, as it is evolutionary simpler to oxidise them further to the lowest achievable free energy level of carbon. Non-carbonic sources of energy such as light, hydrogen oxidation and/or sulphur reduction [6] are commonly employed for the autotrophic fixation of carbon. As biological systems obey the laws of thermodynamics too [14], using this energy for carbon assimilation is always at a net loss. Increasing the carbon or energy conversion efficiency in any organism, even by a fraction, would have a significant impact on every downstream usage of that biomass.

1.2 Natural carbon fixation pathways

As shown in table 1.1, four carbon fixation mechanisms have been documented to occur in archaea, though few details are known about their specific metabolic features. *Dicarboxylate/4-hydroxybutyrate* (dc/hb) and *3-hydroxypropionate/4-hydroxybutyrate* (hp/hb) cycles share seven reactions catalysing succinyl-CoA to acetyl-CoA [15] and both are driven by energy gained from the reduction of sulphur to hydrogen sulphide. Where the dc/hb cycle operates anaerobically, the hp/hb cycle is aerobic but less energy-efficient. The *reductive acetyl-CoA* (or Wood-Ljungdahl) pathway is a non-cyclic carbon assimilation pathway found in methanogenic and acetogenic bacteria and archaea. As it is driven by the oxidation of hydrogen, it is catalytically the slowest carbon-fixation pathway but also the most energy-efficient [6]. The *reductive hexulose phosphate* (RHP) pathway was discovered more recently in methanogenic archaea. With its ribulose monophosphate pathway it converts fructose-6-phosphate (F6P) to ribulose-5-phosphate (Ru5P) instead of through the ubiquitous pentose phosphate pathway [16, 17].

Found in bacteria, the *3-hydroxypropionate* (Fuchs-Holo) bi-cycle consists of three common reactions after which the cycle splits in two interlinking

Pathway	Organisms	Energy source	Reactions	ATP/ CO2	NAD(P)H /CO2	Overall reaction
Dicarboxylate/ 4-hydroxybutyrate cycle	Archaea	Hydrogen/ sulphur	14	1.5	2	CO_2 + HCO_3^- +3 ATP + NAD(P)H + Fdred + 4MVred + CoASH → AcCoA + 2 ADP + AMP + 2 Pi + 2 PPi + $NAD(P)^+$ + Fdox + 4MVox
3-hydroxypropionate/ 4-hydroxybutyrate cycle	Archaea	Hydrogen/ sulphur	10	2	2	2 HCO_3^- + 4 ATP + 4 NAD(P)H + CoASH → AcCoA + 3 ADP + 3 Pi + AMP + PPi + 4 $NADP^+$
Wood-Ljungdahl pathway	Anaerobic bacteria	Hydrogen	8	0.5	2	2 CO_2 + ATP + 2 NAD(P)H + 2 Fdred + CoASH → AcCoA + ADP + Pi + 2 $NADP^+$ + 2 Fdox
Reduced hexulose pathway	Archaea	Hydrogen	9	2	1	CO_2 + 2 ATP + NAD(P)H → CH_2O + 2 ADP + NAD(P)+
3-hydroxyproprionate	Green non-sulphur bacteria	Light	16	1.67	1.67	3 HCO_3^- + 5 ATP + 5 NAD(P)H → Pyruvate + 3 ADP + 2 AMP + 3 Pi + 2 PPi + 5 $NAD(P)^+$
Reductive TCA cycle	Green sulphur bacteria	Light/ sulphur	9	1	2	2 CO_2 + 2 ATP + 2 NAD(P)H + FADH + Fdred + CoASH → AcCoA + 2 ADP + 2 Pi + 2 $NAD(P)^+$ + FAD^+ + Fdox
Reductive glycine pathway	Sulphur-reducing bacteria	Hydrogen/ sulphate	7	1	1.5	2 CH_2O + NH_3 + 2 ATP + 3 NAD(P)H + CO_2 = 2 ADP + $NAD(P)^+$ + Ser
CBB cycle	Plants, algae, cyanobacteria	Light	13	3	2	3 CO_2 + 9 ATP + 6 NAD(P)H → G3P + 9 ADP + 6 $NAD(P)^+$ + 8 Pi

Table 1.1: An overview of the known naturally occurring carbon fixation pathways and their characteristics. Molecular quantities are stoichiometrically balanced, based on full recovery of its smallest carbon-containing metabolite present prior to carbon incorporation. Energy requirements are based on expenditure of reductive power per fixed molecule of CO_2. Based on data obtained from [8] and expanded with additional developments [16, 18].

branches [19]. One that requires extra reductive power and produces a molecule of glyoxylate, where the other recovers it to produce a molecule of pyruvate, together allowing for the aerobic photo-heterotrophic metabolism found in green non-sulphur bacteria. Catalysed by 13 enzymes, it is the most complex known carbon assimilation pathway, though it captures two CO_2 molecules per complete cycle. The *reductive tricarboxylic acid* (rTCA) cycle is found in anaerobic bacteria where it generally uses a photosystem adapted for low light intensities to oxidize hydrogen or sulphur as a source of energy [19]. This cycle employs enzymes adapted to operate the ubiquitous tricarboxylic acid (TCA) cycle in both directions, making it the carbon-fixation pathway with the fewest composing enzymes. This adaptation of the TCA cycle allows for dynamic carbon fixation and energy generation on demand from the same ubiquitous metabolites. It anaerobically assimilates two molecules of CO_2 per cycle, of which isocitrate dehydrogenase is the most efficient known carboxylase. A second non-cyclic carbon fixation pathway was recently theorized, implemented [20], and then discovered in nature [18]. The *reductive glycine pathway* reduces CO_2 to formate, is converted to methylene while bound to tetrahydrofolate and gets converted to glycine via a double inter-reaction involving ammonia. The glycine can then be easily converted to acetyl-CoA for use via the central carbon metabolism.

The *reductive pentose phosphate* pathway, otherwise known as the Calvin-Benson-Bassham (CBB) cycle (fig. 1.1) is found in the chloroplast organelles of plants and algae, and the cytoplasm of cyanobacteria. Its carboxylating enzyme is Ribulose-1,5-bisphosphatase carboxylase/oxygenase (RuBisCO), and together with phosphoribulokinase (PRK) it circularizes the ubiquitous gluconeogenesis and pentose phosphate pathways to harvest gaseous CO_2 [21]. The CBB cycle has evolved the most recent of the presently documented carbon fixation pathways [22]. It is also thought of as the most successful carbon fixation pathway to have evolved on Earth as it is encountered in the vast majority of its autotrophic organisms [23]. RuBisCO is likely to be the most abundant enzyme on Earth [24], providing most heterotrophic life with the energy and biomass they acquire using it to support the entire ecosphere [23]. Despite its success as the only natural carboxylating pathway operational under chemically and thermally mild conditions, it is notoriously inefficient [25] and requires more energy than the other discovered carbon fixating pathways.

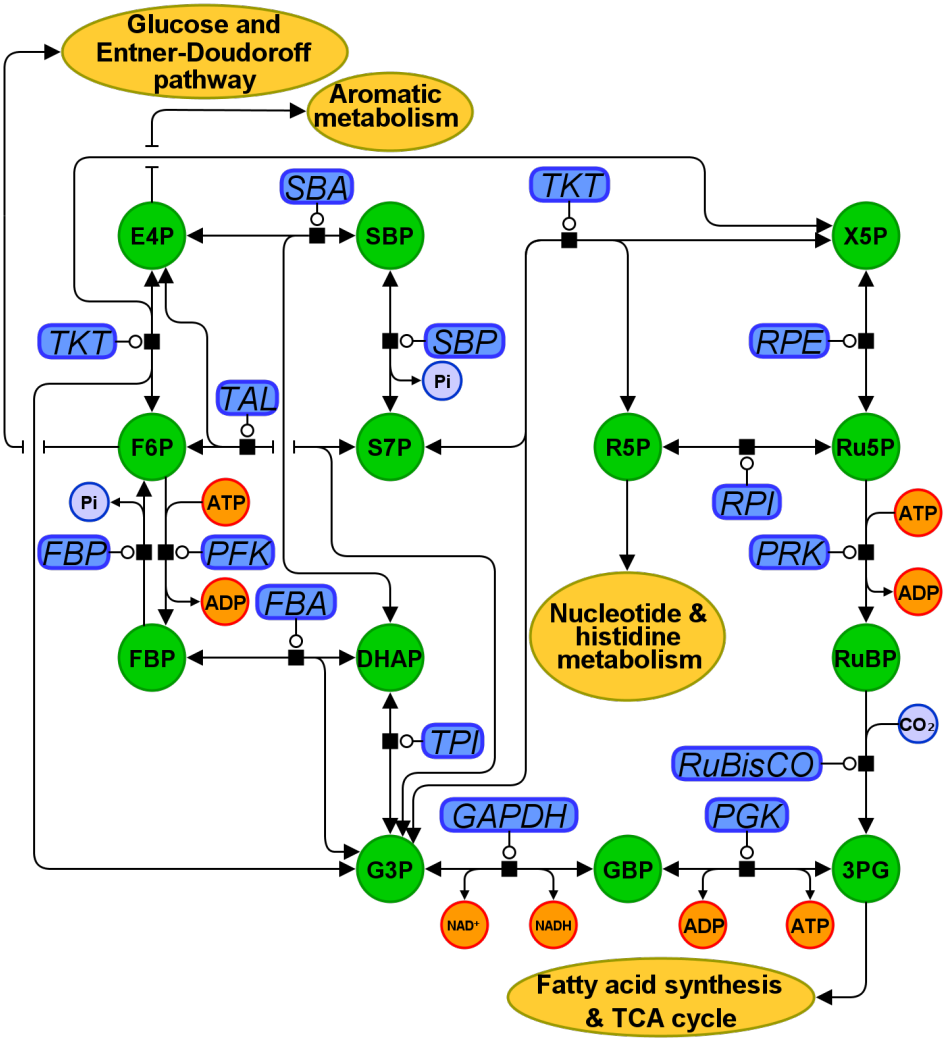

Figure 1.1: A Schematic overview of the canonical CBB cycle. Major connected pathways are displayed in yellow bubbles leading from metabolic branch-points. Enzymatic reactions are shown in blue bubbles while major metabolites are in green, with energetic cofactors in orange and inorganic molecules in light-blue.

1.3 CBB cycle

The CBB cycle allows for operation under a wide range of conditions, offering algae and plants the ability to exploit a large surface-area of the biosphere and sustain all heterotrophic life forms. At the (thylakoid) membrane photons are absorbed and their energy transferred to an antenna complex. The gained energy is transported to the photosystem where protons and oxygen are split from water. The electrons gained from it are transferred

by menaquinones until accepted by a cytocrome protein to pump protons across the membrane, creating a proton gradient. A second photosystem generates additional reductive power by the reduction of nicotinamide adenine dinucleotide phosphate (NADP) to NADPH or further increase the osmotic pressure by expending its energy at the cytochrome. Adenosine triphosphate (ATP) synthase exploits the created proton motive force to photo-phosphorylate adenosine diphosphate (ADP) into the organism's universal energy carrier. The energy the organism gains is used for metabolic maintenance or used for stored in carbon compounds of increasing complexity, like sugar is polymerized through gluconeogenesis. In the absence of usable external energy, the process is easily reversed by the process of glycolysis, to yield ATP in the TCA cycle.

Light-independent reactions incorporate environmental CO_2 into organic components either usable for energy storage or growth. The fixation of one molecule of CO_2 requires the reduction of 3 ATP and 2 nicotinamide adenine dinuclotide (NADH) [26] molecules. The CBB cycle is also referred to as the C3 photosynthetic metabolism, as the cycle is completed when ribulose 1,5-bisphosphate (RuBP) is directly carboxylated by RuBisCO into 3-carbon molecules: 3-phosphoglycerate (3PG). Several variations to this basic pathway have evolved to handle differences in the CO_2/O_2 ratio in the molecular neighbourhood of RuBisCO to increase energy, water, or carbon-efficiency and are collectively referred to as carbon concentrating mechanisms (CCMs).

1.3.1 CBB cycle enzymes

The CBB, in its forward direction, consists of the near-universally present gluconeogenesis and non-oxidative pentose phosphate pathways, in the central carbon metabolism. The former consumes energy and is used for the processing of metabolites into storable polymers for use in energy-scarce times and reproduction. The pathway's reversal, during glycolysis, yields energy and can generate even more when its products are directed towards the TCA cycle. The pentose phosphate pathways direct sugars towards the synthesis of nucleotides and amino acids, or vice versa. In the CBB cycle, these two pathways are connected by PRK and RuBisCO, requiring the gluconeogenesis to operate in its energy-consuming direction to capture CO_2. This interconnectivity makes the CBB cycle mostly reversible, with the excep-

tion of these two enzymes unique to it, metabolite flow is largely dependent on metabolite availability.

As reflected in table 1.2, starting from the smallest molecule in the CBB cycle: 3-phosphoglycerate (3PG) is converted to glycerate-1,3-bisphosphate (BPG) and glyceraldehyde-3-phosphate (G3P) by phosphoglycerate kinase (PGK) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). These reactions require the oxidation of energy carriers ATP and NADH, respectively. Next, G3P is converted by triose phosphate isomerase (TPI) into dihydroxyacetone phosphate (DHAP) and together one of it and its precedent are converted into fructose-1,6-bisphosphate (FBP) by the multifunctional (fructose bisphosphate) aldolase (or FBA). The gluconeogenetic part of the CBB cycle is completed as fructose-1,6-bisphosphate (FBPase) is converted into F6P. All these reactions are reversible except for FBPase which has an agonistic reaction completed by phosphofructokinase (PFK) and catalysed by the reduction of ATP.

Next, in the non-oxidative pentose phosphate part of the pathway, F6P and G3P are converted into xylulose-5-phosphate (Xu5P) and erythrose-4-phosphate (E4P) by the multi-functional transketolase (TKT). Ribulose phosphate-3-epimerase then catalyses the reaction of Xu5P to Ru5P. F6P can also be converted together with E4P into G3P and sedoheptulose-7-phosphate (S7P) by transaldolase. E4P is also united with DHAP to form sedoheptulose-1,7-bisphosphate (SBP) by fructose bisphosphate aldolase. This in turn is dephosphorylated to S7P by sedoheptulose-1,7-bisphosphatase (SBPase), exclusively found in photosynthesizing organisms, and catalysed by a dual-functioning FBPase in cyanobacteria. Transaldolase further catalyses the reaction from S7P and G3P into Xu5P and ribose-5-phosphate (R5P). The latter of which is also converted into Ru5P by ribose-5-phosphate isomerase (RPI). Metabolic flux in the CBB cycle is divided from F6P on, until TKT brings both branches together with the formation of Ru5P. These interacting reactions not only recycle the E4P and Xu5P from each other but the involved proteins, TKT and transaldolase, regulate the capacity of both junctions of the cycle as most of its metabolites are flux branch-points to other pathways.

PRK and RuBisCO complete the metabolic loop of the CBB cycle to regenerate its metabolites with the gain of carbon. The expense of ATP by

Enzyme	Abbreviation	Gene(s)	EC number	*E. coli* native	Reaction	stoichiometry
Phosphoglycerate kinase	PGK	*pgk*	2.7.2.3	Y	ATP + 3PG = ADP + BPG	10
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	*gapA*	1.2.1.12	Y	BPG + NADH + H+ = G3P + PPi + NAD+	10
Triose phosphate isomerase	TPI	*tpiA*	5.3.1.1	Y	G3P = DHAP	4
Fructose-1,6-bisphosphate aldolase	FBA	*fbaA*/ *fbaB*	4.1.2.13	Y	DHAP + G3P = FBP	3
					DHAP + E4P = SBP	1
Phospho-fructokinase	PFK	*pfkA*	2.7.1.11	Y	ATP + F6P → ADP + FBP	
Fructose-1,6,-bisphosphatase	FBP	*fbp*/ *glpX*/ *yggF*	3.1.3.11	Y	FBP + H2O → F6P + Pi	3
Sedoheptulose-1,7-bisphosphatase	SBP	*sbp*/ *fbp*	3.1.3.37	N	SBP + H2O → S7P + Pi	1
Transketolase	TKT	*tktA*/ *tktB*	2.2.1.1	Y	S7P + G3P = R5P + Xu5P	2
					F6P + G3P = E4P + Xu5P	2
Ribulose phosphate-3-epimerase	RPE	*rpe*	5.1.3.1	Y	Ru5P = Xu5P	4
Transaldolase	TAL	*talA*	2.2.1.2	Y	S7P + G3P = E4P + F6P	1
Ribose-5-phosphate isomerase	RPI	*rpiA*	5.3.1.6	Y	R5P = Ru5P	2
Phosphoribulo-kinase	PRK	*prkA*	2.7.1.19	N	Ru5P + ATP = RuBP + ADP	6
Ribulose-1,5-bisphosphate carboxylase/ oxygenase	RuBisCO	*cbbM*	4.1.1.39	N	RuBP + CO_2 → 2 3PG	6

Table 1.2: An overview of the enzymes involved in the canonical CBB cycle. Gene names are as used in *E. coli*, even if the organism doesn't harbour any functional variants of the gene naturally. Stoichiometry refers to the amount of conversions required to regenerate the original carbon input of the cycle, excluding that gained from CO_2. The PFK reaction does not participate in the forward direction of the pathway and is therefore not included in its stoichiometry.

PRK converts R5P to RuBP, which can be toxic in *E. coli* [27]. RuBisCO finally assimilates a CO_2 molecule and converts it together with one RuBP in two 3PG molecules. This final carboxylation step is catalytically the most inefficient of the cycle as it seems to have evolved from an ancient protein with a completely unrelated function [28], it is therefore highly expressed to prevent bottlenecks at the carboxylation stage of the pathway. The assimilation of the similar-structured oxygen molecules instead of carbon dioxide is named photo-respiration and yields one 3PG and one 2-phosphoglycolate (2PG) molecule. Converting the latter back to the initial 3PG is a slow process that takes place in three different cell compartments, requires additional reductive potential and produces toxic intermediates.

The stoichiometrical balance of the 13 reactions the CBB cycle is composed of gains a total of one 3PG, 8 phosphate, 3 H_2O, 9 ADP and 6 NAD^+ while consuming 3 CO_2, 9 ATP, 6 NADH and 6 hydrogen molecules. Its low specific activity and high energy demand per sequestered carbon atom explain the high abundance of RuBisCO in its hosts. Despite its energetic disadvantages, the CBB cycle has enabled its utilizing organisms to account for the largest share of carbon fixation of all encountered pathways because light has a near-limitless availability. Its presence at or near the surface allows for its operation in multi-cellular aerobic organisms under relatively mild chemical and thermal conditions without dependence on localized sources of carbon or energy.

In the 2.5$\sim$3 billion years since its emergence [29], RuBisCO is found in a substantial portion of the tree of life and several distinct isoforms have been identified throughout. As the primary method of carbon fixation in terrestrial life it has been the subject of great scientific interest though its abundance only allowed for the thorough study of only a fraction of these variants. Form I, found in cyanobacteria, algae and plants consists of eight large catalytic subunits (L) and eight of the regulatory small (S) subunits in a heterohexadecamer. In plants and some algae it can be inactivated by the binding of RuBisCO activase to its active site, requiring ATP (and therefore light) for its release. Form IA and IB are found in cyanobacteria and proteobacteria, B in algae and plantae, C in proteo-bacteria, and D in non-green algae. Form II is encountered in a species of protists as an L2 homo-dimer, form III in methanogenic archaea (in the RHP pathway) and

form IV is widely distributed in archaea [30]. Little else is known about the latter RuBisCO-like proteins, except that they facilitate functions unrelated to carboxylation.

1.3.2 Regulation and related pathways

As oxygen and carbon dioxide share two oxygen molecules and differ only a relatively small carbon atom, the former will occasionally get catalysed with RuBP too. Lacking a carbon atom, the products of this photo-respiration produces one 3PG and one 2PG molecule instead (and is therefore also referred to as the 2C cycle). In plants and algae, the photorespiratory pathway regenerates 3PG from 2PG for re-use in the CBB or TCA cycle. Here, 2PG is first metabolized to glycolate and transported to the peroxisome for conversion to glycine. Next, it is decarboxylated to serine in the mitochondria, releasing ammonia (NH_3) and CO_2 while reducing NADH. Back in the peroxisome, the toxic NH_3 is further metabolized to glutamine at the cost of ATP and NADH. Serine is further reduced to glycerate, which finally gets phosphorylated by ATP in the chloroplast to 3PG. This is an inefficient process due to the loss of 3PG out of the CBB cycle, re-release of CO_2, additional consumption of energy and slow recovery time due to the diffusion between organelles.

The total carbon loss due to photo-respiration in plants is estimated at approximately 21% [31], and increases with temperature. Even in cyanobacteria, with a more controlled oxygen-exposure of RuBisCO, this loss is estimated at up to 1% [32]. Suppression of photo-respiration displayed an increased growth but proved unsustainable for the plants unless grown in an environment with increased CO_2 or decreased oxygen concentration [33]. However, there are indications of metabolic advantages to the pathway as it is linked to an increased availability of NADH, greater nitrogen assimilation capacity and is hypothesized to work as a protection from the oxygen radicals that can originate from an excess of photosynthetic energy [34].

As a vital process, the CBB cycle is well integrated in an organism's central carbon metabolism, as well as the gluconeogenesis and pentose phosphate pathway that comprise it. The cycle's activity is primarily dependent on the reaction rates of its enzymes, their catalytic equilibria, availability of metabolites [35, 36] and expression level of its encoding genes but it is also

regulated by the metabolic demands of connected metabolic networks. The most apparent catalytic bottleneck to the CBB cycle is RuBisCO as the vital carboxylating step has a low efficiency, though this drawback is partially negated in nature by its high expression. Transketolase [35] and transaldolase [37] were found to be the pathway's major regulators.

In *E. coli* cytosol, most of the involved enzymes' substrates are available at concentrations comparable to their documented K_M values, making the flux through the heterologous CBB cycle largely regulated by its metabolite concentrations [36]. The flux branch-points connecting the CBB cycle to other pathways are critical for controlling the metabolites coming in and out of it. The main two points of metabolite influx are F6P and 3PG as they connect the flow from the sugar and starch [37, 38] metabolism from glucose to pyruvate in the energy generating TCA cycle or the cell maintenance pathways involved with lipids and certain amino acids. Regeneration of 3PG in the CBB cycle requires energy carriers and flux in this direction is therefore dependant on the availability of ATP and NADH. F6P can also be shuttled towards the TCA cycle through the oxidative pentose phosphate pathway followed by the Entner-Doudoroff pathway which is an ancient alternative route to pyruvate yielding two ATP less than glycolysis [39]. Both F6P and R5P can supply flux towards the generation of nucleotides and histidine are both catalysed by the interconnecting TKT, which thereby largely regulates flux in the cycle. Transaldolase can also pose as the limiting reaction as it interconnects most of the metabolites of both TKT reactions. The E4P used by both these enzymes can also enter a pathway leading to the synthesis of aromatic molecules and amino acids.

In plants and algae, specific isozymes are expressed in the chloroplast stroma for localized optimization of the CBB cycle and regulate it with less interference from metabolically adjacent pathways. Little is known about their transcriptional regulation except that their operation is regulated by light intensity via the ferredoxin/thioredoxin [40] system. An additional method of regulation was shown to sense NADH availability via the CP12-GAPDH protein complex [38, 41]. An additional method controls carboxylation activity directly by expression of RuBisCO activase which catalyses the formation of a substrate blocking the RuBisCO active site, that is released upon ATP reduction, to prevent activity in the without a surplus of light-

derived energy [42]. Photosynthetic protobacteria often express the CBB cycle genes on a single operon, regulated by a single gene activated by the intra-cellular concentration of phosphoenolpyruvate [43].

In *E. coli*, the gluconeogenesis is regulated by the catabolite repressor activator (Cra) and cyclic adenosine monophosphate (cAMP) receptor protein (CRP) systems [44, 45]. The former directs flux from the TCA cycle towards the formation of sugars upon detection of fructose sugars, where the latter global regulating complex controls many genes, including those promoting the breakdown of starch and glucose for energy in the pentose phosphate pathway and glycolysis [46] when the absence of rapidly metabolizing sources of carbon is detected via cAMP. Both suppress reactions in the opposing direction, though *fbp* isozymes are not susceptible to this regulation [47]. Little is known about the expression of its pentose phosphate pathway genes except for the regulation of transaldolase and transketolase by environmental phosphate and the *)ppgpp-dksA* complex, which regulates the stringent response in reaction to nutrient limitations [48]. It is also largely regulated by *phoB* which is sensitive to the cytoplasmic availability of phosphates. The regulation of transaldolase by *CreCB* involves phosphoenolpyruvate [49] and RPI, which can be ribose-dependently regulated by *RpiR* [50].

1.3.3 Carbon concentrating mechanisms

The biochemical success of RuBisCO has led to several variations of the carbon-concentrating mechanism in order to to compensate for the innate inefficiency of the CBB cycle. These mechanisms are mostly understood to compensate for the limited carbon fixation rate and selectivity inherent to the carboxylating enzyme resulting from its evolutionary history [51, 52].

Carbonic anhydrase (CA) is an enzyme that allows a cell to increase the amount of CO_2 available in the cytoplasm by converting the freely diffusing gaseous form into the dissolved HCO_3^- (bicarbonate) and reverse the reaction according to its equilibrium constant in areas of lower relative CO_2 pressure. Bicarbonate has a much higher solubility in water than CO_2, allowing for cytoplasmic storage of a larger amount of carbon and for increasing the CO_2 availability near its areas of consumption. Furthermore, it enhances pH regulation which can be beneficial in the proton-dependent generation

of ATP across the thylakoid membrane. Eukaryotic phototrophs were found to assemble variants of the enzyme on the inner mitochondrial membrane to enable efficient utilization by their chloroplasts [53]. Bicarbonate transporters pump HCO_3^- across the membrane at the cost of pumping sodium or potassium in the opposite direction. In combination with carbonic anhydrase, this brings increased control over cytoplasmic CO_2 concentrations and a larger carbon fixing potential in times of photosynthetic activity. These transporters naturally occur in green algae and cyanobacteria [54].

Carboxysomes are micro-compartments with a polyhedral outer shell, found in cyanobacteria and proteobacteria. They are assembled from thousands of proteins that encapsulate RuBisCO and carbonic anhydrase proteins. Their pores allow for the diffusion of bicarbonate, RuBP and 3PG while resisting that of oxygen and CO_2. Transporting, isolating and trapping the CO_2 converted from the freely diffusing HCO_3^- by carbonic anhydrase inside the carboxysome increases the effective CO_2 concentration in the compartment. As variants of RuBisCO that evolved without CCM's have higher CO_2 affinities than those that did, but with a lower carbon fixation rate, the carboxysome allows for an increased efficiency while avoiding photo-respiration. *Pyrenoids* are chloroplastic micro-compartments functionally similar to carboxysomes encountered in algae and the hornwort group of plants [55]. Although the structure is estimated responsible for nearly half of global carbon fixation, little is known about the structure or composition of their CCM other than that it mainly contains linked RuBisCO proteins, is usually surrounded by starch and closely associates with the photosynthetic thylakoid.

Phosphoenolpyruvate carboxylase is an enzyme occurring in cyanobacteria and plants, able to capture dissolved bicarbonate and add it to phosphoenolpyruvate (PEP). This results in an easily transportable and a storable 4-carbon metabolite, typically malate or aspartate. Under light availability, this process is reversed in the proximity of RuBisCO to create a low oxygen/high CO_2 environment, improving its efficiency. This pathway requires more energy and is therefore more effective in warmer conditions, where CO_2 becomes less soluble and the generated ATP would otherwise get wasted on increased photo-respiration. In plants, the C4 metabolite is transported from mesophyll cells to photosynthesising bundle sheath cells and is named the

C4 carbon fixation metabolism. The crassulacean acid metabolism (CAM) fixation in plants employs a mechanism similar to C4 metabolism. Here, CO_2 is captured via stomata that are only opened at night, reduced to malate and stored in vacuoles. This CO_2 is then re-released by the enzyme and fixated with photosynthetic energy available by day, thereby preventing high water losses during carbon collection in arid climates [51]. Some CAM plants can also switch to either C3 or C4 metabolism, depending on season. The C4 metabolism has evolved at least 40 separate times [56], represents about 5% of all Earth's plant biomass but accounts for 23% of its carbon fixation [57]. These pathways require higher expression levels of enzymes natively present in C3-plants to minimize photorespiration while maximizing RuBisCO efficiency, and allow for a lowered RuBisCO expression.

1.4 CBB cycle engineering

1.4.1 Over-expression of native enzymes

Engineering the natural variants of the CBB cycle for increased efficiency could result in desirable mitigation methods to the current global increase of atmospheric CO_2. It has been shown that most of the enzymes involved in the cycle are available in abundance and require significant down-regulation to become limiting, there are however bottlenecks that impose limitations on its flux. SBPase has been shown to be a limiting reaction in the CBB cycle and up-regulating it increased carbon fixation in plants [58, 59] and cyanobacteria [60]. Overexpression of FBA has also been demonstrated beneficial in plants, and even more so under elevated CO_2 concentrations [59]. Increasing the expression rate of FBPase has yielded similar results [61]. Transketolase however has a more complicated effect in the cycle as both its up- and down-regulation have demonstrated to negatively impact carbon fixation [62]. Experiments in a cyanobacteria aimed to direct maximal flux into the formation of RuBP under mixotrophic conditions yielded a carbon output 136% of chemoautotrophic maximum [63]. This was achieved by the deletion of the *cp12* encoding gene and over-expression of glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase in the oxidative pentose phosphate pathway.

1.4.2 RuBisCO engineering

The overexpression of RuBisCO has also been met with mixed results as over-expression had no effect in plants but did increase total activity in cyanobacteria. This could be due to the observation that plant RuBisCO is only limiting under high light-intensities, increased temperatures or CO_2 concentrations [35]. Regardless, the enzyme has been a target for various engineering strategies to improve its catalytic properties as it could reduce the protein investment required for carbon fixation and prove beneficial under high irradiation or in combination with other CBB flux-increasing modifications. Efforts at mutagenesis have thus far yielded humbling improvements to reaction rate, at the cost of CO_2 selectivity [29]. This indicates that most RuBisCO variants investigated have already approached their maximal adaptation to their environment and that a trade-off reaction rate and selectivity exists in the enzyme [64, 65]. It has been shown that modifications to the thermal stability of RuBisCO activase can increase yields under specific circumstances [66]. Ideas to re-evolve RuBisCO's postulated ancestor [67] or engineer an improved RuBisCO [68] have been met with positive initial results.

1.4.3 Heterologous expression of CBB cycle genes

The introduction of non-native enzymes could increase the efficiency of an organism's CBB cycle due to evolutionary constraints posed on its pathway that might not have been present in its donor. The expression of a cyanobacterial FBA in a green microalga showed a $\sim$20% increased carbon fixation rate and cell growth [69]. Functional cyanobacterial and proteobacterial RuBisCO's have been engineered in *Nicotiana tabacum* without the need of extra chaperones [70]. Although requiring a heightened CO_2 concentration and demonstrating a lowered growth rate, the plants contained $\sim$10-fold lower amounts of RuBisCO with the expected superior catalytical properties. Expression of a cyanobacterial RuBisCO and its chaperones resulted in a near doubling of growth rate and bioproduction yield in *Cupriavidus necator*, making it a suitable platform for experimentation due to its redundant RuBisCO operons [71].

1.4.4 CCM engineering

Functional expression of RuBisCO, carbonic anhydrase [72], bicarbonate transporters and carboxysomes [73] in the heterotrophic *E. coli* have been reported, marking important steps on the way to bacterial CCM expression in plants [74]. The recombination of CCM's into plants could theoretically be superior to CAM and C4 metabolism as it wouldn't require intercellular transport, its associated metabolic conversions or energy requirements [75]. Successful expression of cyanobacterial transporters in a variety of plants has shown an increase in carbon fixation rate and biomass [76]. Other, similar, experiments demonstrated there are still knowledge gaps in this area of research [77]. The expression of cyanobacterial carboxysomes in tobacco resulted in their correct assembly in chloroplasts but the co-expression of the corresponding RuBisCO didn't associate with the structure [78].

Other efforts focus on the engineering of C4 or CAM metabolism in C3 plants native to hot/arid climates. As C4 metabolism has independently emerged at least 60 different plant species [79], it is regarded as a feasible effort. Despite progress, severe hurdles such as identifying the regulatory genes responsible for the phenotypical characteristics found in these plants need to be overcome [80, 81]. Despite the disappointing catalytic properties of contemporary RuBisCOs, it is not a majorly limiting factor under natural conditions as a relatively high expression, CCM's and tight regulation provide enough carbon and energy to the organism to compete in their respective milieus. Increasing the flux going into the CBB cycle and its efficiency without sabotaging its supporting metabolism would increase carbon fixation beyond evolutionary restraints. Efforts of engineering the photosystem, reducing antenna size and shifting its effective spectrum [82], or circumventing it entirely by providing the required electrons directly through electrodes have also shown promising initial results [83].

1.4.5 Synthetic carbon fixation pathways

Database mining from all documented enzyme reactions has shown that it's possible to construct various synthetic carbon fixation pathways that are at least in a few aspects superior to the CBB cycle. Redesigns of carboxylation pathways show great promise as the use of alternatives allows for higher catalytic rates with lower protein investments [52]. The CETCH cycle consists

of 13 enzymes originating from 9 different organisms. It is oxygen insensitive, fixes more CO_2 per photon, and requires less time and energy than the photo-respiring CBB cycle [84]. It has been demonstrated in vitro but its in vivo functionality would require drastic metabolic engineering and would be likely to see a decrease in efficacy.

Using an environmental CO_2 concentration of 20% and an RPE knockout, it proved feasible to reverse the heterotrophic direction of the *gnd* reaction to carboxylate Ru5P to 6-phosphogluconate in *E. coli*. The cycle is completed via the Entner-Doudoroff pathway, though overexpression of all three the (GED) enzymes was required for growth [85]. The reductive glycine pathway has expression problems in plants but has been demonstrated functional in *E. coli* [20]. Here, four genes are introduced to carboxylate formate and convert it into 3PG via serine, using a 3-step recovery cycle, in total consuming just 1 ATP and 3 NADH per fixed molecule of CO_2. A follow-up study enabled growth of the strain using CO_2 at a reduced concentration, using either formate or methanol as its sole carbon source [86]. Partial expression of the 3-HP bi-cycle proved functional in *E. coli* [87]. the full expression of which could allow for an aerobic carbon fixation of improved energy-efficiency.

Strategies to overcome the innate inefficiencies of the CBB cycle have largely focussed on the photo-respiration pathway, and have been met with some successes [88]. Expression of just glycolate dehydrogenase in plants showed increased photosynthesis and higher carbon fixation, even while the lacking tartronate semi-aldehyde reductase needed to facilitate metabolization back to 3PG, indicating a native plant glyoxylate metabolism able to complete the cycle [89]. *Synechocystis* cyanobacteria were shown to utilize this decarboxylation pathway to completely reduce glyoxylate to CO_2 in a plant-like C2 cycle [33]. The second pathway is less desirable as it releases CO_2 but might be advantageous under high light conditions as it requires no reductive power. An alternative pathway was developed to oxidise glycolate to malate at the cost of NAD(P)H and release the full molecule as CO_2 in a cycle with pyruvate and acetyl-CoA [90]. A more recent design employed the rational adaptation of new functions of an engineered glycolyl-CoA carboxylase and hetrolougously expressed tartonyl-CoA reductase. The pathway regenerates glycerate without the release of CO_2 or ammonia with com-

parative kinetic characteristics to the natural photorespiratory pathway [91]. Varioius promising theorized alternatives are still awaiting practical implementation of their designs [82, 92, 93].

1.4.6 Heterologous CBB in E. coli

It was previously shown that a functional CBB cycle could be expressed in *E.coli* [16] but simulation-assisted metabolic design has enabled the creation of an *E. coli* strain with a functional CBB cycle, able to use CO_2 as its primary carbon source [94]. Three genes are reportedly expressed as part of the same plasmidic operon. Regulated by a pTet promoter, the transposon-interruption of its *tetR* repressor mutated to effectively result in constitutive expression of the CBB operon. The large subunit of RuBisCO (*cbbM*) and carbonic anhydrase (*CA*) originating from *Rhodospirillum rubrum* and phosphoribulokinase (*prkA*) from *Synechococcus elongatus* complete the CBB cycle with the native pentose phosphate pathway and neoglucogenesis. Several additional gene deletions in the native genome were predicted as necessary to drive flux through this new pathway.

Knock-out of the two phosphoglycerate mutase (*gpmA/B*) genes severed the pathway between the carbon and energy metabolism, preventing the TCA from supplying 3PG to the CBB cycle instead of recycling it via RuBisCO. Phosphofructokinase was deleted to prevent it from phosphorylating F6P back to FBP in the opposite direction of the CBB cycle at the expense of ATP. Glucose-6-posphate 1-dehydrogenase (*zwf*) was removed to prevent flux escaping for the cycle through the Entner-Doudoroff pathway. Additionally, the *aceBAK* operon was knocked out to prevent carbon from entering via the glyoxylate shunt and tartronate semialdehyde pathway. With the carbon metabolism able to convert inorganic CO_2 into biomass, only the energy metabolism required an organic substrate. As it is directly usable in the TCA cycle, pyruvate was used as energy-source to drive the CBB pathway, this novel artificial trophic mode was named *hemi-autotrophy* [94] as part of its metabolism is still dependent on an organic substrate.

Despite its indication from metabolic modelling, the strain did not grow in a minimal medium culture containing only pyruvate and CO_2 as its substrates. Supplementing a low concentration of xylose sustained the culture while allowing the chemostat cultures to evolve to utilize its CBB cycle. In

the three experiments, strains took between approximately 60 and 120 days to adapt to these conditions and significantly increased culture density to allow continued xylose-independent hemi-autotrophic growth. Upon sequencing, the strains were found to have gained between 22 and 33 mutations. A single common mutation was found in the ribose-phosphate-diphosphokinase gene (*prs*) and cultures lost their hemi-autotrophic phenotype upon restoration of the gene. The phosphoenolpyruvate synthase regulatory gene (*ppsR*), inhibiting the expression of PEP synthase, was mutated in two of the three strains. It was found to improve CBB-dependent growth through an unknown mechanism but most likely by preventing the leakage of pyruvate into the now-futile pathway leading towards the phosphoglycerate mutase deletion. Several other mutations in flux branch points or carbon metabolism regulators, varying between experiments, were proven advantageous but not essential for directing flux towards the CBB cycle and obtain the hemi-autotrophic phenotype. In a fourth experiment, *ppsR* was knocked-out and the *prs* mutation replicated, the strain rapidly obtained six background mutations to support its new hemi-autotrophic metabolism [95]. Similar strategies have since been used to further develop the *E. coli* metabolism to utilize formate [96] instead of pyruvate.

As *E. coli* lacks SBPase, its FBA also doens't catalyse the formation of S7P as it occurs in plants, and an otherwise controlling branch of the canonical pathway is not present in the synthetic CBB cycle. The competing transaldolase and transketolase branch still metabolise S7P resulting in a balanced cycle with the same final stoichiometry per fixed carbon as the one found in photosynthesising organisms. Despite the strain fixing a significant portion of its accumulated biomass from atmospheric CO_2, various metabolites are synthesized from TCA cycle metabolites, and therefore from the supplemented pyruvate. It is possible that some of these metabolites leak into the carbon metabolism via unforeseen, undocumented or promiscuous reactions. By use of 13C-labelling it was demonstrated that no detectable quantity of carbon leaked from the carbon-fixation module to the energy module and vice versa. Furthermore it was calculated that $\sim$30% of accumulated biomass contained labelled carbon, in accordance with the predicted fraction of biomass generated from CBB cycle metabolites [94].

The reported hemi-autotrophic doubling time of approximately 9 hours

is a 13-fold increase compared to using glucose as a carbon source [97]. It is unclear whether or not this is directly due to its dependency on carboxylation, CO_2 diffusion rate, the metabolic impact of the other genomic deletions or the unoptimized supporting pathways. The same strategy also saw success in introducing the CBB cycle into *Methylobacterium extorquens* AM1 [98] and *Pichia pastoris* [99]. Both gaining their reductive power from methanol and both were also reported with severely lengthened doubling times. Nonetheless, the construction of this strain provides a suitable platform for proof-of-concept experiments in the manipulation of the CBB cycle in a relatively fast-reproducing, well-characterised, easily manipulatable, single-celled, photo-independent organism.

1.4.7 Isofunctional enzymes

The bottleneck reactions of the CBB cycle are catalysed by enzymes that are ubiquitously present throughout the metabolism of nearly every organism in the documented tree of life, of which only a small fraction has been investigated so far. Therefore, evolution might already have provided variants of these enzymes, able to relieve the imposed limitations on the presently investigated variations of the CBB cycle [22]. Functionally similar enzymes with superior catalytic properties for a specified reaction could be evolved or active sites engineered for specific substrates. Considering the size of the tree of life though, a wide variety of solutions is likely to already exist.

One interesting isozyme is the pyrophosphate-dependent phosphofructokinase found in various bacteria [100, 101]. The *pfp* gene (WP_010960535.1) isolated from the RHP-fixing *M. capsulatus* bath [102] has uniquely been documented to reversibly catalyse three reactions involved in its gluconeogenesis and pentose phosphate pathway: FBP to F6P, S7P to SBP and RuBP to Ru5P. The multi-functionality of the encoded PPi-PFK could explain its lack of encoding genes with homology to FBPase, SBPase and PRK [103]. Though kinetic characterisation of the enzyme demonstrated that both directions for each reaction occur in the same order of magnitude, a lowered overall reaction rate was found when compared to its CBB cycle homologs.

The lack of ATP-consuming PPi-PFK isozymes in glycolysis increases its energy-efficiency and the reported equilibria increase its metabolic flexibility, making flux directionality largely dependent on metabolite availability

for these reactions. The gene is encoded from a bicistronic operon, where the second gene *hpp* (WP_010960535.1) is a putative H^+-pyrophosphatase pump, suggesting an interdependent interaction between the two. Recombination of this operon into the CBB cycle has the potential advantage of saving ATP by producing inorganic pyrophosphate (PPi) from dephosphorylation of FBP and SBP while using it in the phosphorylation of Ru5P. The overal supply of PPi is generated by the membraneous H^+-pyrphosphatase pump (HPP) from the binding of two orthophosphate (Pi) molecules using the proton-motive force.

The *pfp* gene was successfully expressed and functionally purified from *E. coli* for characterisation using a coupled enzyme assay, though the functionality of *hpp* remains inferred. Despite the *M. capsulatus* optimal culturing temperature of 37°C, these kinetic assays were performed at 30°C [102]. These enzymes natively function in the RHP pathway of a methanotroph, and it is unknown how dependent their functioning is to these specific metabolic conditions.

1.5 Hypothesis and project objective

The hypothesis that this research is built on purports that the multi-functional PFK-PPi enzyme originating from *M. capsulatus* bath [102] functions in its ribulose monophosphate pathway and could therefore functionally substitute for the catalytic activities of its CBB cycle isozymes . Replacing the FBPase and PRK reactions with a single protein (while also introducing those of PFK and SBPase) while driving the latter reaction with reductive power originating from PPi instead of ATP would have several benefits. The most obvious advantage is a significant increase in the energy-efficiency of the CBB cycle by saving one ATP per fixed CO_2 molecule, and a decreased metabolic burden from protein maintenance would add to this advantage.

The CBB cycle has been demonstrated functional and stoichiometrically balanced in the relatively fast-reproducing, well-documented, easily manipulatable, single-celled, photo-independent *E. coli* [94]. The objective of this research is to evaluate the functional substitution of FBPase and PRK with PFP in the CBB cycle of a hemi-autotrophic strain of *E. coli* and test the feasibility of this hypothesized pathway, as shown in figure 1.2. Achieving functionality of the proposed synthetic CBB cycle would signify a leap in the efficiency of every process dependent on the pathway, competitive with any other carbon fixation research. It would also signify an increase in the applicability of a very adaptable CO_2-fixing strain with numerous potential downstream biotechnological applications.

Evaluation of the strain should be evident when cultured in comparison to the parental hemi-autotrophic CBB strain [94] as its enhancement would allow for an increased growth rate in the laboured strain. This 33% decreased expenditure of ATP in a pathway of major flux should also result in an increased growth density as the strain converts its limited organic carbon substrate into biomass with greater efficiency. When this objective is achieved, further applications of the system will be investigated.

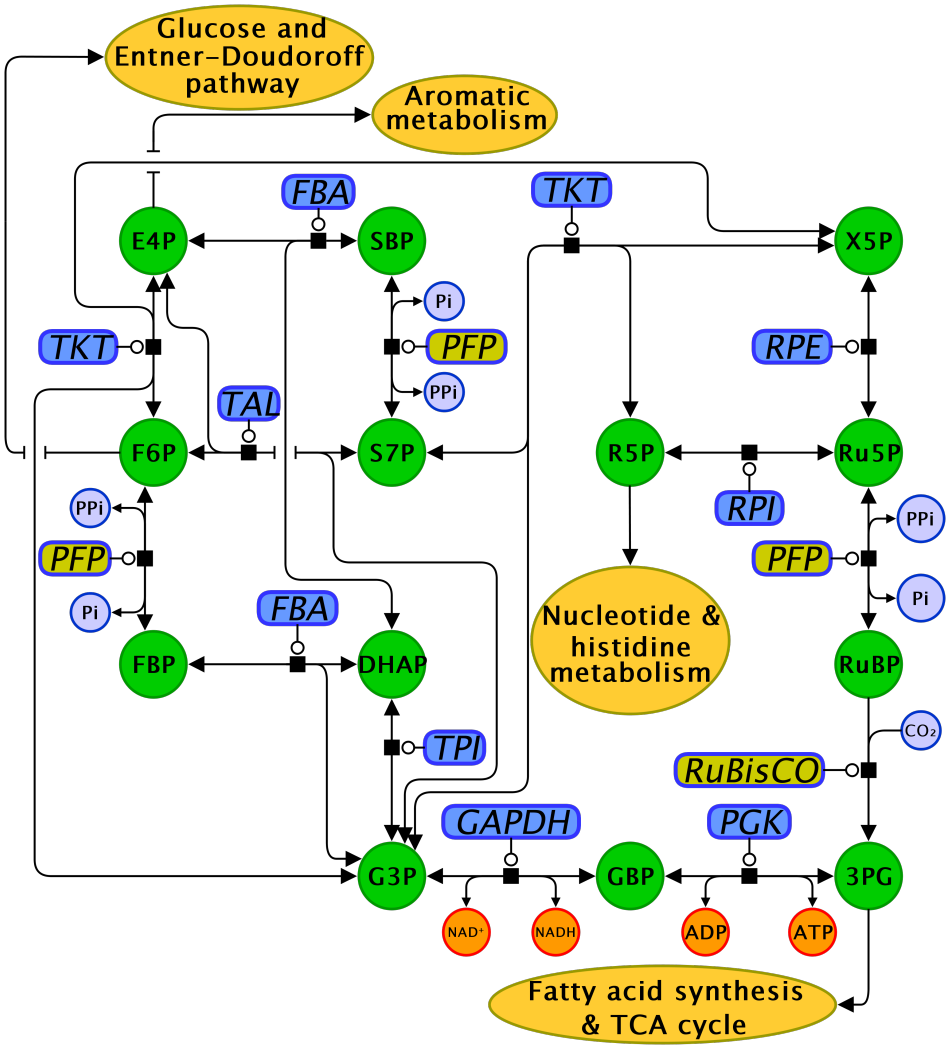

Figure 1.2: Schematic overview of the proposed synthetic CBB cycle in the hemi-autotrophic strain of E. *coli*. Non-native enzyme reactions are shown in yellow. The reactions of FBP and PRK are replaced with those catalysed by the *pfp*-encoded PFK-PPi.

Chapter 2

Materials and methods

> "Push a button, a light comes on inside, so it's a light. Push a different button and stick your hand inside, it burns you, so it's a weapon. Learn to open and close the door, it's a place to hide things. Never grasping what it actually did, and maybe not even having the framework necessary to figure it out. No monkey ever reheated a frozen burrito."
>
> *James S.A. Corey*
> *Abbadon's Gate*

2.1 Biological methods

2.1.1 Culturing

Standard microbiological techniques were used to reduce the risk of contamination of any handled culture [104]. Bacterial cultures were usually grown in autoclaved lysogeny broth (LB) medium (Fluka) with the addition of 30 μg/ml chloramplenicol, 50 μg/ml ampicillin, 20 μg/ml kanamycin and/or 100 μg/ml spectinomycin for marker selection. Incubations were performed at 37°C for up to 48 hours with temperature-insensitive strains and 30°C for up to 5 days with temperature-sensitive strains. Agar plates were prepared by the addition of 15 g/L Oxoid bacteriological agar No. 1 (ThermoFisher) to the medium. The plates were poured at an approximate temperature of 40°C after autoclavation. Antibiotics were added just before pouring and stored, solidified agar was re-melted in a micro-wave. For growth on plates, up to 50 μL of a liquid pre-culture was spread on an LB agar plate containing the same antibiotic concentration and grown under similar incubation conditions as liquid cultures.

Culture media growth was monitored by measuring the optical density at 600 nm (OD_{600}) with 1 mL of the inoculated medium on a Thermo Scientific Genesys 10S Vis in a 1 cm cuvette. Cultures were stored at -80°C after mixing to 20% (v/v) glycerol, inoculation from a glycerol stock was performed by scraping the still-frozen culture with a pipette tip and resuspending the scrapings in the appropriate culture medium with antibiotic.

For hemi-autotrophic cultures, M9 minimal medium was prepared just before use from several stock solutions to yield the following final concentrations in the working solution: 2.6 g/L $(NH_4)_2SO_4$, 1.0 g/L NH_4Cl, 0.5 g/L NaCl, 15.0 g/L $Na_2HPO_4 \cdot 12\ H_2O$, 3.0 g/L KH_2PO_4, 28.1 mg/L $FeSO_4$, 65.0 mg/L $EDTANa_2$, 1.8 mg/L $ZnSO_4 \cdot 7\ H_2O$, 1.8 mg/L $CuSO_4 \cdot 5\ H_2O$, 1.2 mg/L $MnSO_4 \cdot H_2O$, 1.8 mg/L $CoCl_2 \cdot 6\ H_2O$, 0.24 mg/L $MgSO_4$, 0.022 mg/L $CaCl_2$, 50 mg/L thiamineHCl, Biotin 5 mg/L, p-aminobenzoic acid, 5 mg/L, nicotinic acid 500 mg/L and 5 g/L pyruvate. The medium was CO_2-equilibrated in a Forma Scientific model 311 Direct Heat CO_2 incubator set to 37°C and 10% CO_2 for at least 40 minutes, or $NaHCO_3$ was added to a final concentration of 5 mg/mL. The medium was inoculated at $OD_{600} \approx 0.1$ and incubated for up to 2 weeks. The qualitative growth evaluation of a hemi-autotrophic culture was performed in 5 mL medium, including $NaHCO_3$ and antibiotic, in a 50 mL tube in a normal shaking incubator. Cultures for quantitative evaluation were grown in a 250 mL baffled flask with foam stopper, allowing for the aseptic diffusion of gases in a Kuhner LT-X shaking CO_2 incubator.

2.1.2 Batch evolution

Cultures were grown in M9 medium with 0.5% sodium pyruvate, 0.5% sodium bicarbonate and 30 µg/ml chloramphenicol in 5 mL of 50 mL screw-cap tubes. Opacity of the cultures was checked once a day and if cloudy, its OD_{600} was measured. Measurement times were not recorded and therefore assumed at the same time, offset by 5 hours depending on weekly shift, the estimated maximum deviation is 3 hours. Every culture was confirmed by polymerase chain reaction, confirming culture purity by both chromosomal knock-out of ppsR and the CBB genes on the plasmid by primers ”2868” and ”CA5”. The amount of cell divisions in a culture were calculated by subtracting the multiplication of the culture volume and its OD_{600} from its inoculation volume multiplied by its OD_{600} and dividing the whole by the log of 2.

2.1.3 Preparation of chemically competent cells

An LB pre-culture from a single colony or glycerol stock was grown overnight at 37°C, shaking at ∼200 RPM. 1 mL of this culture was inoculated in 100 mL LB and incubated under the same conditions until an OD_{600} of $0.3 \sim 0.5$ is reached. Then the culture was placed on ice for 15 min, followed by a centrifugation at 3300×g at 4°C for 10 minutes. The pellet was resuspended

Strain	Properties	Source
MG1655	K-12 *F⁻ λ⁻ ilvG⁻ rfb-50 rph-1 seqA-eYFP:kmR*	[105]
Keio*Δfbp*	MG1655*Δfbp*	[106]
DH5α	*F⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 Φ80dlacZδM15 Δ(lacZYA-argF)U169, hsdR17($r_K^-m_K^+$), λ⁻*	New England Biolabs
TOP10	*F⁻ mcrA δ(mrr-hsdRMS-mcrBC) Φ80lacZδM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ⁻*	ThermoFisher
BW25113	*$lacI^+rrnB_{T14}$ Δ$lacZ_{WJ16}$ hsdR514 Δ$araBAD_{AH33}$ Δ$rhaBAD_{LD78}$ rph-1 Δ(araB–D)567 Δ(rhaD–B)568 ΔlacZ4787(::rrnB-3) hsdR514 rph-1*	[107]
CBB4	BW25113 *serA(210Q), crp(M190K), malT(W317R), xylA(K264Q), glmU(1336+A), pgi(G378C), prs(310+AGTGATTGGTACACGAGC), ΔaceBAK, ΔgpmA, ΔgpmM, ΔpfkA, ΔpfkB, Δzwf*	[94]
CBB4*Δfbp*	CBB4*Δfbp*	This work

Table 2.1: The strains used in this work.

in 20 mL ice cold 0.1M $CaCl_2$ and kept on ice for 30 min. The cells were centrifuged as before and resuspended in 2 mL 0.1M $CaCl_2$ + 15% glycerol. The resultant competent cells were stored at -80°C until immediately before use.

2.1.4 Preparation of electro-competent cells

An LB pre-culture from a single colony or glycerol stock was grown overnight at 37°C, shaking ≈ 200 RPM. 1 mL of this culture was inoculated in 100 mL LB and incubated under the same conditions until an OD_{600} of 0.3 ∼ 0.5 is reached. Then the culture was placed on ice for 15 min, followed by a centrifugation at 4,000×g at 4°C for 15 minutes. The pellet was harvested and resuspended in 100 mL pre-chilled ultra-pure water and then harvested by another centrifugation. The pellet was then resuspended in 2 mL pre-chilled ultra-pure water containing 10% (v/v) glycerol, which was again harvested by centrifugation. A final resuspension in 300 µL pre-chilled ultra-pure wa-

ter containing 10% glycerol was divided in 1.5 mL tubes in 50 µL aliquots that were stored at -80°C until immediately before use.

2.1.5 Transformation

Competent cells were thawed on ice. 50 µL of an aliquot was mixed with 20 to 200 ng of purified plasmid DNA or a mix of ligated and/or assembled fragments and incubated on ice for 30 minutes. For chemically competent cells, a 42°C heat-shock was applied to the sample for 30 seconds and placed on ice for 2 minutes. For electro-competent cells, the sample was transferred to a 1 mm electroporation cuvette and shocked at 1.8 kV using the Bio-Rad MicroPulser electroporator, ensuring a pulse time of ∼5 milliseconds 950 µL of room temperature SOC medium (Invitrogen) was added (immediately after elctroporation) and the culture incubated at 37°C at ∼200 rpm for 1 hour. For high-copy number transformations the medium was spread undiluted onto an LB plate with the appropriate selection marker. For low copy number transformations, the medium was pelleted by centrifugation at 10,000×g for 2 minutes, resuspended in <50 µL and spread on the plate. The commercially obtained cloning strains DH5α (New England Biolabs) or TOP10 (Thermo Fisher) were used for the expression, validation and storage of modified constructs before their testing in a CBB-derived strain.

2.2 Molecular methods

2.2.1 Plasmid purification

Plasmid isolations were performed using the New England Biolabs Monarch Plasmid Miniprep Kit and protocol. For downstream use, 1 µL of the eluate was measured on a ThermoFisher NanoDrop Lite spectrophotometer for an approximate plasmid DNA concentration. Samples were typically stored at -20°C and handled on ice until use.

2.2.2 Restriction enzyme digestion

Reaction volumes of 10 µL were used containing 300 – 1,000 ng of DNA, 1 µL of 10× buffer and 0.5 µL of each required enzyme. Ultra-pure water was added to reach the total required volume. Enzymes and the appropriate buffer were checked for compatibility using information provided by NEB.

Plasmid	NCBI ref.	Properties	Source
pCBB	KX077536.1	*cbbM*, *prkA*, *CA*, chl^R, amp^R	[94]
pCP20		FLP recombinase, amp^R, chl^R	[107]
pKD46	AY048746.1	*gam*, *bet*, *exo*, thermo-sensitive replicon, amp^R	[107]
pKD3	AY048742.1	Template plasmid for chl^R	[107]
pKD4	AY048743.1	Template plasmid for kan^R	[107]
pCR2.1	222717	Cloning vector, amp^R, kan^R	ThermoFisher
pUC19	31851	Expression vector, amp^R	ThermoFisher
pMTL71301		pBAD:*ppa*, chl^R	Minyeong Yoo
pIB4		*cI857*, pR	Philippe Soucaille

Table 2.2: The plasmids used in this work.

Diagnostic digests were incubated at 37°C for approximately 1 hour, for further cloning up to 3 hours or at room temperature overnight.

2.2.3 Polymerase chain reaction (PCR)

Diagnostic PCR's were conducted using OneTaq Quick-Load 2× Master Mix with Standard Buffer (New England Biolabs), addition of 0.5 µL of forward and reverse primer, template (picked colony =0 µL), 1 µL LB culture, 10 to 100 ng of purified plasmid DNA or 50 to 100 ng of purified genomic DNA) and sterilized water to a total reaction-volume of 25 µL. The reaction was initiated at 94°C for 3 minutes. The mix was denatured at 95°C for 30 seconds, annealed at 45 to 68°C (depending on primer-pair) for 30 seconds and extended at 72°C for 1 minute/kbp, this cycle was repeated 33 times and finished with a 2 minute final extension. Cloning PCR's were conducted using Q5 high-fidelity 2x master mix (New England Biolabs), 1.25 µL of each primer, <100 ng of template and sterilized water to a total reaction volume of 25 µL. The reaction was initiated at 98°C for 30 sec. The mix was denatured at 98°C for 30 sec, annealed at 50 to 72°C (depending on primer-pair) for 30 seconds and extended at 68°C for 1 minute/kbp, this cycle was repeated 25 times and finished with a 2 minute final extension. Samples were stored at

4°C or frozen until use.

Name	Sequence
CBBsyn HiFi F	AATAATTTTGTTTAACTTTAAAGTTAAGAGGCAAGAA TGGCGGCGCGGAATGCTTTCTATG
CBBsyn HiFi R	AATTATTTCTATTAACTAGCTCAGAGCAAGGGCACGA TCAAGAGAG
pCBBs spacer F	TGCACGGCCGGCCGGTACCCA
pCBBs spacer R	TATGGGTACCGGCCGGCCG
pCBBs pTet F	GAAGATCAGTTGGGTGCACGGCCGGCCGGTACCCAT ATGCGGATTAGAAAAACAACTTAAATGTGAAAGTGG GTCTTAAAAGCAGCATAACC
pCBBs pTet R	GGCCTCGCCGGTCTCGTCCTGGGCCCGCCTCATGGC GTCGGCGACCAGGG
ci857 hifi f	GCACGGCCGGCCGGTACCCATATGTTAACCTATAAAA ATAGGCGTATC
ci857 hifi r	TGCATTCTTGCCTCTTAACTTTAGTACATGCAACCATT ATCACCGCC
Ci857 int r2	CTACGAGATGTATGAAGCGGTTAGTATGCAG
ci857 ext r2	GGTGAAATCGTCGGTGGTGCAGACC
Hifi3 R	GAAATAATTTTGTTTAACTTTAAAGTTAAGAGGC
Hifi3 F	TCCTTCCCCGGCGACGCCGAC
P15a NotI F	TTAGCGGCCGCTTCCGTCACAGGTATTTATT
P15a XhoI R	CTACTCGAGTAAGTTAGCTCACTCATTAGGC
Hifi3 R2	TAAAGTTAAACAAAATTATTTCTATTAACTAGCTTACG CCGG
cbbM site directed mutagenesis F	CTGGTCGCCGACGCCATGAGGCGGGCCCAG
cbbM site directed mutagenesis R	CTGGGCCCGCCTCATGGCGTCGGCGACCAG
Hifi FC pfp hpp F1	GGCGGGCGTCGAGATGGCTCTGAAAG
Hifi FC pfp E hpp R	GACAGAACCGAGCAAGCCATCGCCGCCGGTCCTGCT TATA

Hifi FC pfp E hpp F	TATAAGCAGGACCGGCGGCGATGGCTTGCTCGGTTCTGTC
Hifi FC pfp RBS hpp R2	GCCGCGAGCCCTCCCGATACGATCACGC
Hifi FC pfp prkA F	GGAAGGCGGGCGTCGAGATGGCTCTGAAAG
Hifi FC pfp E prkA R	CTTATAAAGTTAAACAAAATTATTCTAGATCAGACGCTGAACGAAGCGG
Hifi FC pfp E prkA F2	CCGCTTCGTTCAGCGTCTGATCTAGAATAATTTTGTTTAACTTTATAAG
Hifi FC pfp prkA R2	CTTCCCCCTGCGAATAAAGTTAATCTAGAACAAAATTATTTCTATTAAC
Hifi CI857 prkA R2	GAGACTTCGTGACCGTAGTAGGTGTCAG
HF CI857 D prkA R1	TGCATCTTCCCCCTGCGAATAAAGTTAAACAAAATTATTTC
HF CI857 D prkA F2	CTTTATTCGCAGGGGGAAGATGCATCATCACCATCACCAC
HF CI857 C prkA R1	GCATTCTTGCCTCTTAACTTTAAAGTTAAACAAAATTATTTC
HF CI857 C prkA F2	CTTTAAAGTTAAGAGGCAAGAATGCATCATCACCATCACCAC
HiFi pfp F	GAAATAATTTTGTTTAACTTTAGTTAGGCGCCAGCCAACGATACTCAAAG
HiFi pfp R	ATCGCCGCCGGTCCTGCTTAGTTGGTGTTACGGATTTGAGTGAGGAACTCAG
HiFi fbp F	GAAATAATTTTGTTTAACTTTAGTTGCGGTCGCTTTACTCCATAAACATTG
HiFi fbp R	ATCGCCGCCGGTCCTGCTTAGAGGCGCAGGAGATTACGCGTC
ppa KO cassetteF	TTGCAGGCGGTGGTGGAATAAGGTGTGTTTATTTATCGCGGGCATAAAAAAACCCTTACTAACCGAAGCCCGGCGTTCAGGGTTATTACGCCAGAAGAACGTGTAGGCTGGAGCTGCTTC
ppa KO cassette R	TGAAAAAGGAAAATGAAGACGAAAACAAGCGAAGACATTCGGCGCGAGTTGGCTATAATACTCGGCACTTGTTTGCCACATATTTTTAAAGGAAACAGACATGGGAATTAGCCATGGTCC
dupr spacer F	GATATCTCGAGCATATGGCCGGCC

dupr spacer R	TAAGGCCGGCCATATGCTCGAGATATC
dupr CI cloning F	GCTGCAGCCCGGGGGATCCCATGGTA CGAAGCAGCTCCAGCCTACAGGGCCCT
dupr CI cloning R	GCAACCATTATCACCGCCAGAGGTAAAATTG
dupr pfp cloning F	CCTCTGGCGGTGATAATGGTTGCACGCGTGT TTAACTTTAAAGTTAAGAGGCAAGA
dupr pfp cloning R	CCTTTCGTTTTATTTGATGCCTCTAGCACGCGTGTTAA ACAAAATTATTTCTATTAACTAGC
dupr fbp cloning F	CCTCTGGCGGTGATAATGGTTGCACGCGTCTCCATAA ACATTGCAGGGAAAGTTTT
dupr fbp cloning R	CCTTTCGTTTTATTTGATGCCTCTAGCACGCGTGGTCC TGCTTAGAGGCGCAGGAGA
dupr fbp native F	CTGCAGCCCGGGGGATCCCATGGTAACGTCACTATCC TTTTTTAA
dupr fbp native R	TTCGTTTTATTTGATGCCTCTAGCATTACGCGTCCGGG AACTCAC
dupr lacZ CI F	CCTCTGGCGGTGATAATGGTTGCAGGAAACAGCTATG ACCATGATTACG
dupr lacZ R	CCTTTCGTTTTATTTGATGCCTCTAGCATTATTTTTGAC ACCAGACCAACTGG
dupr lacZ native F	CCTCTGGCGGTGATAATGGTTGCGGCTTTACACTTTAT GCTTCCGGCTCG
dupr lacZ native2 F	CTGCAGCCCGGGGGATCCCATGGTAGGCTTTACACTT TATGCTTCCGGCTCG
dupr lacZ CI2 F	CCTCTGGCGGTGATAATGGTTGCCAATTTCACACAGG AAACAGCTATG
dupr fbp native2 F	CTTATTCGGCCTTGAATTGATCAAAACGCCTCTCCGTG TGGAGAGG
dupr fbp native2 R	TTAAGTTGTTTTTCTAATCCGCAACGTCACTATCCTTTT TTAATCTGCC
FLP fbpKO F	ATCGTCAGAGTTCGCCAGTGACAAGAGTCAACATTG CGCAGCAATAAAAACGCCTCTCCGTGTGGAGAGGCG CAGGAGATTACGCGTCCGGGAACTCACGGTGTAGGC TGGAGCTGCTTC
FLP fbpKO R	GTAATTTGCTGGCGCGATTCACCTGAGCGCAACATTG AGATTATTTGTTAAGATTGTTGCGGTCGCTTTACTCCAT AAACATTGCAGGGAAAGTTTTATGATGGGAATTAGCC ATGGTCC

FLP ppaKO F	GGTGTGTTTATTTATCGCGGGCATAAAAAAACCCTTA CTAACCGAAGCCCGGCGTTCAGGGTTATTACGCCAGA AGAACTTATTTATTCTTTGCGCGCTCGTGTAGGCTGGA GCTGCTTC
FLP ppaKO R	AAAAGGAAAATGAAGACGAAAACAAGCGAAGACAT TCGGCGCGAGTTGGCTATAATACTCGGCACTTGTTTG CCACATATTTTTAAAGGAAACAGACATGATGGGAATT AGCCATGGTCC
prkA sdm F	CGATCCGCCCGAAAAAATCGAACCCAATCG
prka sdm R	CGATTGGGTTCGATTTTTTCGGGCGGATCG
fbp term f	GCAAATTACGGATTATCCTGAAATGCGTTTCAAAAAT AAGAGTTACCATTTAAGGTAACTCTTATTTTT
fbp term r	TTACGCGTCCGGGAACTCACGGATAAAGCAAAAATA AGAGTTACCTTAAATGGTAACTCTTATTTTT
fbp term f2	AAAAATAAGAGTTACCTTAAATGGTAACTCTTATTTTT GAAACGCATTTCAGGATAATCCGTAATTTGC
prka sdm 2 F	AAAAAATCGAACCCAATCGCATCATTGTGATCGAG
prka sdm 2 R	CGGGCGGATCGATCAAGCC
prka sdm 2 R2	GGGCGGATCGATCAAGCC
ppa compl spec F	GCATTTGCAGGCTTCTTATTTTTATGGGCTAGCATGCG CTCACGCAACTGG
ppa compl spec R	CAACTTGCCCACTGGCCGGCCGGCTAGCACGCGTTC CATAGCGTTAAGG
prkA codon change F	GACAAGATCGAACCCAATCGCATCATTGTGATCGAG
fbp compl F	GAAACGCATTTCAGGATAATCCGTAATTTGC
fbp compl R	GCAGCAATAAAAACGCCTCTCCGTGTGGAGAG
LacZ tetR homology F	ATAGAAATAATTTTGTTTAACTTTACAATTTCACACAG GAAACAGCTATG
LacZ tetR pr F	CTGCAGCCCGGGGGATCCCATGGTAGACATTAACCTA TAAAAATAGGCGT
LacZ tetR pr R	CATAGCTGTTTCCTGTGTGAAATTGTAAAGTTAAACA AAATTATTTCTAT

Ppa Kan F	CAGGCTTCTTATTTTTATGGGACTGGGCGGTTTTATGGACAG
Ppa Kan R	CTTGCCCACTGGCCGGCCGGCGTCGCTTGGTCGGTCATTTCG

Table 2.3: The cloning primers used for this work. Primers are ordered in chronological order, primers may not have exclusively been used for their initially designed purpose as applicable.

2.2.4 Ligation

DNA fragments originating from a PCR reaction or gel extraction were inserted into the vector backbone using a total reaction volume of 10 µL containing 1 µL of T4 ligase (New England Biolabs), 1 µL of its 10× buffer and approximately 1 µg of DNA. Insert and vector backbone were added to the reaction in approximate equimolar amounts. Incubations were performed at room-temperature for 1 hour, or 2 hours for blunt-ended ligations.

2.2.5 Plasmid dephosphorylation

The reaction volume contained 1 µL of 10X buffer solution, 1 µL (5U/µL) of Antarctic phosphatase (New England Biolabs) and 8 µL of purified, restricted vector backbone. The mix was incubated for 30 min at 37°C, followed by a heat inactivation for 2 min at 80°C.

2.2.6 Gibson assembly

Assemblage of plasmid fragments was performed using either the New England Biolabs NEBuilder HiFi DNA assembly kit and manufacturers protocol. This is accomplished via homologous recombination and thus requires overlapping homologous regions between the assembling fragments. These were designed manually or using the NEBuilder online tool, amplified using PCR, then confirmed and purified from gel. Purified fragments were incubated appropriately with a 2:1 insert to backbone ratio and transformed into competent cells. Primers were annealed by diluting both nucleotides in ultrapure water to double the molar concentration of the backbone. In a PCR cycler the mixture was heated to 95°C and lowered to room-temperature over the course of 45 minutes, then added in a 2:1 ratio to the backbone in a conventional ligation reaction.

2.2.7 TOPO cloning

A PCR is performed using Q5 polymerase for high quality cloning and primers external to the region of interest, the amplicon is confirmed and purified by gel electrophoresis. Mix 6 μL of the pufiried reaction mix with 0.2 μL deoxynucleside triphosphates and 0.2 μL taq mastermix and incubate for 15 minutes at 72°C. Of this mixture, 4 μL is mixed with 1 μL of linearised pCR2.1 vector (Invitrogen), incubated at room temperature for 5 – 15 minutes and wholly transformed into TOP10 chemically competent cells. Insertion of the sequence of interest was confirmed by colony PCR.

2.2.8 Site directed mutagenesis

A 25 μL PCR reaction using 1 μL of a sample of isolated plasmid containing the mutation was run using primers specifically designed for its DNA sequence in need of correction. Q5 polymerase was used in an otherwise normal 25-cycle PCR program. The product resulting mix was either treated with KLD solution (New England Biolabs) and incubated at room temperature for 5 minutes or incubated with DpnI (New England Biolabs) for 30 minutes at 37°C to digest the mutated template. Finally, 5 μL was transformed into chemically competent cells and cultured while selecting for the plasmid. Results were analysed by subculturing of colonies, plasmid purification and Sanger sequencing of the previously mutated region.

2.2.9 P1 transduction

An overnight LB culture of the donor strain containing the gene knock-out from the Keio collection [108] was inoculated in 5 mL of LB + 0.2% glucose and 5 mM $CaCl_2$, and incubated for 1 hour at 37°C. 50 μL of P1 phage lysate was added to the culture and incubated at 37°C for up to 4 hours until the culture had clearly lysed. 50 μL chloroform was added to the culture and vortexed to lyse the remaining cells. After 10 minutes centrifugation at 4.500×g, lysate was extracted from the upper liquid phase and stored at 4°C until required. 2 mL of an overnight culture of the receptor strain was pelleted by centrifugation at 16,000×g for 2 minutes, and resuspended in 1 mL of 10 mM $MgSO_4$ (Sigma-Aldrich) and 5 mM $CaCl_2$. 100 μL of these cells were mixed with 100 μL of lysate and dilutions with sterile water thereof, controls without cells and lysate were made as well, and incubated at 30°C without shaking. 100 μL of 1M sodium citrate was mixed into each tube, fol-

lowed by 1 mL of LB medium and incubated at 37°C for 1 hour. The culture was harvested by centrifugation at 16,000×g for 2 minutes, resuspended in <100 µL and spread on a kanamycin LB plate.

The kanamycin cassettes replacing genes from the Keio collection are flanked by FRT sites. Removal of the cassette was performed by transforming the thermo-sensitive pCP20 plasmid encoding FLP recombinase and a temperature-sensitive origin of replication into the relevant strain and growing it on an LB + amp plate at 30°C. A single colony was inoculated in LB + ampicillin and incubated at 43°C to induce both the recombinase expression and prevent the plasmid from duplicating. Three serial dilutions of each were plated on LB without antibiotic and incubated at 30°C. Colonies from the resulting plates were picked and sequentially patched on LB + kanamycin, LB + ampicillin and LB, in that order to select for sensitivity to both, and grown at 30°C. Colonies sensitive to both were cultured for further work.

2.2.10 Phage λ-Red recombination

Used for the replacement of a chromosomal gene with an antibiotic marker, a target strain was transformed with pKD46. Transformants were grown at 30°C, as the plasmid contains a temperature-sensitive replicon, selection was performed by addition of ampicillin to the culture. Electro-competent cells were made from this culture while supplemented with 1 M L-arabinose during their growth to induce the expression of its recombinatory gam, bet and exo genes. PCR amplicons of the pKD3 (chloramphenicol) and/or pKD4 (kanamycin) with 100 bp homology to the gene of interest were eletroporated into the competent cells and selected for both antibiotics at 37°C. Successful exchange of the gene was screened by colony-PCR using external primers.

2.3 Analytical methods

2.3.1 Agarose gel electrophoresis

0.8-1% (w/v) agarose (Sigma) gel, dissolved in 1X TEA buffer (Lonza) and stained with 10,000× SYBR safe DNA gel stain (Invitrogen) were used for DNA fragments separation. Typically, samples containing 200∼800 ng of DNA mixed with 6× Purple Loading Dye (New England Biolabs) were

loaded as well as 4 μL of pre-diluted 1 kb or 2-log DNA ladders (New England Biolabs) as marker. Electrophoresis was performed at 80-120 V. and gels were imaged using a Biorad gel doc XR+.

Bands of interest for further cloning, after confirmation, were excised from the agarose gel on a Syngene LED Blue Light transilluminator, purified using the Zymo Research Zymoclean Gel DNA Recovery Kit according to the manufacturer's protocol, eluted in 10 μL and its concentration measured using a NanoDrop light spectrophotometer. Samples were stored at -20°C and handled on ice until use in downstream application.

Name	Sequence
FBP EXT F	TTATGGGGCTTCTGATCGACC
FBP EXT R	GTGCCGCTCATTTTACGCATA
FBP INT F (r)	ACTCACGGATAAAGCGTTCG
FBP INT R (f)	AACTCGACTTGTTCGCTAATGAA
KAN INT F	CTAGACTGGGCGGTTTTATGG
KAN INT R	GTCGATGAATCCAGAAAAGCGG
gpmA F	TCGCGGGTAAGGTTTAATTG
gpmA R	TCACAACAACCTTACACTGC
gpmM F	GGGTAACAACTCCCGACGTA
gpmM R	TTACGGGTGGCTTCAGAGAT
pfkA F	CTCTCTGCCTTTGGTTCAGG
pfkA R	TCGCTTGTTTACCTGAGCCA
pfkB F	TTCGTATTGCCTGATTGTGC
pfkB R	CTATTCCATTCCTCCCAGGT
zwf F	GCGAGATAAGCAGAGCGAGT
zwf R	GCAGTTTTGTCAGATATTACGCC
aceBAK F	TATTTCCCGCACAATGATCC
aceBAK R	AGCCCCACAAAAATACTCTG
New ppsR For	ACATAACGTTCTACGCTTCTGTGCG
New ppsR Rev	TGCTGATTTCAAGTGCGAGGTGTG
pfp 5pr Rev	CCCTTTGAGCTTGTAGCGGCAGGAAC
hpp 3pr For	GTCCTCGTGGGGCTGCTGCTC
CA 5pr Rev	AGACGGTTCAGGGCGGCGCGAA
ppa ext. F	GGATCTACACGCCCGGCTGGCAAAGC
ppa ext. R	CCACATAGGGTTGTCCTCGTCGGGGTAG
Cm int F	CCACCGTTGATATATCCCAATGG
Cm int R	CCATCACAAACGGCATGATGAACC

Table 2.4: The analytical primers used for this work. Primers are ordered in chronological order, primers may not have exclusively been used for their initially designed purpose as applicable.

2.3.2 Sanger sequencing

The sequencing of PCR and plasmid samples was performed out-house, by Eurofins Genomics. Any DNA sample sent to them originated from a media culture inoculated or sub-cultured from a single colony. These samples contained 15 µL of 50 ng/µL DNA and 2 µL of 10 µM of the appropriate primer. Chromatograms of the sequencing results were obtained, aligned to the *in silico* construct, and differences from the plasmid map were checked for errors or ambiguous peaks in its raw data and corrected where necessary. At times low quality results were obtained but samples were similarly re-sent until a sample of usable quality was obtained.

Name	Sequence
cbbM For	GCACCAACGTCGAGGTCTGCACCAC
cbbM Rev	CCAGGTTCTCGAAGAAGCCGGGC
up cbbM For	GGTGCAGAGCCAGCCTTCTTATTCGG
pfp hpp seq 700	AGGTCCGGTCACATGGCTATTGTTCG
pfp hpp seq 1400	GGAATTTCTGGCGCTCGGCCTG
pfp hpp seq 2100	TGACCATCGTCGCCACCATGCT
pfp hpp seq 2800	ACACCACCAAGGCGGTCACCAA
pCBB AN 1442BP	CCCTTTCGTCTTCACCTCGAGTC
pCBB AN 2165BP	CTGGGCGGCGACTTCATCAA
pCBB AN 2868BP	ACGGGGTTCCGGTTCTGGACTAT
pCBB AN 3583BP	CAAGGCCTACATTGAGCCC
pCBB AN 4294BP	AATGAAATCGTCGGCATGCAGC
pCBB AN seq	CCGCTCATGAGACAATAACCCTGATA
pCBB AN seq4973	CTGGCATTTCAGTCAGTTGCTCA
pCBB ANseq30	CTGAGTAGGACAAATCCGCCG
pCBB ANseq631	CAAGCCAGTTACCTCGGTTCAA
pCBB AN seq rev	GCGATCAGATCCTCTTCCTTGAG

pfp400 seq	GATTCCGCCGATACCTGTCTGAAAG
pfp rev	GGTGATCGGCAGGTCGTTGTCC
fbp 940 F	GAAGATGTCGAACGCTTTATCCGTGAG
seq ORI R	GGTCTAGGGCGGCGGATTTGTCCTAC
seq CI int	CCCACAACGGAACAACTCTCATTGC

Table 2.5: The sequencing primers designed for this work. Primers are ordered in chronological order, primers may not have exclusively been used for their initially designed purpose as applicable.

2.3.3 Genome sequencing

To obtain high quality genomic DNA, the Sigma Aldrich GenElute Bacterial Genomic DNA Kit was used according to its protocol. The obtained DNA sample was run on a 0.7% agarose gel for $\sim$3 hours and confirmed to have little to no smear below its heavy-weight band, its concentration was determined on a NanoDrop photospectrometer. A single tube containing 100 µL of 170 ng/mL DNA was sent to MicrobesNG for sequencing using their standard service. The obtained reads were mapped using the *E. coli* K-12 BW25113 genome, sequence CP009273.1, as a reference. This alignment and its subsequent variant calling of the obtained genome data was performed with Geneious Prime Build 2022-03-15 11:43, using a minimum variant frequency of 0.2 for its SNP analysis. The alignment of reads to the reference was manually checked for unannotated gaps, none were found. For the pCBB plasmid, the GenBank sequence KX077536.1 was used as reference.

2.3.4 β-galactosidase promoter assay

Cultures of the appropriate strain harbouring a pCBB plasmid, independently expressing the *lacZ* gene under the control of the investigated promoter, were grown under standard conditions. The OD_{600} of overnight cultures were measured and 20 µL of homogenous culture (or dilution) was added to a tube containing 80 µL of permeabilization solution, consisting of 100 mM Na_2HPO_4, 20 mM KCl, 2 mM $MgSO_4$, 0.8 mg/mL hexadecyltrimethylammonium bromide, 0.4 mg/mL sodium deoxycholate and 5.4 µL/mL 2-mercaptoethanol. These mixtures were incubated at 30°C for 20 to 30 minutes, after which they were transferred to a 2 mL cuvette, 600 µL of substrate solution was added and the time accurately measured. This solution consists of 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 1 mg/mL o-nitrophenyl-β-D-Galactoside and 2.7 µL/mL 2-mercaptoethanol. Upon the visible yellowing of the solution, 700 µL stop solution consisting of 1 M

Na_2CO_3 was added and the time carefully noted. In order to prevent overshooting of the saturation-point cultures were measured in duplex, with the first stopped at the first sight of any colour-change and the second a 20∼40% greater interval. Development of the stopped reaction was measured at OD_{420}. Dividing this value by the original culture density multiplied by culture sample size, multiplied by the measured reaction time, multiplied by 1,000 yields the culture's *lacZ* activity in Miller units.

2.4 In silico methods

2.4.1 Gene sequence analysis

For the prediction of promoters in a sequence, the web-tool bprom [109] was used. For rho-indpendent transcription terminating sequences the web-tool ARNold [110]. The rho-dependent tools found didn't yield unique results. For RBS and translation rate prediction, Salislab's RBS calculator version 2.1.1 was used with *E. coli* MG1655 as set organism [111]. The Salis lab promoter predictor was excluded as it produced noisy data that didn't directly correlate to verified annotations.

2.4.2 Sequence alignment

For the comparison of a known sequence to unknown sequences the BLAST web-tool [112] was used to find similar sequences, allowing specification of organism, accuracy and sequence translation. The (multiple) sequence alignments of known sequences were performed using ClustalO [113]. The alignment of obtained sequencing data to plasmid maps, constructed in Ugene [114] were performed in the web-tool Benchling [115].

2.4.3 Thermodynamic pathway analysis

The web-tool eQuilibrator [116] version 2.2 was used to obtain thermodynamic data of chemical conversions, compare the free energies between them and combine the results into an pathway overview. The tool uses standard conditions of 25°C, 1.0 bar, pH = 7.0, ionic strength = 0.1 M and 1mM concentration for all metabolites.

2.4.4 Kinetic pathway modelling

The software tool COPASI 4.27.127 [117] was used to simulate enzyme kinetics by Michaelis-Menten and mass action equations, with a single (ir)reversible molecular conversion per reaction, necessitating the omission of complex interactions. For reactions without available data on closely related paralogs, the thermodynamically derived k_{eq} was used and if larger than 10^4, the reaction was considered irreversible. For mass action equations, the k_{cat} was derived from V_{max} values using $k_{cat} = V_{max}/[Et]$. Reaction rates were calculated by V_{max} = ((K_{cat}·1,000) / monomeric molecular weight)·60. The initial metabolite concentrations were inserted as previously described [36] and reaction rates multiplied by protein copy numbers [118]. Only the main isozymes were considered and molecular weight or cell volume corrections were performed on the reactions involved, making the time axis on the simulations arbitrary.

A kinetic model of *E. coli* central metabolism [119] was imported and adapted using COPASI. Knock-out reactions were omitted from the model and new reactions were implemented using Michaelis-Menten reactions using best available kinetic parameters. If unavailable, a k_{eq} derived from Gibbs free energy differences between substrates and products obtained from eQuilibrator was used. The equation $\Delta G° = -RT \cdot \ln \cdot k_{eq}$, was used for this conversion with T = 310.1°K and R = 8.314 J/K·mol. For introduced reactions, documented V_{max} was initially used but as protein expression isn't accounted for, this was included in the variables to optimise for. Changes were balanced out by the optimization function and particle swarm method on the variables to find the objective function set to maximise flux through the growth reaction, conditional on a steady-state model. Model changes without optimization were kept to the absolute minimum. Swarm size was set to 50 and standard deviation to 1×10^{-6}. This process was considered as failed if no progress of its solution was observed during the first 400 evaluations and considered completed with 1,000 to 2,500 evaluations with less than 10% increase from previous in each of the last 2 solutions found. Each optimisation was performed on a single core of an AMD FX-6300 processor and would typically take 1–3×10^4 CPU-seconds.

2.4.5 Visualization

Krayon4SBGN [120] was used to manually draw metabolic pathways of the integrated synthetic CBB model, using the published model [119] and classic pathway as template. The creation of chematic representations, image editing and data assembly were performed using LibreOffice [121] and GIMP [122], this text was finally compiled in LaTeX [123] using TexStudio [124].

Chapter 3

Optimization of hemi-autotrophic cultures

"Power corrupts. Knowledge is power. Study hard. Be evil."

Unknown

3.1 Introduction

As the CBB-dependent growth of the hemi-autotrophic strains of *E. coli* only emerged after extended chemostat evolution and were halted almost immediately after their adaptation, the obtained CBB4 strain exhibits a reported doubling time of 20 hours [94]. The typical doubling time of one hour for *E. coli* in minimal medium [97] indicates opportunities for significant optimization of hemi-autotrophic growth. As pyruvate and CO_2 are readily available under these culturing conditions, the introduced CBB pathway, the metabolic modifications that support it or the used culturing conditions can be further optimized to increase growth characteristics [72].

The objective of the experiments described in this chapter is to improve on the growth characteristics and experimental throughput of the hemi-autotrophic strain of *E. coli*. Increasing the strain's viability would decrease the delay in experiments utilizing the strain, increase its experimental throughput and form the basis for functional modifications of its CBB pathway. Such improvements are likely to also improve carrying capacity and to enable same-media transformations. Developments to the CBB strain are measured first by the ability to maintain hemi-autotrophic growth and then compared to the growth rate, density, viability and transformability of the parental CBB4 strain.

3.2 Optimization of growth conditions

The CBB4 strain used in this work was evolved in a chemostat fed by M9 minimal medium supplemented with 5 g/L sodium pyruvate and 10% CO_2 with a doubling time of 20 hours. The strain grew on agar plates with the same

composition successfully, though it was prone to contaminations and sub-culturing remained unreliable. It was not feasible to re-constitute a hemi-autotrophically grown culture from glycerol stock directly back to these stringent conditions, so an intermediate culture in selective lysogeny broth (LB) medium was required for sub-culturing to CBB-dependent growth. 50-mL screw-cap tubes containing 5 mL of medium proved highly reliable with negligible contaminations and was manageable in the required culturing quantities without the need for larger-scale cultures sealed with a foam stopper in a CO_2-incubator. Electromagnetically stirred flasks containing 10 mL of medium sealed by foam stoppers in a CO_2-incubator proved unreliable and prone to contamination. 2 mL cultures in 5 mL tubes supplemented with 5 g/L of sodium carbonate in a shaking incubator resulted in fewer contaminations but at higher rates of culture failure. The found contaminations occurred despite the careful usage of antiseptic techniques [104] and were resistant to the chloramphenicol and kanamycin most commonly used in this project. Their origins were not closely investigated but were eventually mitigated by the avoidance of working in labs where experiments with sporulating organisms was also performed.

A hemi-autotrophic culture of the CBB-utilizing strain of *E. coli* is typified by a dull-white haze that shows a slight yellow tinge when pelleted. The optical density at 600 nm (OD_{600}) reached in batch cultures of this strain grown in hemi-autotrophic medium was typically between 1.0 and 1.2, and would be reached in 5 to 10 days after its inoculation. Consecutive sub-cultures would occasionally see improvement to both these characteristics. Contaminants were identifiable by a distinguishable difference in medium colouration, growth without increased CO_2 pressure, faster growth or higher cell density as they can utilize the supplied pyruvate more efficiently than the CBB strain. Regardless, successful cultures of experimental significance were verified by polymerase chain reaction (PCR) using primers external to the chromosomal *ppsR* (replaced with *kanR* in CBB4) and *prkA* on pCBB, as examplified in figure 3.1.

Batch cultures were reliably grown in 50 mL of a 250 mL baffled flask inside of a shaking incubator maintaining 10% CO_2 and were typically inoculated from LB cultures at an OD_{600} of approximately 0.1. High-pressure liquid chromatography analysis of the fully grown and uncultured medium

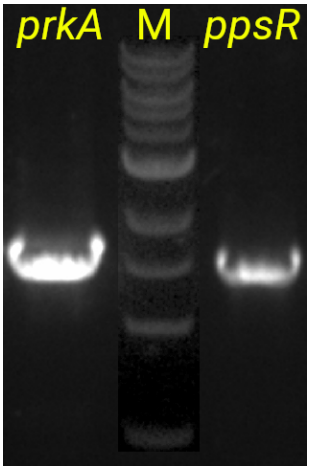

Figure 3.1: A typical agarose gel showing the PCR amplicons of *ppsR*, located on pCBB, and the chromosomal *ppsR*. Both were amplified using primers external to the genes of interest. The 1kb DNA ladder was used as marker (M) with fragments of 10.0, 8.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.5, 1.0 and 0.5 kbp.

showed a 115-fold decrease in pyruvate concentration to 37 µM, indicating pyruvate as a limiting factor under the used hemi-autotrophic conditions. Data suggested the pre-culture medium, its time in stationary phase and inoculation density contributed to the sub-culture's stationary density and lag phase. Cultures expressing the CBB strain but inoculated from either an LB or hemi-autotrophic pre-culture resulted in divergent growth profiles, as shown in figure 3.2. Plasmid-cured cultures of the CBB strain, re-transformed with pCBB and grown from a LB pre-culture (fig. 3.2a) typically reached stationary phase slower than a similarly cultured plasmid-expressing stock would(fig. 3.2b). The hemi-autotrophically pre-cultured strains would reached their stationary phase significantly faster but proved highly dependent on the growth phase of its pre-culture, as illustrated by a culture shown in figure 3.2c. Their maximum doubling times would similarly differ in the culture's adaptation to its medium.

Several modest improvements to the hemi-autotrophic culture media were discovered over the duration of experimentation with the strain. The addition of 5 g/L sodium bicarbonate that allowed for growth inside tubes also made pre-incubation of plates and larger cultures dispensable. Hemi-autotrophic medium wasn't prepared from it's stock solutions until the day it was needed as long-term storage of even sterile-filtered M9 medium caused its precipitation. The usage of ultra-pure water instead of autoclaved,

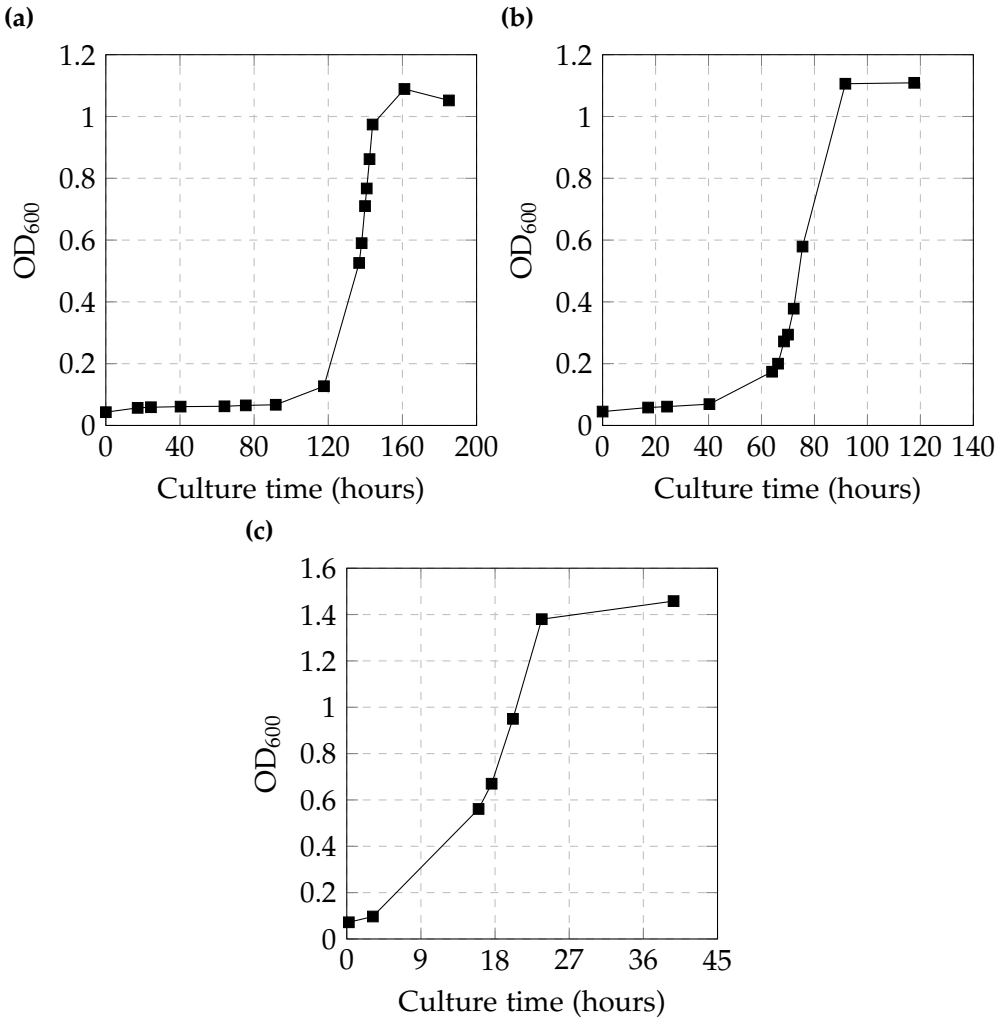

Figure 3.2: Hemi-autotrophic growth-curves of CBB-utilizing *E. coli*. **(a)** re-transformed CBB strain pre-cultured in LB medium and inoculated at $OD_{600} = 0.043$ was measured with a minimum doubling time of 6.86 h. **(b)** Inoculated at $OD_{600} = 0.045$ from LB-grown pre-culture and with a minimum measured doubling time of 5.42 h. **(c)** Inoculated from hemi-autotrophic pre-culture with an initial OD_{600} of 0.072 and a minimum measured doubling time of 5.069 h.

de-ionised water and buffering of the phosphate- and ammonium containing stock to pH 6.8 prevented precipitation issues. These modifications had no influence on subsequent culturing as its final pH was measured to stay between 7 and 8 in a grown culture. Doubling the CO_2 concentration saw no significant effects on growth rates or final densities of both plate and media cultures, halving them resulted in a decrease of culture density. Similarly varying pyruvate concentrations only resulted in decreased growth.

The highest temperatures tested for the hemi-autotrophic strain were 40°C for plates and 39°C for media cultures, where the lowest was 30°C for media; all grew successfully, though no growth-curves were recorded for these non-standard culturing conditions. Using a previous report [125] these temperatures were deemed as a practical limit for the cultures considering the combined viability reduction from a temperature increase and hemi-autotrophic metabolism. Addition of 0.2% (v/v) xylose to hemi-autotrophic plates saw a great increase of colonies but didn't lead to a subsequent increase in colony count after sub-culturing to plates without. Chloramphenicol was later added to both media and plate cultures to increase the selective pressure on plasmid maintenance. All otherwise reported cultures were (re-)cultured with hemi-autotrophic media containing only ultra-pure water, the buffered stock solution, sodium bicarbonate and antibiotic selection.

3.3 Transformation optimization

The transformation of CBB strains harbouring pCBB (and its derivatives) by heat-shock proved relatively time-consuming and culturing via rich media prone to mutations as elaborated in section 4.3 and 3.4.1. Culturing transformants from rich to the selective M9 medium, and potentially to plate, could take between one and three weeks for the established pCBB plasmid as shown in figure 3.3. The rich plates selected for the plasmid by antibiotic resistance and grew large numbers of colonies, inoculation from LB to hemi-autotrophic medium was reliable but time-consuming while its inactive CBB operon proved prone to mutations. Reduction of the strain's adaptation time could greatly increase experimental throughput and efficiency while decreasing sample handling.

In attempts to finding a method of transformation directly to selective media, the recovery was performed for 2.5 hours in either SOC, M9 + pyruvate or M9 + pyruvate + 0.5% xylose. This was followed by washing with medium and split over hemi-autotrophic plates, with or without the addition of 0.2% xylose. These experiments did not result in any colony growth. A follow-up, extending the SOC recovery to 16 hours before washing and inoculation into hemi-autotrophic medium didn't yield colonies either. Competent cell stocks prepared 1 day, 2 and 6 months before re-transformation with equal concentrations of pCBB showed no significant

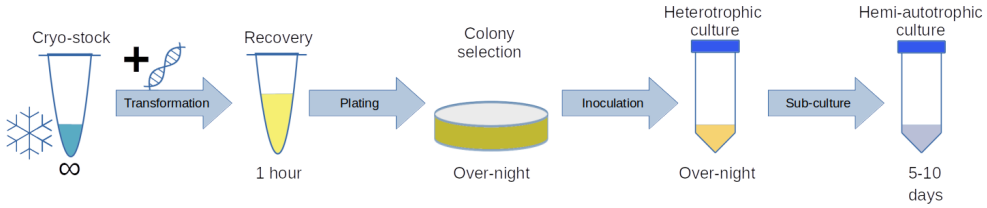

Figure 3.3: A schematic overview of the established method for transforming pCBB or its derivatives and brining it into hemi-autotrophic culture. A cryo-stock of competent CBB cells is incubated and transformed (typically by heat-shock) with the plasmid of interest. A one hour recovery in SOC medium is plated with a selective antibiotic, grown over-night and a colony picked for over-night culturing in LB medium. Depending on experiment, colony PCR's could be performed after the plating step to screen the desired genotype. After accumulating biomass in the rich medium a OD_{600} equivalent of approximately 0.1 would be inoculated into the restrictive hemi-autotrophic culture and incubated until growth is observed.

reduction in their transformation, subs-culturing efficiency or growth speed. No difference was found in the transformation efficiency or sub-culturability between chemically transformed of electroporated cultures of re-transformed CBB4 cells either.

Assuming the metabolic state of competent cells was responsible for these inefficient transformations, an alternative method was devised without the need for changing medium. The pCBB chloramphenicol resistance gene was exchanged for a spectinomycin resistance gene (*aadA1*) by restriction of pCBB. Digestion by NheI and MluI resulted in a pCBB backbone corresponding tot he predicted 7367 basepairs (bp) and a $spec^R$ fragment approximating the calculated 1013 bp. Growth in spectinomycin-containing media and subsequent sequencing confirmed several purified plasmids had ligated correctly in all eight tested colonies. For transformation, a hemi-autotrophic control culture expressing pCBB was pelleted, washed three times in ultra-pure water, resuspended in 10% glycerol and electroporated with the newly constructed plasmid. The recovery was performed in SOC medium, washed and transferred to plates and media incubated under hemi-autotrophic conditions. Four separate attempts using this method, two of which were specifically harvested in their log phase, didn't yield any transformed colonies. A final heat-shock transformation via rich medium demonstrated that spectinomycin-dependent growth on both hemi-autotrophic plate and media were viable using this plasmid.

3.4 Batch evolution

An increase of performance was observed in hemi-autotrophic sub-cultures. In order to evolve a strain with improved growth rate, viability and transformability, a batch of the received CBB strain 4 [94] was consecutively sub-cultured. The hemi-autotrophic media were inoculated into 50 mL falcon tubes containing hemi-autotrophic medium with added sodium bicarbonate in a shaking incubator as established in section 3.2. To increase selective pressure and reduce the introduction of contaminants, these cultures were grown in the presence of chloramphenicol. A total of 37 subcultures were performed over 136 days. Optical density and the volume propagated in the next culture were measured at estimated times. Using this data as displayed in supplementary table C.1, the inoculation density, average growth rate and doubling time between inoculations were retro-actively approximated. The estimated doubling time is plotted for each of the sequential subcultures in figure 3.4a, showing a gradual decrease in doubling time.

Culture 16, inoculated at day 67 of the series, was partially grown at 30 degrees due to the sudden inaccessibility of the lab and this is clearly reflected in its calculated doubling time. The estimated number of cell divisions occurring in each culture is shown in figure 3.4b and indicates a lasting negative effect that was not recovered from until about 6 sub-cultures later. This could be indicative of a reverted adaptation in the culture's or its subpopulations. Overall this metric of the CBB4 strain was also shown to have improved during the course of the batch evolution experiment.

Smaller inoculation delays due to restricted weekend access generally show lower estimates of growth rate than the others. Facilities for continuous monitoring weren't available so a culture's approach to maximum density was estimated from its inoculum size and adjusted to approximate a culture time of 3-4 days. Despite the experiment's limitations, the development of the culture's characteristics and a decrease in the data variability is apparent. Inoculation volume has shrunk by 1,000-fold, growth time per inoculum by nearly 100-fold, growth rates doubled and final culture densities saw a moderate increase. The growth curve of a hemi-autotrophic sub-culture of the 37th generation strain showed a significant increase in specific growth rate as its measured doubling time of 3.96 hours, shown in

figure 3.4c, was over an hour below that typically measured of its parental strain.

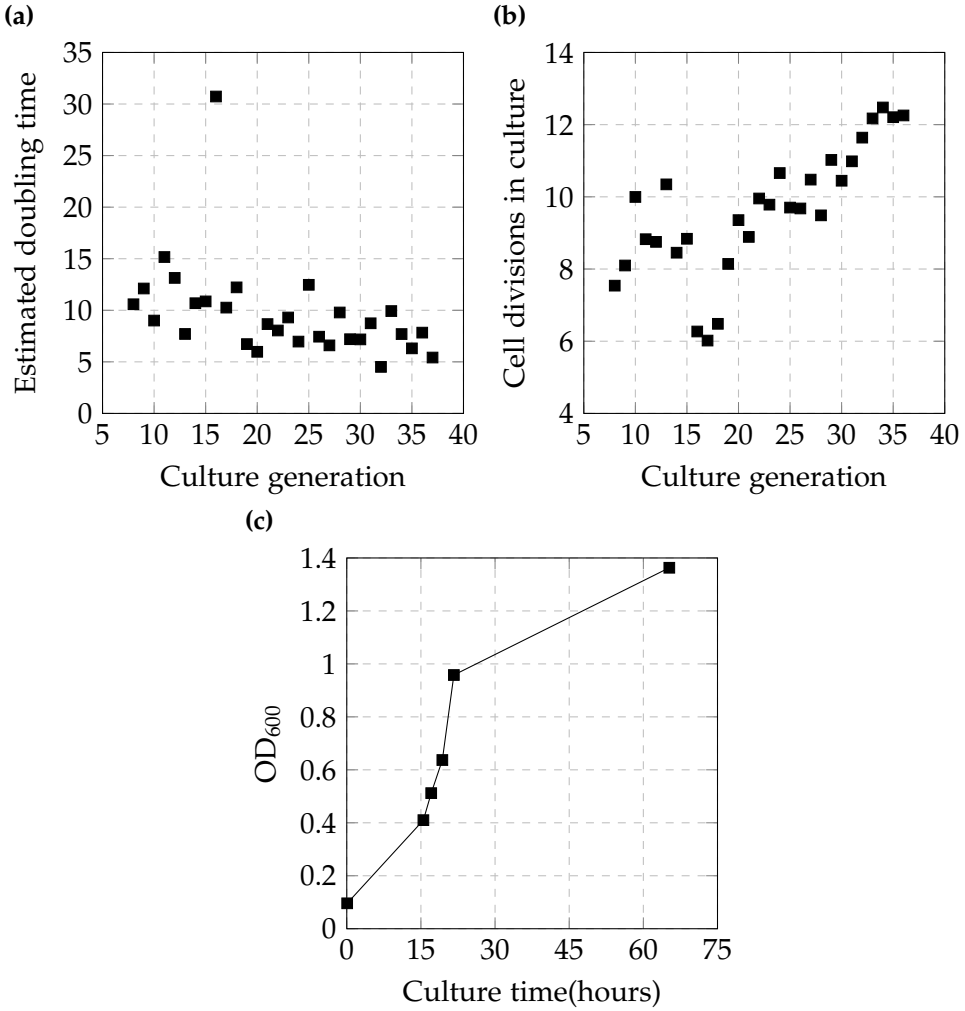

Figure 3.4: Measured growth characteristics of the batch-evolved hemi-autotrophic strain. **(a)** Estimated doubling time over culture generations (the first six were unmeasured). **(b)** Each culture's growth time divided by its estimated specific growth rate, plotted against its sub-culture generation. The r^2 of its linear trend line $= 0.640$. **(c)** Hemi-autotrophic growth curve of the final batch-evolved strain, re-cultured from a 5% DMSO cyro-stock. Minimum observed doubling time was measured at 3.96 hours. Inoculation density of $OD_{600} = 0.096$, from a hemi-autotrophic pre-culture.

A glycerol stock taken at generation 26 was sub-cultured while only selecting for its chromosomal kanamycin resistance. Two generations without the selective pressure to maintain its plasmid saw its loss in of over 99% of colonies when compared to chloramphenicol cultures. Picking and patch-

ing these colonies confirmed loss of the plasmid in specific colonies, two of which were used for the subsequent preparation of chemically competent cells. Re-transformation with a fully sequenced pCBB stock, followed by a recovery in rich medium, yielded few colonies upon direct transfer to hemi-autotrophic plates and none in medium. The only viable colonies obtained from selective plates still had to be cultured via LB medium and these took seven days to grow to a maximum OD_{600} of 1.1 to 1.2. No change in transformation efficiency, growth rate or maximum growth density was found in the two generations of sub-culturing that followed this hemi-autotrophic adaptation.

Long-term stocks of *E. coli* cultures are traditionally mixed with 20% glycerol and kept at -80°C. As glycerol can provide carbon to this strain instead of CO_2 it could temporarily relieve selective pressure on the plasmid during unavoidable this storage and as culturing from glycerol stock directly to hemi-autotrophic conditions has never been successful, alternatives were explored. A 5% (v/v) mixture of dimethyl sulfoxide (DMSO) and the CBB culture was tested instead, to grow directly from stock to hemi-autotrophic cultures. No significant decrease to optical densities in stationary phase or growth times were observed. These DMSO stocks were proven viable for at least nine months of storage, as exemplified in figure 3.4c.

3.4.1 Whole genome sequencing of the evolved strain

An LB culture grown directly from a cryo-stock of the batch-evolved CBB strain was sent for whole genome sequencing. The obtained contigs were aligned to the plasmid reference of *E. coli* MG1655 (NC_000913.3) with a mean coverage of 44.6 (standard deviation = 19.1) to cover all nucleotides and revealed no mutations. The reads mapped to the parental genome reference with a mean coverage of 4,621,066/4,641,144 bases, averaging at 66.2-fold with a standard deviation of 38.1. A total of 401 polymorphisms were found as attached in supplementary table C.2. 165 of these were associated with the modifications and mutations documented for CBB strain 4 [94]. These amounts are not absolute as deletions leaving an FRT site were flagged as several individual variants as the software attempted accurate alignment and all infrequently detected nucleotide deviations above the threshold were included. Screening sequences by annotation description and their associated regulatory region left 56 open reading frames (ORF)

containing at least one mutation over a 168 polymorphisms. Filtering these further for variant frequencies <80% left just 16 uniquely annotated regions containing a total of 55 unique polymorphisms in these regions as displayed in table 3.1.

The most notable deviation from its reference in the evolved strain genome is the *ychH* deletion as its remaining nucleotides align perfectly with the FRT sequence [126] also found in the documented knock-outs performed on the CBB4 strain. This is striking as the strain has been sequenced and documented prior to publication. The evolution experiment was performed on the original tube received for this work and has not been in contact with any component required for genetic recombination during this time, nor has *ychH* been considered as a target. This gene's function is reported as related to hydrogen peroxide and cadmium resistance [127] and has no currently published relation to the conditions this strain has been modified for.

Of the previously reported mutations in CBB derived strains, only *glpR* and *gppA* were seen in the dataset obtained here. The previously mutated *malT* and *crp* both had gained an additional mutation at different base-pairs causing a frameshift in the former, all four were found with ∼100% frequency. As a frequently occurring variant, the membraneous *ompC* gene was found with 15 mutations in the last 14% of its CDS, including the deletion of 2 codons over 3 indels. The natively IS5 transposon-interrupted *yhiS* gene had a total of 7 mutations (6 in its ORF) and local sequencing coverage increased to 940-fold. The found *menD* nucleotide insertion caused a frameshift in the final 72 base-pairs of the gene involved in menaquinone synthesis. The 2706 bp *malT* gene contained a frameshift after 1026 bp, likely causing a nonsensical translation for the remainder of the *mal* regulon transcription factor, associated with sugar catabolism. A substitution mutation caused a 30 codon premature truncation of *gppA*, coding for a protein involved in the stringent response. The *lrhA* gene, transcriptionally involved with cellular motility was found to contain a 3-codon deletion. The remaining 9 coding substitution mutations were in genes involved with vitamin B12 and ion transport, sugar, global energy, phosphate, fatty acid (acetyl-CoA) metabolism, or arginine biosynthesis (α-ketoglutarate dehydrogenase). The latter two transcriptionally regulate flux branch-points in the TCA cycle.

CDS annotation	Unique mutations	Type(s)	Gene description/regulation
ompC	15	Substitution, frameshift, truncation	outer membrane porin protein C
aroG	1	Substitution	3-deoxy-7-phosphoheptulonate synthase
ychH	20	Knock-out (FRT)	DUF2583 family putative inner membrane protein; stress-induced protein
ddpD	1	Substitution	ATP-binding component of ABD superfamily
menD	1	Frameshift	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic-acid synthase
lrhA	1	9 bp deletion	transcriptional repressor of flagellar, motility and chemotaxis genes
ppk1	1	Substitution	polyphosphate kinase
argP	1	Substitution	transcriptional regulator for arginine transport and DNA replication genes
crp	1	Substitution	cyclic AMP receptor protein
malT	1	Frameshift	*mal* regulon transcriptional activator
glpR	1	Substitution	DNA-binding transcriptional repressor
yhiS	7	Frameshift, substitution	pseudogene
pstC	1	Substitution	phosphate transporter subunit
gppA	1	Truncation	guanosine pentaphosphatase; exopolyphosphatase
fabR	1	Substitution	transcriptional repressor of fabA and fabB
btuB	1	Substitution	vitamin B12; cobalamin and ion outer membrane transporter

Table 3.1: Gene polymorphisms found in the batch-evolved hemi-autotrophic strain with a detection frequency over 80%. Mutations are ordered by genome location and grouped by CDS annotation, including its regulatory region, if applicable.

The less frequent (20-80%) gene variants found are shown in table 3.2. A greater diversity of gene functions is represented in this group, compounding to the likelihood at least some of these occurred in the non-restrictive

rich medium the sequenced culture was grown in. 9/39 coding enzymes found are involved in sugar or energy metabolism with only a single polymorphism each. Eleven different ribonucleic acid (RNA)-coding regions were found to contain multiple mutations each on average, where 3 pseudogenes were found to contain almost 10 polymorphisms each. In the remaining polymorphisms transposases, a toxin-antitoxin system, heme lyase (involved with respiration), DMSO reductase, proteins involved in amino acid and nucleotide metabolism and 4four CDS's of unknown function were found.

CDS annotation	Unique mutations	Type(s)	Gene description/regulation
rayT	3	Frameshift, start codon loss	REP-associated tyrosine transportase
lacA	1	Truncation	Galacaside O-acetyltransferase
	1	Truncation	Regulatory region of *lacZ*; beta-galactosidase
tRNA-Gln	5	Deletion	
tRNA-Met	1	Deletion	
tRNA-Leu	5	Deletion	
	1	Transversion	Regulatory region of *chiP*; chitoporin
ybfD	2	Frameshift	H repeat-associated putative transposase; putative DNA ligase
tRNA-Lys	5	Deletion	
tRNA-Val	2	Deletion	
RtT sRNA	6	Deletion	
dmsA	5	Frameshift, deletion	dimethyl sulfoxide reductase anaerobic, subunit A
puuD	1	Substitution	Gamma-glutamyl-gamma-aminobutyrate hydrolase
	7	Insertion, substitution	Regulatory region of IS5 family transposase and transactivator insH1

pseudogene *ydbA*	9	Frameshift, deletion, noncoding substitution	autotransporter domain-containing protein CDS
mlc	1	Truncation	Putative NAGC-like transcriptional regulator
ydjY	1	Substitution	Putative ferrodoxin-like lipoprotein
pseudogene *wbbL*	13	Substitution	Lipopolysaccharide biosynthesis protein, N-ter fragment
wzxC	5	Frameshift, deletion	putative colanic acid exporter
	1	Transversion	Riboswitch for *thiM*, hydoxyethylthiazole kinase
hypothetical gene *gtrS*	2	Substitution	serotype-specific glucosyl transferase
dsdC	2	Substitution	D-serine dehydratase (deaminase) transcriptional activator
tRNA-Arg	3	Deletion	
thyA	1	Substitution	Thymidylate synthase
ibsD	1	Start codon loss	Type I toxin-antitoxin system *Ibs* family toxin CDS
ibsE	3	Frameshift	Toxic membrane protein
	2	Substitution	Regulatory region of *ygjJ*; putative periplasmic protein
pseudogene *yhcE*	6	Substitution	Pseudogene
rrfF	1	Deletion	rRNA-5S ribosomal RNA of *rrnD* operon
tRNA-thr	1	Deletion	
rrfD	1	Transition	rRNA-5S ribosomal RNA of *rrnD* operon
gspF	1	Substitution	General secretory pathway component, cryptic
metE	1	Substitution	5-methyl-tetrahydropteroyl-triglutamate-homocysteine S-methyltransferase
rrfB	3	Substitution	rRNA-5S ribosomal RNA of *rrnB* operon

nrfE	5	Frameshift, deletion	Heme lyase (NrfEFG) for insertion of heme into c552, subunit *NrfE*
mscM	1	Truncation	Mechanosensitive channel protein; miniconductance; putative
yjfM	1	Noncoding substitution	DUF1190 family protein
yjjN	1	Noncoding substitution	L-galactonate oxidoreductase
ettA/yjjK	1	Frameshift	Putative transporter subunit of ABC superfamily: ATP-binding component

Table 3.2: Overview of *E. coli* BW25113 annotated regions found in the batch-evolved strain in a frequency of 20 to 80%, unique polymorphisms are grouped by their relevant gene name.

Considering the culture time, volume, dilution rate and generously estimated average culture densities of the four hemi-autotrophic strains evolved [94], results in a mutation rate of 1.1 - 5.2×10^{-8} mutations/bp/cell, with 1.79×10^{-8} mutations/bp/generation for the CBB4 strain. This is over 1000-fold above the random mutation rate of 8.9×10^{-11} in *E. coli* [128]. Performing the same calculation on the batch-evolved strain required replacement of the dilution rate with a base-2 logarithmical function to determine the amount doubling events in between inoculation and its next sub-culture, which was an average of 8.81 generations. Despite its low total volume, the batch-evolved strain had gone through an amount of generations comparable to the chemostat cultures with a calculated mutation rate of 3.76×10^{-8} mutations/bp/generation for the high-frequency mutations and 1.92×10^{-7} mutations/bp/generation when considering all new found polymorphisms in the sequenced culture. This could indicate that the hemi-autotrophic metabolism has not stabilized in the strain. Alternatively, it could signal the the mutations found here and those previously published [95] represent only a small portion of possible mutations relevant to this phenotype, or that there are significant inconsistencies between the used culturing conditions and they end up selecting for different genotypes.

3.5 Plasmid optimization

To relieve a portion of the metabolic burden the pCBB plasmid (visualized in fig. 3.5a) imposes and create space for further modifications, it was decided to exclude a large inactive region from the construct. This shortened pCBB plasmid (pCBBs), shown in figure 3.5b) was digested on the unique restriction digestion sites of NdeI and ApaLI, and re-ligated with a spacer consisting of two annealed primers of the appropriate sequences. This resulted in a deletion of 2531 base-pairs containing a large 5′ fragment of the β-lactamase (*BLA* gene and both halves of the transposase-interrupted *tetR* gene, whose expression isn't required for this project and would create more room on the plasmid for eventual inserts. Sequencing of pCBBs revealed a 343 bp deletion, starting from the NdeI restriction site to the end of the promoter of the CBB operon, just upstream of *cbbM*[1].

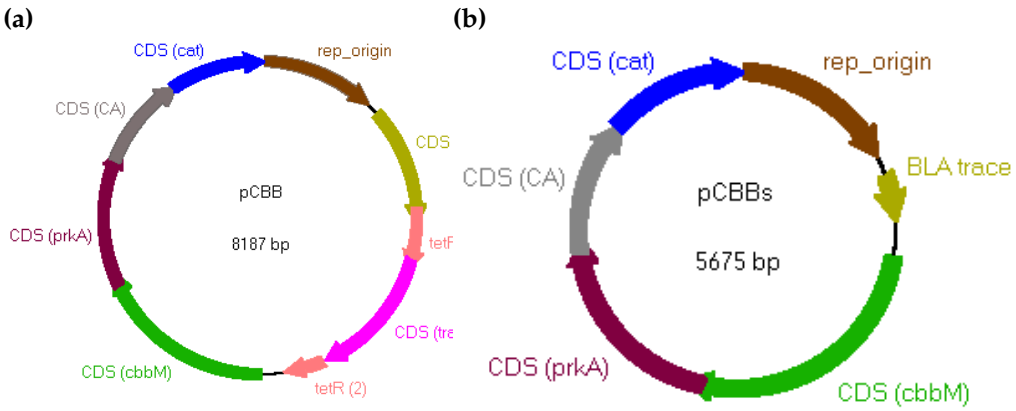

Figure 3.5: Schematic plasmid maps of **(a)** pCBB [94] and **(b)** the derived pCBBs, lacking the BLA, tetR and transposase regions of the former.

Early efforts of this work started with the correction of the deletion, by digestion of the FseI site (introduced on the spacer) and the ApaI site located in *cbbM*, to remove the 744 bp region. This was ligated to the backbone with the 1087 bp PCR amplicon from pCBB between the same restriction sites. Correct assembly of the construct was confirmed by digestion of XhoI and AvrII to produce bands of 1411 and 4265 bp as observable in figure 3.6. Local sequencing showed several clones still contained partial deletions of the promoter but sequencing of several more colonies found a plasmid with the

[1]Work presented in this section up until this point was conducted by Doctor Minyeong Yoo.

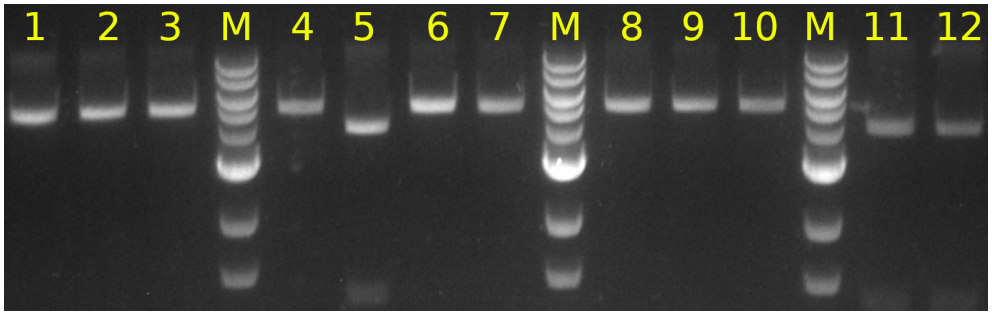

Figure 3.6: Agarose gel of 12 assembled pCBBs clones after re-introduction of pTet digested by XhoI and AvrII, expecting fragment sizes of 1411 and 4265 bp. The 1kb DNA ladder was used as marker (M) with fragments of 10.0, 8.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.5, 1.0 and 0.5 kbp.

correct insert. Whole plasmid sequencing showed no further mutations were introduced, though transformation of the plasmid into the CBB background strain never led to a successful hemi-autotrophic culture.

Slightly less than three years after the construction of the shortened plasmid and culturing proficiency of the CBB strain had improved, another attempt was made. As part of a series of constructs described in section 4.6, pCBBs was constructed as the intermediary to a control plasmid. A mutation that had accrued during the process was eventually corrected and the plasmid was transformed into the CBB background strain. Via an intermediary LB culture, the strain was successfully grown under hemi-autotrophic conditions to a maximum OD_{600} of 0.72. Upon a second sub-culture, under now-established growth conditions, the strain's highest recorded doubling time was 10.36 hours, as shown in figure 3.7. The background strain with the unmodified pCBB as control doubled significantly faster at 5.07 hours. Later re-transformations of the initially constructed pCBBs plasmid stock still resulted in the failure of hemi-autotrophic growth after an LB pre-culture where the second plasmid as a control succeeded.

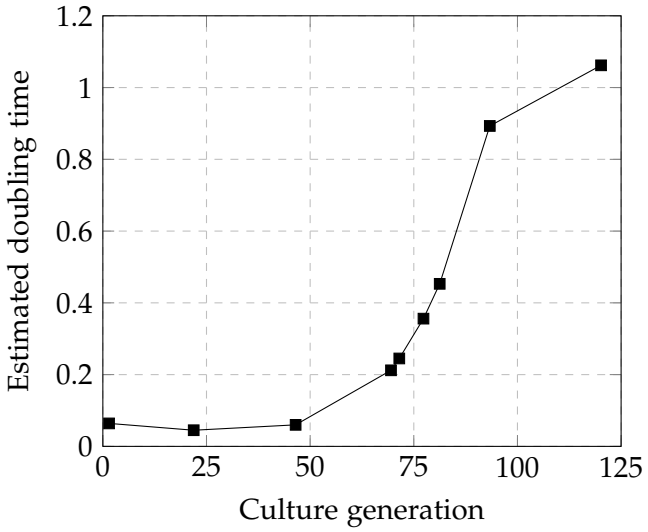

Figure 3.7: Growth-curve of the second pCBBs construct. Inoculated at $OD_{600} = 0.064$, from a hemi-autotrophic culture that grew to a maximum OD_{600} of 0.72, minimum observed doubling time of 10.35 hours.

3.6 Discussion

Few variations to the previously documented growth conditions of the hemi-autotrophic *E. coli* improved culturing characteristics. An increase in reliability was observed with the application of antibiotic selection (despite already restrictive growth conditions), ensuring availability of dissolved CO_2 in the medium and minimization of growth media switches. The ability to grow these cultures in 50 mL tubes in a traditional shaking incubator has greatly increased the throughput and reliability of tested cultures during the project. Transfer of hemi-autotrophic cultures to plates saw a low success rate but as transformants could be selected during the still required growth on rich substrates and could reliably be sub-cultured to M9 medium, the need for it was decreased. The capability to store DMSO culture stocks was a great development as it negates the need for re-adaptation to selective media of grown strains. Efficiently transforming or sub-culturing to these same conditions is still a much-needed development for the strain but was demonstrated reliable when performed and grown in an environment with sources of contamination vulnerable to traditional methods of sterilization.

The transformation efficiency of the CBB4 strain was not improved during the course of experiments. Increasing survivability of the transformed cells by variation of recovery time, media or immediate transfer to selective media (with selection-friendly supplementation) didn't result in increased adaptation rates under hemi-autotrophic conditions. The basic M9 medium formulation, apart from supplementation, has remained unchanged as it was already supportive of hemi-autotrophy but it is possible variations to it could improve growth characteristics and/or transformation efficiency. Curiously, the exact composition of M9 medium used during the development of the strain was not specified [94, 95] and could be a factor in the results obtained. Though, as the used medium supported growth of the strain and permissive growth conditions were confirmed as susceptible to mutations via correspondence with the developers of this strain, it is unlikely significant improvements would be gained by its variation.

Avoiding the use of non-selective culturing conditions during transformation of the hemi-autotrophic strain by electroporation of freshly grown, competent cultures didn't result in growth dependent on the replacement

plasmid. If the transformation method was viable, these experiments should have yielded colonies whose growth was supported by the newly introduced antibiotic resistance. The constructed pCBB variant with its chloramphenicol resistance exchanged for spectinomycin was validated by sequencing of its marker and via the established heat-shock transformation protocol to hemi-autotrophic medium. A reversed control of the electroporation experiment, replacing the $spec^R$ variant of pCBB with its original, should rule out this modification as its unlikely cause and (dis)prove the feasibility of this transformation method.

The severely reduced growth rate under pressure of the unique hemi-autotrophic culturing conditions is likely indicative of the CBB4 strain's lowr viability, with its replacement rate marginally higher than its mortality rate and / or a diminished overall metabolic efficiency. The additional cellular burden of a transformation, plus its accompanying change in metabolism, likely push this rate below a sustainable level. Further modifications of this strain therefore demand a transformation method that cause minimal disruption to its already precarious metabolism. Significant development in overall growth and/or metabolic efficiency might allow for the improved cell viability required for hemi-autotrophic transformations (and plate growth). Electroporation is more efficient than the otherwise already effective heat-shock but was also proven insufficient under these sensitive conditions; even more efficient techniques might prove worth experimenting with.

The successful continuous sub-culturing of the CBB strain without failure proves the hemi-autotrophic culturing methods used were reliably sufficient for the CBB4 strain. Despite the absence of optical density density data for the first seven cultures went unrecorded, as its data wasn't expected to be impactful, these are likely to have poorer characteristics considering their inoculation volumes (supplemental figure C.1). Only measurement dates were recorded, lacking specific times, meaning there is an unknown error rate to the estimated growth rates. The observed increase in growth rate and the window in which measurements could be taken limit this range and allowed for the observations that indicate strain adaptation over this period of decreasing growth rates and shrinking of inoculation volumes to previously infeasible quantities. Regardless, the re-transformation of the strain at generation 26 didn't show any subsequent qualitative improve-

ments. This could be due to the limited development of the strain, as it hadn't yet reached the performance that was observed a few generations after. The estimated growth rate and inoculation density at this point were significantly improved from their initial observations and these weren't replicated beyond this evolution experiment. Except for the decrease of inoculum volume, it is difficult to attribute the increasing growth rate exclusively to evolution. It is possible that the strain's metabolism adapted to the used culturing regiment as this has been shown as a significant factor in other experiments (illustrated by the different inocula of figure 3.2). Epigenetic changes in the strain could also explain these observations, and they were not considered in the analyses for this project. An increase of cell viability in stationary phase could have a similar phenotype to the desired improvements in carbon and energy metabolism. The 5-generation drop in growth rate per culture time following the culturing issues at generation 16, shown in figure 3.4b, supports this hypothesis or could signal a dominant community possessing an advantageous but sensitive mutation. Unfortunately, concluding project schedules didn't allow for a transformation test of the last-obtained sample of the evolved strain, at sub-culture 37 or further investigation into the observed developments.

The sequencing of the batch-evolved CBB strain revealed a FRT knock-out of *ychH* that wasn't mentioned in the strain's documentation and has no obvious relation to its distinctive phenotype. It is likely an artefact from previous modifications as it possesses the same left-over sequence, indicative of FRT recombination, as the documented knock-outs. A hitch-hiker region from another P1 transduction knock-out is unlikely as the gene is outside of the 100 kbp range of the rational knock-outs. Though recovery primers surrounding the gene are documented [95], they are otherwise unmentioned and are clearly not the primers used for this knock-out as the region closely matches its Keio primers [108]. Regardless, there is no current indication this modification is relevant to hemi-autotrophic growth. Why it is undocumented while every other mutation was carefully noted, is unknown. It could have been a minor community that gained dominance in the few cultures surrounding its transfer to the work documented here.

Of the remaining major genomic changes in the batch-evolved strain, many are directly relatable to the maintenance of the hemi-autotrophic

growth phenotype as they are involved in sugar, energy or phosphate metabolism, or transport. Additional mutations such as those found in *malT* and *crp* to further fine-tune the strain to its unnatural metabolism were expected. Mutations in *glpR* and *gppA* were found as previously published in CBB-derived strains [94, 95]. Similarly, the found *fabR* is likely to be functionally related to the published *fabH* mutation in fatty acid metabolism. The found *pstC* mutation could be analogous to the documented *pstS* in phosphate transportation/regulation. All these mutations are relatable to optimization of flux through pathways connecting to the TCA and CBB cycles. All these mutations are relatively similar to those previously found and no major unexpected mutations were discovered.

Several of the lower-frequency polymorphisms occur in metabolic pathways similar to the major mutations. The more mutable annotated regions of pseudogenes, transposases, tRNA's, rRNA's and a toxin-antitoxin system account for a large fraction of these minor mutations. Their effects are difficult to estimate in this context but could be attributable to genetic drift; the varying of genetic polymorphisms due to random chance. Heme lyase, anaerobic DMSO reductase and the four uncharacterised genes would need further investigation to determine their relation to hemi-autotrophic growth conditions. Given the increased mutation rate when including these low-frequency mutations, it is also possible these are the result of the non-restrictive LB culture performed to harvest cells for genome sequencing. Considering its increased culture time, this was difficult to practically avoid.

Finding only two genes of the collective 105 mutated genes published [94, 95] in CBB-derived cultures, was unexpected. With a minimal genetic background to support the hemi-autotrophic phenotype in the starting strain, a significant amount of mutations was predicted. But such a small overlap with the published fine-tuning mutations from this strain was unexpected. This divergence could be due to the translation of experimental methods, though a previous batch-cultured analysis [95] showed little overlap as well [94]. Due to the sequencing of only a single evolved culture (despite the increased variability of a culture instead of a purified colony), it is impossible to derive any certainty from these differences. The discovered high frequency mutations rule out the LB culture the deoxyribonucleic acid (DNA) was purified from as their origin. Those of lower occurrence potentially are,

the sequencing of multiple purified colonies would decrease this likelihood. The lack of polymorphisms in the plasmid went against expectations as the CBB operon and its toxic gene(s) were assumed an easier target for expression modulation than several accommodating background mutations. Although the plasmid didn't accrue any fine-tuning mutations during chemostat evolution either, where the naturally evolving genomes gained between 22 and 34 [94].

Though the ancestral CBB strain was successfully adapted to the hemi-autotrophic conditions, the process was halted as soon as CO_2-dependent growth by the strain was detected. It is therefore unlikely to have reached its optimal adaptation to these conditions, especially considering strain 4 (used for this work) was modified for minimal phenotypical replication only gained 6 mutations during its evolution where the preceding strains gained up to 33. The mutation rates of the batch-evolved strain falling within the maximum 5-fold range difference found between the 4 chemostat strains, they are remarkably similar considering they are at least 100-fold higher than the highest reported background mutation rate [129] This suggests there is still a great selective pressure acting on the strain and that there are likely to be more mutations or modifications able to further improve on the hemi-autotrophic phenotype. The $\sim$20-fold higher mutation rate calculated for all detected polymorphisms includes those originating from the non-restrictive heterotrophic culture grown before harvesting of the strain for genome sequencing. These could be excluded from the population by performing several whole genome sequencing on several single colonies and comparing mutation frequencies. Regardless, it is confirmed the hemi-autotrophic strain still mutates rapidly to either overcome or adapt to its synthetic metabolic burden and as demonstrated before [94]. This increased mutation rate is likely caused by the toxicity of the introduced *prkA* [27] and / or the modifications made to the central carbon metabolism. It is also indicative that significant improvements to the growth characteristics of strain CBB4 can eventually be developed or evolved.

Decreasing the size of the hemi-autotrophic growth enabling CBB plasmid would allow for increased metabolic efficiency as it reduces the amount of DNA that requires maintenance [130] and is reported to increase transformation efficiency [131]. It would also create more space for the further

incorporation of genes onto the 8.2 kbp plasmid. The deletion of the promoter upstream of *cbbM* that resulted from the initial modification could be indicative of transcriptional disruption by removal of toxic plasmid regions, when grown in rich media. The mixed results in growing the two separately constructed (and fully sequenced) pCBBs constructs under hemi-autotrophic conditions is explainable by an impure stock of the former, though this went unnoticed on gel analyses and purifications, or could hold a connection to the promoter that was restored in this strain. Regardless, the latter construct eventually proved to support hemi-autotrophic growth albeit at approximately half the rate of the full-sized pCBB, making it less desirable for further cloning work. This is an indication that this deleted region likely contains elements that interact with CBB synthetic operon. The transposon-interrupted *tetR* present here still transcribes the 381 base-pairs before concatenation. The HTH-type DNA binding domain of this protein is within the first 63 amino acids (UniProt AC P03038) of the full *tetR*-encoded protein and likely still interacts with the tetO region present in pTet. What exact effect this potential residual binding could have is presently unreported and beyond the scope of this project.

The experimental objectives for this chapter were partially achieved. The hemi-autotrophic CBB4 strain of *E. coli* was found to grow with high reliability in 5 mL culture volume when sodium bicarbonate was added, relieving the necessity of culturing larger volumes in CO_2 incubators. This reduction in culture volume significantly increased experimental throughput capacity CBB4 variants. No tested method of hemi-autotrophic transformation proved successful. This is likely related to the low growth rate of the strain and the effects of modifying its trophic mode has had on the strain's viability. DMSO was proven as a viable cyroprotectant that doesn't require an intermediary heterotrophic culture to continue selective growth. A batch evolution experiment conducted over 37 generations saw modest improvements to the strains specific growth rate, carrying capacity and required inoculum size. These characteristics didn't persist upon re-transformation with the same plasmid, though due time restrictions this experiment was not performed on the last generation. Full genome sequencing revealed new but predictable mutations at a rate well above the background mutation rate.

To adapt this strain further to hemi-autotrophic conditions, the other

three evolved strains [94] should strongly be considered as only the fourth strain was experimented on in this research. As an artificial minimal replication of the common mutations in the preceding strains introduced to replicate hemi-autotrophic growth. The other background mutations weren't essential to hemi-autotrophic growth but are likely to provide the metabolic fine-tuning required for the increased growth densities and rates reported in CBB strains 1 and 3. These growth characteristics are still beyond the mild improvements achieved in the batch-evolved strain 4. Elucidating the most relevant of these mutations to hemi-autotrophic growth and implementing them might be more time- and cost effective than optimizing culturing conditions for the strain. With an increased viability of the strain, under these already stringent conditions, further modifications to the already fragile hemi-autotrophic metabolism could prove beneficial to both its culturing and transformation efficiency.

3.7 Acknowledgments

The efforts of Dr. Minyeong Yoo in establishing early hemi-autotrophic culture methods, performing and characterizing the initial shortening of the CBB plasmid, and providing continued support were invaluable to this project. Professor Philippe Soucaille laid the ground-work for this project, contacted the researchers involved in developing the hemi-autotrophic *E. coli* and obtained CBB strain 4 for this work from them. Niv Antonovsky subsequently confirmed the strain's mutability when removed from hemi-autotrophic pressure in an insightful correspondence.

Chapter 4

Development of a synthetic CBB cycle in hemi-autotrophic E. coli

"Science isn't about why, it's about why not."

Portal 2
Valve

4.1 Introduction

Following the establishment of culturing techniques for the hemi-autotrophic strain of *E. coli*, its comparative use to evaluate the performance of its CBB cycle is possible. As an isozyme to both FBPase and PRK, the replacement with pyrophosphate-utilizing phosphofructokinase (PFP) is theoretically feasible. As PFP doesn't have the same ATP requirement as phosphoribulokinase, it might even outperform the canonical pathway. As the native *E. coli* metabolism might not support the isozyme's unique requirement for pyrophosphatase HPP, predicted to catalyze its formation and originating from the same operon as PFP, is also introduced. An improved energy and/or carbon metabolism of the strain is expected to improve on the relatively poor growth characteristics of the hemi-autotrophic strain.

The objective of experiments described in this chapter is to develop the strain and constructs to functionally include PFP into the hemi-autotrophic CBB cycle, and comparatively evaluate the performance of these strains. Cultures expressing these synthetic CBB cycle modifications are evaluated by their ability to grow under the same CO_2-dependent conditions as their parental CBB strain. Their growth characteristics will be compared to the derivative strains to determine any qualitative or quantitative changes.

4.2 Fructose-1,6-bisphosphatase knock-out

To reduce interference for the PFP reactions and confirm its dual functionality, the *fbp* gene located on the *E. coli* genome required inactivation. As a non-essential gene [132], the most effective method to do so was deemed deletion by P1 transduction using the Keio collection. The genome of the

hemi-autotrophic strain still contained the kanamycin resistance cassette as a result of the removal of *ppsR* during the development of the strain [94]. This kan^R requires removal as successful P1 transduction knock-outs performed with the Keio collection are selected by their resistance to kanamycin, imparted by the exchange with this marker gene. As the cassette is flanked on both sides by flippase recognition target (FRT) sites, it was removed from the wild-type (ppsR) site by expression of flippase (FLP) on temperature-dependently expressed pCP20 plasmid (with ampicillin and chloramphenicol resistance) and consecutive selection for antibiotic sensitivity. Figure 4.1 shows the successful removal of the kanamycin cassette from the *ppsR* site.

Several P1 transductions [107] were needed to obtain clones containing a pure, unambiguous knock-out of *fbp*. As shown in figure 4.2, PCR's performed on three of these colonies, using external primers and their parental strain as control, resulted in the approximate amplicon size of 2220 bp, expected for a successful replacement of the *fbp* site (1870 bp) with the kan^R cassette. Implementation in the correct strain was confirmed by performing similarly designed PCR amplifications on the other reported knock-outs.

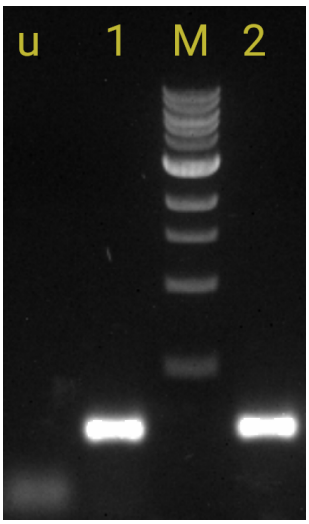

Figure 4.1: Agarose gel of the PCR amplicons of two colonies selected by their sensitivity to both kanamycin and ampicillin following the loss of both the chromosomal resistance cassette (of 992 bp) and pCP20 plasmid following flippase recombination. Primers external to *ppsR* were used for the numbered reactions, expecting a 129 bp amplicon, and an unrelated (u) failed reaction showing only primer dimers. Colony 2 was used further in the knock-out of *fbp*. The 1kb DNA ladder was used as marker (M) with fragments of 10.0, 8.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.5, 1.0 and 0.5 kbp.

The obtained Δ*fbp* strain transformed with pCBB failed hemi-autotrophic growth where the parental fbp^+ control strain grew. This demonstrates that the FBPase isozymes encoded by *yggF* and *glpX* could not functionally compensate for the loss of *fbp* and would not significantly interfere with the FBPase reaction catalyzed by PFP.

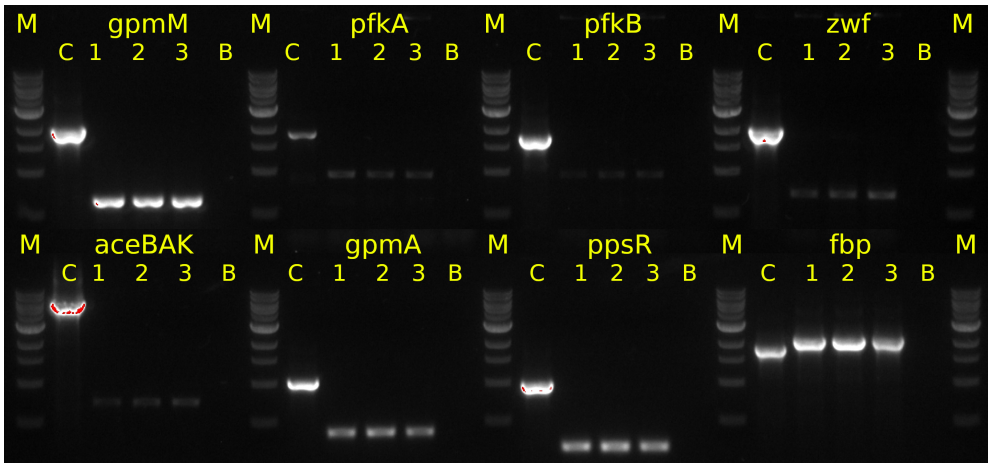

Figure 4.2: The PCR confirmation of all knock-outs of in the CBBΔ*fbp* strain. *gpmM* KO: 539, WT: 2061; *pfkA* KO: 872, WT: 1821; *pfkB* KO: 1770, WT: 861; *zwf* KO: 614, WT: 2067; *aceBAK* KO: 593, WT: 5425; *gpmA* KO: 290, WT: 1042; *ppsR* KO: 183, WT: 992 and *fbp* KO: 2220 (kan^R), WT: 1870.; M: 1kb ladder with fragments of 10.0, 8.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.5, 1.0 and 0.5 kbp.

4.3 Heterologous operon integration

The integration of the *pfp-hpp* operon originating from *M. capsulatus* bath[1] into both the unedited pCBB and the shortened pCBBs were performed by Gibson assembly. The operon's two coding sequences and each ribosome binding site (RBS) were amplified from a TOPO plasmid[2] by PCR using primers "CBB reco F" and "CBB reco R", resulting in an amplicon of a predicted 3447 bp, including homologous overhangs. The plasmid backbones were digested by the single cutter restriction enzymes PsiI and BlpI, excising a 760 bp region comprising the first 739 bp of *prkA* and its RBS. Transformation of the assembly mixes into competent cells only yielded viable colonies for those performed using the pCBB backbone. Integration of the operon into the shortened plasmid (pCBBs, as described in section 3.5) failed to grow under these same culturing conditions, even when re-performed at

[1]The *M. capsulatus* bath strain was generously provided by Professor Philippe Soucaille
[2]Constructed by Doctor Minyeong Yoo before the presented work commenced.

room-temperature. Correct integration in pCBB was screened by plasmid purification and subsequent restriction digestion. The resulting clones that displayed the expected restriction profile were sent for sequencing. After finding identical mutations in the operon it was discovered they originated from several TOPO clones. A TOPO clone with the unmutated operon sequence was found and used for subsequent cloning. Its integration using the same strategy was confirmed by the sequencing of multiple clones but all contained mutations varying from single point mutations to deletions approaching an entire ORF in the operon (as exemplified in fig. 4.3), most of them in the *hpp* gene. These clones were obtained before the start of this project and the origins of these mutations can only be speculated upon but it is also possible this heterologous operon is toxic in *E. coli*. The development of multiple independent mutations in the *pfp-hpp* operon upon its introduction into pCBB in a cloning strain and consequent growth in rich media was attributed to toxicity [27] resultant of its unregulated expression or non-restrictive growth conditions.

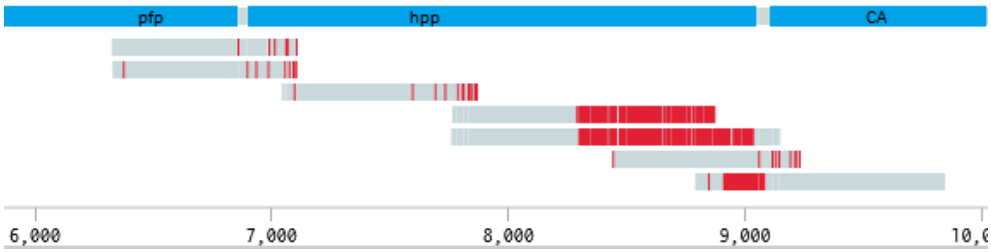

Figure 4.3: Alignment of seven different pSCBB clones to its designed plasmid map, displaying a variety of mutants.

4.4 Temperature-sensitive operon

The *pfp-hpp* operon was expressed under the hemi-autotrophic conditions that support *prkA*-dependent growth while the catalytical properties of the encoding enzyme differ significantly from it [102, 133]. These two enzymes originate from a methylotrophic organism and its native expression is unlikely to translate to a correctly functioning level in *E. coli*. To achieve the project's objective it was decided to dynamically modulate transcription of the heterologous operon by use of a thermo-sensitive promoter, controlling the operon's activity as a function of the temperature its harbouring culture is grown under. This mechanism was preferred due to the strain's reduced growth rate; the strain would otherwise require continuous addition of in-

ducer or the additional knock-out of its metabolizing enzyme. The *cI857* repressor associates to the *pR* promoter to block transcription rate depending on temperature. In *E. coli*, expression is reduced to negligible amounts at 30°C, expression is moderate at the standard culturing temperature of 37°C and expression is high from 39°C [125]. The pR promoter with its constitutively expressed temperature-sensitive repressor *cI857*, downstream on the opposite strand is documented to reduce expression to negligible amounts at 30°C, with moderate expression at 37°C and high expression beyond 39°C [125].

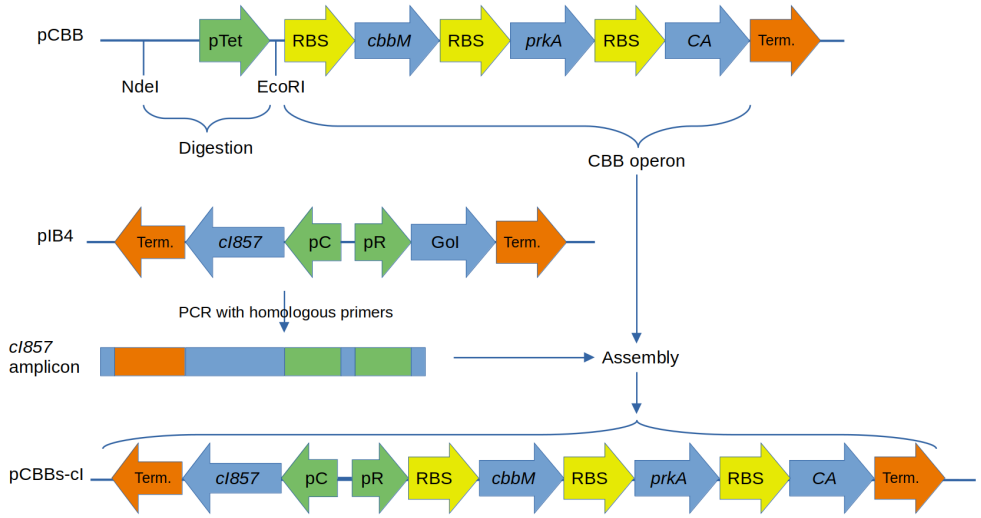

Figure 4.4: Schematic overview of the procedure to exchange the pTet promoter, regulating the CBB operon on pCBB as documented [94], with the *cI857* temperature-sensitive promoter-repressor system. The pCBB is digested and the *cI857* system amplified by PCR using primers sharing homology to the pCBB on the region flanking the excised sequence. Finally, the two fragments are assembled by homologous recombination. Red genetic elements signify a terminator (Term.); blue, genes (of interest); green, a (constitutive) promoter; yellow, an RBS; and line, an arbitrary nucleotide sequence. Displayed genetic element sizes don't correspond to their nucleotide lengths.

The sequence of the *cI857* expression system was amplified by PCR from the *pIB4*[3] plasmid using primers "cI857 hifi F and R", as illustrated in figure 4.4. The resultant 965 bp amplicon was used in a Gibson assembly containing the 5378 bp backbone of the fully sequenced pCBBs vector digested by EcoRI and NdeI. The assembly mix was transformed into DH5α competent cells and plated at 30°C. The resulting colonies were screened by PCR using the "pCBB-AN-seq and *cI857*-int-R" primers and the expected 772 bp am-

[3]pIB4 was generously provided by Professor Philippe Soucaille

plicons were present, as shown in figure 4.5. These clones were sequenced around the insertion site and integration of the anticipated sequence was confirmed.

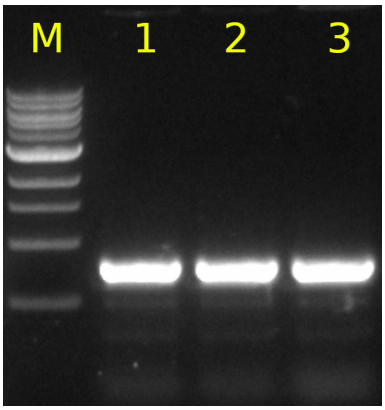

Figure 4.5: Agarose gel of three PCR's performed on colonies obtained from the assembly of pCBBs-cI with primers "pCBB AN seq" and "Ci857 int r2", expecting a 772 bp amplicon. The 1kb DNA ladder was used as marker (M) with fragments of 10.0, 8.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.5, 1.0 and 0.5 kbp.

The *pfp-hpp* operon was inserted into the newly constructed pSCBB-cI plasmid by use of a Gibson assembly. As the new promoter introduced a second PsiI site, the thermo-sensitive pCBBs intermediate was digested by BglII and BlpI instead, leaving a 5439 bp backbone. An additional 120 bp fragment was amplified using primers ”HF3F and HF3R2”, to restore the digested part of the *cbbM* gene. Its reverse primer was designed with a 5′ region complementary to the inserted amplicon initially included on the temperature insensitive plasmid. These fragments were assembled in a Gibson reaction, transformed and plated at 30°C to keep the operon inactive until necessary and prevent potential toxicity. Three of the screened colonies displayed the bands corresponding tot he predicted 4036 bp PCR amplicon using external primers “pCBB-AN2868 and ”CA 5pr Rev” (as shown in fig. 4.6a) and the correct restriction digestion profile (fig. 4.6b). One clone was determined to be correctly integrated by sequencing but failed to grow under hemi-autotrophic conditions using established sub-culture methods, where pCBB controls grew successfully.

Upon full sequencing of the plasmid, a 664delA mutation in *cbbM* was found to be responsible for the inability to hemi-autotrophically grow the

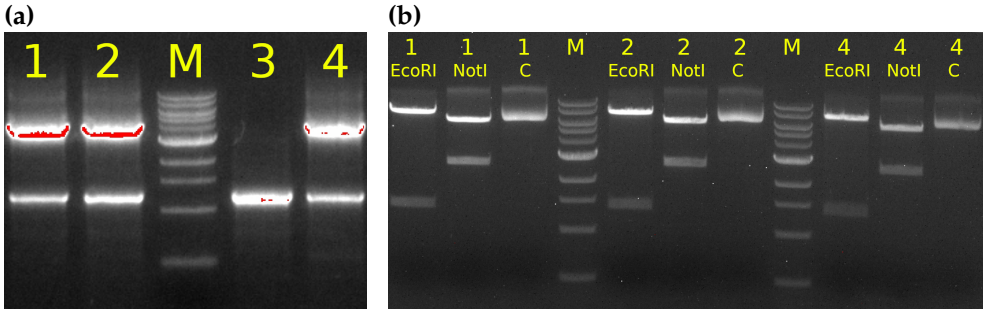

Figure 4.6: Agarose gels displaying the analysis of pSCBBs-cI performed by **(a)** PCR using primers "pCBB AN 2868BP" and "CA 5pr Rev", expecting a 4036 bp amplicon. The 1.2 kbp band is not associated with any pCBB-derived constructs during this work and is likely the result of unspecific amplification. Analysis of pSCBBs by **(b)** restriction digestion of the plasmids isolated of positive colonies, using EcoRI (expecting 1.4 and 7.3 kbp fragments) and NotI (expecting 2.6 and 6.1 kbp fragments) and its negative controls. Oversaturated signal is displayed in red and the 1kb DNA ladder was used as marker (M) with fragments of 10.0, 8.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.5, 1.0 and 0.5 kbp.

strain while harbouring the plasmid. Site-directed mutagenesis was performed to correct this nucleotide deletion. The plasmid was amplified using mirrored primers with the mutation site at their centre, amplifying the entire plasmid, digesting the template by DpnI, performing gel purification and transforming it into competent cells to ligate the linear plasmid *in vivo*. Upon re-sequencing of the successfully cultured colonies, three clones containing the corrected plasmid sequence were found. Culturing the both Δ*fbp* and CBB4 transformants expressing this plasmid by established methods did not support hemi-autotrophic growth where the re-transformed pCBB control did. Over-expression of the operon by culturing at 39°C didn't result in any growth either, signifying that the achieved expression level of the synthetic CBB operon is not supportive during hemi-autotrophic growth.

4.5 RBS modulation

The failure of hemi-autotrophic growth by the CBB strain with its *pfp-hpp* operon replacing the established *prkA* on pCBB was attributed to the incorrect regulation of individual genes on the operon, as its temperature-modulated expression modulates the whole (starting with *cbbM*). As the PFP-catalysed reaction replacing PRK activity consumes PPi, the FBP reaction produces it but at a 2/3 stoichiometric rate of the former. Assuming sufficient translation, the control construct expressing the synthetic operon in the fbp^+ strain demonstrates a shortage of PPi available for the PRK reaction

of *pfp*. PPi is typically treated as a waste product by the *E. coli* metabolism and efficiently metabolised to two orthophosphate molecules by inorganic pyrophosphatase (PPA) [118, 134]. The insufficient expression of *hpp*, as it is predicted to counter-actively produce PPi, is the likely origin of this shortage. Careful regulation of *pfp* and *hpp* expression will therefore be critical to the success of their heterologous integration in the CBB cycle.

In order to modulate the PPi-supply available to PFP, the expression level of the potentially toxic HPP was modulated. To progress towards the project objective the C, D and E RBSs [135] were decided on for testing an expression range for hemi-autotrophic growth of the (synthetic) construct. As the thermo-sensitive control of the original construct hadn't successfully grown hemi-autotrophically either, and therefore *cI857* hadn't been functionally validated yet, the C and D RBS variants were also cloned (the E RBS variant is originally present) just upstream of *prkA* on the pCBB plasmid. Plasmids expressing *pfp*, without the potentially toxic *hpp*, and *prkA* modulated by the three RBS variants were also constructed to demonstrate the phosphoribulokinase activity of PFP, independently of its fructose-1,6-bisphosphatase activity, and explore the relationship between the promoter and RBS. A schematic overview of these synthetic CBB and control constructs is presented in figure 4.7.

Construction of these strains was performed by modification of the sequenced construct that temperature-dependently expresses the *pfp-hpp* operon constructed in section 4.4: pSCBB-cI. The unique restriction sites BsrGI and BlpI were used to excise the RBS and downstream genes. The restoring coding sequences of *pfp*, *hpp* and a homologous region were PCR amplified using the parental plasmid as template, and primers encoding for the new RBS were annealed. These four fragments were assembled in Gibson reactions for each variant. The same strategy using BamHI instead of BsrGI was employed for the exchange of *hpp* with *prkA*, and using BglII in its stead on the intermediary pCBBs_cI−E_prkA construct allowed for the similar exchange of the RBS just upstream of *prkA*. Clones were screened by PCR using internal and external primers as shown in figure 4.8. The performance of the colony PCR potentially distorted the shown sample bands, obscuring the real size of the amplicon. Regardless, correct assembly of the entire CBB operon was confirmed by Sanger sequencing and eventually the

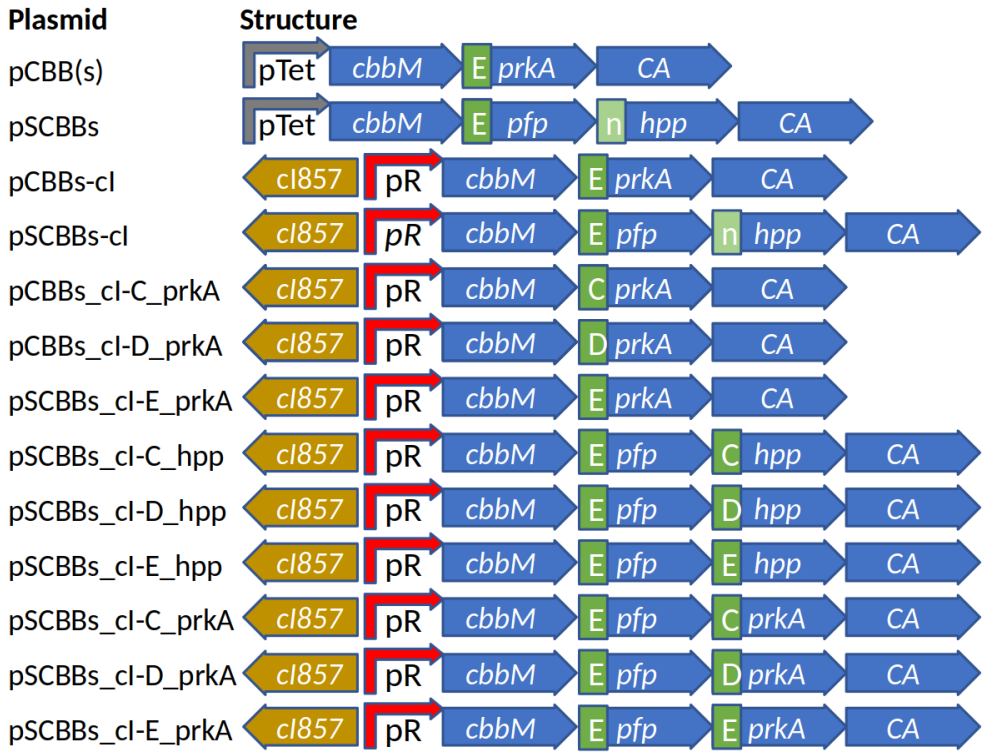

Figure 4.7: Overview of the RBS variant plasmid constructs. Where relevant, RBS type is marked according to Zelcbuch et al. [135], except for 'n' which signifies the native RBS of the downstream gene, unmarked regions remain unedited from pCBB. Only regulatory elements varied in the expression of metabolically relevant genes are displayed. Although the shortened region is not visualized here, the names of successfully assembled constructs are shown. Displayed genetic elements sizes don't represent their nucleotide lengths.

whole plasmid sequence was confirmed for all of these constructs.

The RBS variant constructs were transformed into the CBB4 and Δ*fbp* background strain. Initial cultures were grown in LB at 30°C and upon harvest washed with ultra-pure water before inoculation at $OD_{600} \approx 0.1$ in hemi-autotrophic media incubating at 37°C. Though the pre-cultures grew under plasmid-dependent antibiotic pressure and pCBB re-transformed controls were positive, thorough experimentation with these constructs did not achieve any hemi-autotrophic growth. Replication of these transformations, directly inoculated to hemi-autotrophic plates after recovery were not successful and controls saw an increased rate of contaminations compared to the media cultures. These false positives were eventually identified as contaminations by either PCR or successful growth in the same medium, lack-

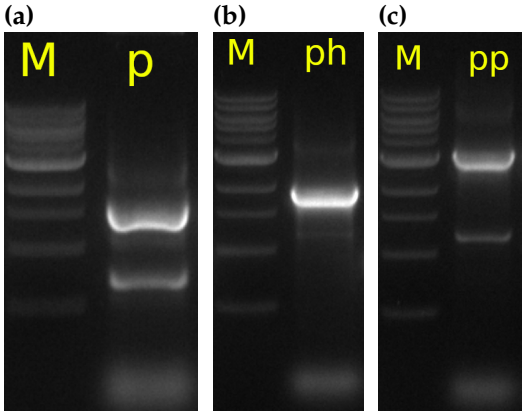

Figure 4.8: Representative examples of PCR analyses performed on RBS variant clones of **(a)** pCBBs_cI-*prkA (p), with a predicted amplicon size of 1607 bp; **(b)** pSCBBs_cI-*syn (ph), with a predicted amplicon size of 3126 bp; **(c)** pSCBBs_cI-*prkA (pp), with a predicted amplicon size of 2019 bp. The 1kb DNA ladder was used as marker (M) with fragments of 10.0, 8.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.5, 1.0 and 0.5 kbp. * denotes an RBS site containing the C, D or E variant.

ing the elevated CO_2 pressure. 39°C was deemed as approaching the strains maximum growth temperature as hemi-autotrophic conditions already impart a notable decrease in cell viability. These constructs were also sub-cultured in the (Δ*fbp*) CBB strain at 34, 36 and 39°C for expression rates of approximately ¼, ½, and 3-fold, respectively, of that at the standard 37°C [125] but none were found to grow on either hemi-autotrophic plates or media using the established methods where re-transformed pCBB controls succeeded.

4.6 Dual operon constructs

Isolation of the CBB operon from any modifications to the plasmid would allow for regulatory independence of further inserted genes while simultaneously functioning as controls for the established pathway. A total of nine sequential constructs were designed, inserting genes for independent expression by utilizing the easily accessible and unique MluI restriction site, located just downstream of the terminator following *cat* on pCBB as shown in figure 4.9. The *cI857* expression system was PCR amplified from pCBBs_cI-D_prkA (with a completely validated sequence) and amplified genes of interest using primers containing homologous regions, and integrated into the pCBB backbone by Gibson assembly. Constructs for the expression of both the *pfp-hpp* operon and *fbp*, in several positions, were designed together with several (intermediary) controls. Variants with deletions of the *prkA* gene and the region from *BLA* to the second tetR fragment (pCBBs) as in section 3.5, both using the same method previously employed, were also included.

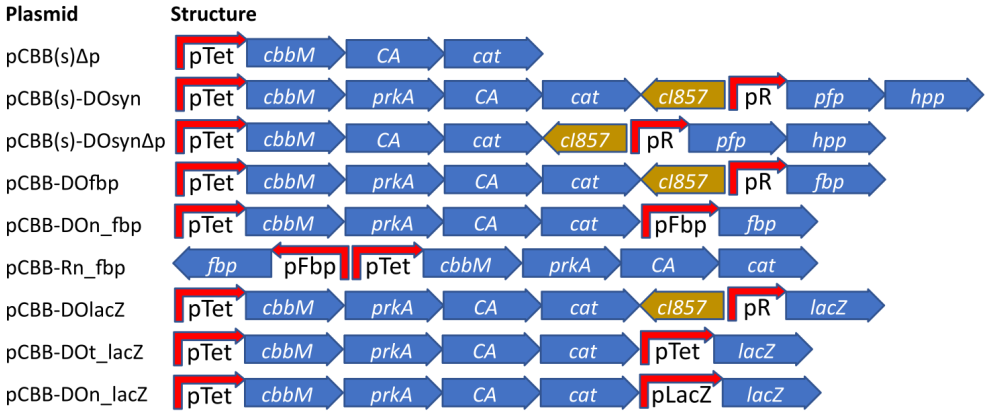

Figure 4.9: Schematic representation of the 12 dual operon (DO) and intermediate constructs. Only regulatory elements variable in the expression of metabolically relevant genes are displayed so the variants with the deletion of *BLA*, transposase and tetR are omitted. The *lacZ* constructs were assembled separately. Displayed genetic elements sizes don't represent their nucleotide lengths.

Construction of the first seven plasmids was performed in parallel, starting with sequence deletions and integrations, followed with subsequent modifications where appropriate, as correctly modified clones were identified by sequencing. Upon confirmation of successful integration, validated by sequencing of the second generation of constructs, whole plasmid sequencing was performed on the first generation of plasmids. These revealed a 312_313insA frameshift mutation in *prkA*, changing a sextet adenosine sequence to septet, in all constructs. Its occurrence in all clones indicates the fully sequenced stock of its parental strain mutated and became the dominant genotype within the single over-night culture used for these constructs. The same glycerol stock of the parental culture has been used for previous experiments and cultures, and this mutation wasn't encountered before or since. The Δ*prkA* constructs saw a drop in sequencing quality and as they had a lower chance of success than most of the controls, fixing the mutations was prioritised.

As two separate attempts at restriction-ligation yielded no clones with the expected amplicon size, site-directed mutagenesis was utilized for the correction of this mutation on the nine constructs containing the *prkA* insert. First, primers were designed to contain the corrected nucleotide sequence in the centre of both primers covering the same area and amplify the remaining plasmid. The resultant PCR mixes were gel-purified, digested by

DpnI to remove methylated template strands and transformed into competent cells. Five corrected colonies of the total 50 were found to have repeat inserts of 69 to 87 bp where the remainder retained the single nucleotide insert. Changing the primer design to amplify the plasmid but omitting the single nucleotide insert and repeating the procedure while applying the KLD enzyme mix and buffer to the PCR mix before transforming, according to its accompanying protocol, showed no success upon sequencing of 10 colonies of each construct.

While retaining the approach of leaving the inserted nucleotide out of the amplified sequence, the original sequence "GAAAAAATC" was changed using synonymous codons to "GACAAGATC" on the forward primer. 28 out of the 40 sequencing samples covering this region contained the correction, though the mistake of reusing the reverse primer caused the insertion of an extra guanosine just preceding corrected mutation. Correcting this error and repeating the process resulted in 23 of 25 sequencing results covering the mutation site, 4 of these were successfully corrected but 3 had nearby insertions. The remaining pCBBs construct was confirmed by restriction digestion (see figure 4.10a) and was found to have no mutations upon full sequencing. One to three more attempts per construct, sequencing five colonies for each construct saw several more corrections but they all contained inserts from the bordering sequence, the two attempts what were cultured at a maximum of 30°C even saw a reduction in correction frequency.

The two dual operon constructs expressing *fbp* under its native promoter and the three expressing *lacZ* were constructed from a separate pCBB culture. Though few (with a maximum 15) colonies were found following the assembly of each of these constructs, restriction digestions showed approximately correct restriction profiles (examplified in fig. 4.10b & 4.10c) and sequencing showed no mutations in either operon. For the *lacZ* construct controlled by pTet, 14 colonies were found of which 4 exhibited the expected 4.1 kbp PCR amplicon using "CmintF + Ori-R" primers as shown in figure 4.10d. Though the downstream integration sites were quickly verified by sequencing, the promoter region of all remaining clones was consistently of low quality and often ended before its sequence region. Using two of the primers used in the assembly of the construct to PCR amplify the region

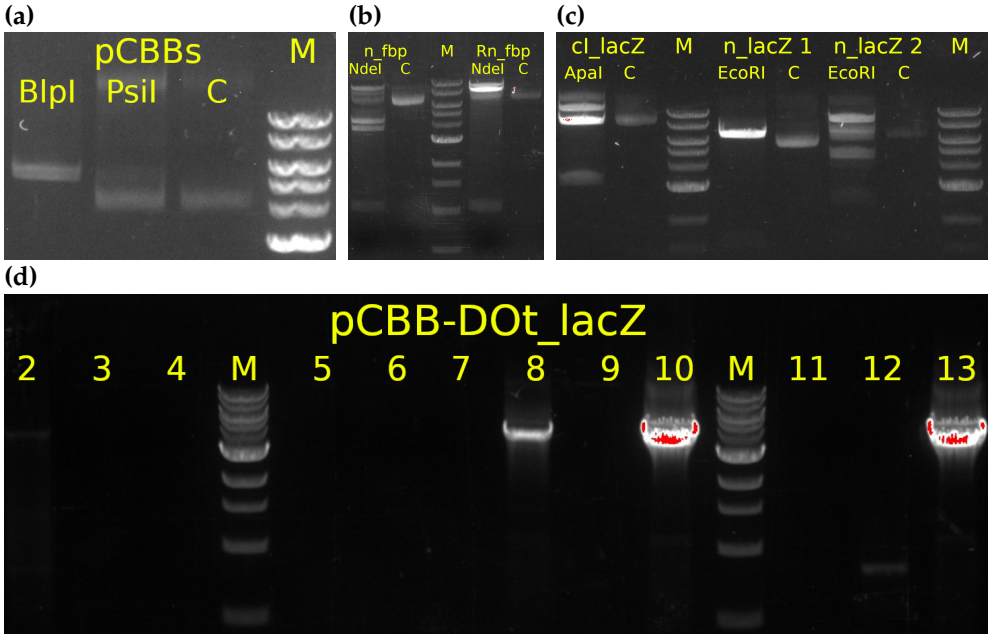

Figure 4.10: Agarose gels showing the analysis colonies expressing pCBB-derived constructs employing dual operons for genes of interest. **(a)** pCBBs digested by BlpI and PsiI (failed), expecting a band size of 5675 bp. **(b)** pCBB-DOn_fbp and pCBB-Rn_fbp digested by NdeI with expected band sizes of 1.1, 4.4 and 3.8 kbp, and 1.1 and 8.3 kbp, resp. **(c)** pCBB-DO_lacZ digested by ApaI to expect 3.4 and 8.8 kbp bands and pCBB-DOn_lacZ digested by EcoRI expecting 1.6, 2.6 and 7.2 kbp bands. **(d)** A PCR performed on pCBB-DOt_lacZ colonies, expecting a 4055 bp amplicon. The 1kb DNA ladder was used as marker (M) with fragments of 10.0, 8.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.5, 1.0 and 0.5 kbp.

eventually allowed for confirmation of the intact promoter present in clone 10 by sequencing. Retrospective analysis confirmed the correct integration and sequence of this construct and, except of the deletion of the tetO region of pTet, in clone 2 and is further discussed in section 4.8.

4.7 FBP complementation

The successful removal of *fbp* from the hemi-autotrophic strain's genome is critical to successful PFP regulation as the enzyme is expected to require a certain cytoplasmic concentration of the pyrophosphate critical to supporting both its reactions in the synthetic CBB cycle. The knock-out was confirmed by PCR, as shown in figure 4.2. Its functional confirmation by a plasmid-expressed *fbp* is of importance as the multifunctional PFP would not rule out dysregulation of its sensitive catalytic equilibrium. Using the single PsiI site on pCBB, *fbp* with its RBS and terminator were PCR-amplified

from the chromosome of the CBB4 strain and inserted into pCBB via Gibson assembly. Successful clones were screened by PCR and confirmed by sequencing of the entire synthetic operon. Throughout a multitude of transformations, no hemi-autotrophic growth was achieved by the complementing construct. Transformations into the *fbp*$^+$ parental strain and sub-culturing to hemi-autotrophic media had no success either, where re-transformed pCBB controls did.

In accordance with the project's objective, three alternative constructs for the verification of the *fbp* knock-out by complementation were included in the series of dual operon constructs as described in section 4.6. All constructs aimed to express *fbp* independently from the CBB operon and incorporate it in two different plasmid regions, regulated it by its native promoter or the temperature-sensitive promoter system. Though integrations of the gene and entire upstream intergenic region were successful, the insertion mutation in *prkA* would prevent hemi-autotrophic growth. An alternative control construct was designed until site-directed mutagenesis optimization allowed for the correction of these plasmids. The *fbp* region and its upstream intergenic region were amplified from the hemi-autotrophic strain by PCR. This 1174 bp fragment was integrated into the pCR2.1 cloning vector by TOPO TA cloning, its integration was confirmed by sequencing. The EcoRI sites flanking the insert were utilized to gel purify the region after restriction digestion and the resulting fragment was ligated into the pUC19 expression vector. Neither sequencing by use of external primers nor culturing to hemi-autotrophic conditions in the Δ*fbp* background strain could confirm the correct integration of the fragment and time constraints prevented further pursuit of qualitatively validating the knock-out.

4.8 Promoter expression assay

The over-expression of the CBB operon is unlikely to result in the failure of hemi-autotrophic growth as the continually maintained presence of CO_2 in cultures should keep the pathway active and prevent toxicity by metabolizing the otherwise accumulating non-native intermediates. Insufficient expression of the operon would however lead to failure of the culture as it can't assimilate the carbon required to maintain its metabolism. Expression levels of the operon as regulated under the *cI857* promoter system can be

predicted using various *in silico* tools but will be indicative at best for conventional strains. In order to measure the *cI857*-dependent activity in relation to the established pTet promoter, *lacZ* was introduced to pCBB (outside of the CBB operon) under three different promoters via Gibson assembly.

The chosen insertion site is located just downstream of *cat* and the first terminator since the start of the CBB operon. The MluI restriction site was utilized to introduce *lacZ* under the regulatory expression of either its native promoter, pTet or the *cI857* promoter system, by use of a two-fragment Gibson assembly as illustrated in figure 4.9. Its construction was performed along with other constructs utilizing a dual operon in section 4.6, and confirmed by restriction digestion for the native promoter and *cI857*-containing constructs as shown in figure 4.10c. Different colonies of the pTet variant were identified by PCR as shown in figure 4.10c, where sequencing confirmed the correct integration in to the plasmid of clone 10 and the exact deletion of its twice-repeated tetO binding sequence *"TCCCTATCAGTGATAGAGA"* in clone 2. The downstream native RBS was included in the *lacZ* amplicon and the terminator just upstream of the origin of replication was left unaltered.

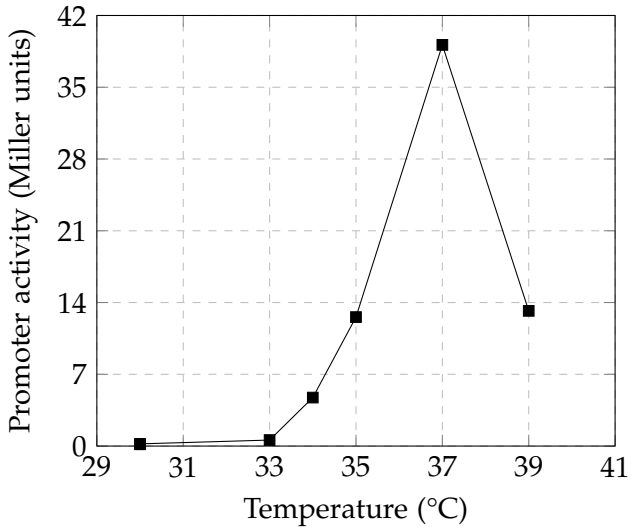

Figure 4.11: Graph of the β-galactosidase activity of pCBB-DO_lacZ cultures expressed in the CBB4 strain plotted against their culturing temperature. Data is normalized

Due to the differential progress in construction of these strains and the

limited availability of temperature-variable incubators, cultures could only be grown at two or three different temperatures at the same time for subsequent assessment in the same β-galactosidase assay. Therefore, either the *cI857* promoter or the native promoter cultured at 37°C were included as an internal reference for any measurement and used normalize temperature variations to. Of the duplex measurements taken per culture, at least one typically fell within the linear range of the spectrophotometer and was used to calculate activity in Miller units (which are independent of OD_{600}). The plot of normalized measurements of the thermo-sensitive promoter system is presented in figure 4.11, with its raw data in supplementary figure D.1. A peak is found at 37°C and since *E. coli lacZ* is documented as comparatively functional in the measured temperature range [125], this activity reduction is likely attributable to the hemi-autotrophic strain's decrease in viability at elevated temperatures as the 39°C cultures grew to approximately half the densities of those grown at lower temperatures.

A trend-line derived from the data-points up to the 37°C peak correlating with an r^2 of 0.943 (as shown in supplementary figure D.2), whereas excluding the single data-point that could only be normalized to a natural *lacZ* promoter measurement (33°C) increased the r^2 to 0.995. Applying these functions within the relevant temperature range resulted in a difference of $<$0.2°C. As the reliable temperature resolution of the used incubators is greater than this potential deviation, the measurement was included. Taking the natural logarithm of obtained and published data [125], and comparing their temperature coefficients after linear regression (r^2 = 0.957 and 0.939, resp.) showed a significant difference despite both using *E. coli* and native *lacZ* as reporters. The published response data is therefore deemed as not directly applicable to its expression in the CBB4 strain or pCBB plasmid.

Assays of three separate cultures of the correctly sequenced pTet-lacZ construct resulted in negligible activity where the pCBB-DOt-lacZ ΔtetO construct displayed a relatively high activity, as shown in table 4.1. Normalizing the obtained β-galactosidase activities in Miller units for this reaction to those found for pLacZ-lacZ in the same assay to the previously derived trend function resulted in a temperature of 36.8°C. This is the equivalent temperature at which the *cI857* system represses pR to the degree its expression equals that of pTet lacking its repressor binding sequence. As the

standard culturing temperature for *E. coli* strains has been 37°C it is deemed as practically indistinguishable from the calculated equivalence point.

Construct	Age (over-night incubations)	pTet activity	Control activity	pTet OD_{600}	control OD_{600}	Assay time (min)
pTet_lacZ	1	0.0047	1.0331	0.676	0.294	78.9
pTet_lacZ	2	0.0045	3.0585	2.131	1.932	156
pTet_lacZΔtetO	1	25.520	0.5718	2.060	1.912	1.55

Table 4.1: β-galactosidase activity of pCBB-DO_lacZ with pCBB-DOn_lacZ as an activity control. Due to low assay inactivity and culture density (consistently observed for this construct) incubation time was extended while also sub-culture for the standard reference, the control regulated by the native *lacZ* promoter was always inoculated on the previous day.

4.9 Inorganic pyrophosphatase knock-out

The most likely cause for the inability of any of the RBS modulated, temperature-dependently expressed constructs containing the synthetic CBB operon was deemed to be the chromosomal *ppa* gene. Constitutively expressed at a relatively high level [118], it efficiently [136] catalyses the splitting of pyrophosphate into two orthophosphate molecules. *E. coli* doesn't significantly express enzymes documented to utilize pyrophoshate and its cytoplasmic concentration is therefore assumed as predominantly dependant on the k_{eq} of inorganic pyrophosphatase. Lacking kinetic data of the enzyme, its equilibrium was thermodynamically approximated and in combination with the concentration of orthophosphate [36], the cytoplasmic concentration of PPi is estimated in to be in the micro-molar scale. This is contrary to the substrate requirements of PFP to catalyse both of its reactions as necessary in the proposed synthetic CBB cycle.

In order to replace the essential *ppa* [132] it will need to be knocked out while expressing a gene of similar effective functionality. HPP is strongly suggested [102] to catalyse the same reaction but due to the lack of published kinetic measurements of either enzyme in *E. coli* or comparable bacteria, the required expression levels are unknowable. As *fbp* is 5.5 kbp removed from *ppa* on the *E. coli* BW25113 genome, the utilization of P1 transduction would have a near-certainty of restoring the Δ*fbp* site. Instead, the gene was attempted for knock-out by performing λ-Red [107] on the *fbp* knock-

out strain from the Keio collection (BW25113) while expressing *hpp* from a plasmid and complete the strain with a P1 transduction of both knock-outs into the CBB4 strain.

The kan^R site that replaces *fbp* in this strain was first removed through FLP recombination. The temperature-induced pCP20 plasmid was transformed into the strain, the plasmid expressed and subcultures screened for loss of pCP20 and kan^R by antibiotic sensitivity. Patched colonies sensitive to both were sub-cultured and confirmed by PCR. Simultaneously, the *cat* gene was removed from pSCBB by restriction digestion with MluI and NheI. The spectinomycin resistance gene *aadA1*, including its promoter and terminator (originating from a TOPO vector generously shared by Minyeong Yoo) was digested by the same enzymes and ligated in. Correct assembly of this pSCBBspec plasmid was confirmed by PCR and sequencing.

Next, λ-RED was applied to the resultant strain by transformation of the Δ*fbp* strain with pSCBBspec, to compensate for the loss of *ppa*, and pKD46[4], for the expression of genes required for homologous recombination, while selecting for both with spectinomycin and ampicillin at 30°C. The knock-out was performed according to protocol, using linearised amplicons of pKD3 (chloramphenicol) and pKD4 (kanamycin)[5], containing a 100 bp homologous region, in parallel. Over a total of four attempts, two plates with chloramphenicol resistant clones were obtained. They identically displayed a dual genotype of both the wild-type and the expected knock-out bands after PCR amplification with external *ppa* primers. Both the expected 1019 bp wild-type and 1521 bp knock-out bands were observable in these colonies, as exemplified in lanes 5 and 6 of figure 4.12. Purification of the knock-out genotype, by sub-culturing of single colonies, was not achieved despite multiple attempts. All derived cultures would eventually lose the knock-out genotype and display the pure wild-type genotype, despite the maintenance of selective pressure with chloramphenicol.

A secondary PPA-expressing plasmid, donated by Dr. Minyeong Yoo, pMTL71301 had been used for the successful complementation of the *Cupriavidus necator* H-16 ortholog under control of the pBAD promoter (plasmid

[4]pKD46 was generously provided by Professor Philippe Soucaille.
[5]pKD3 and pKD4 were generously provided by Professor Philippe Soucaille.

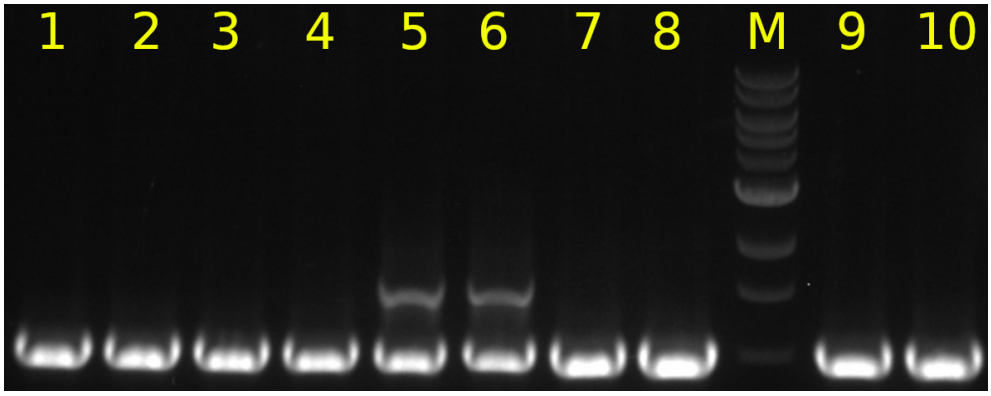

Figure 4.12: Agarose gel of colony-PCR amplifications using external *ppa* primers, two of which display a dual genotype. The wild-type amplicon is 1019 bp and the expected chloramphenicol resistance cassette insert is 1521 bp. The 1kb DNA ladder was used as marker (M) with fragments of 10.0, 8.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.5, 1.0 and 0.5 kbp

map shown in supplementary figure D.1). For future compatibility with pCBB, its *cat* gene was replaced by the kan^R marker originating from pKD4 by use of Gibson assembly and its integration was confirmed by sequencing. This plasmid was transformed into the two double genotype strains, induced by 0.1% L-arabinose and sub-cultured twice more while maintaining counter-selection and induction throughout. Despite maintenance of antibiotic selection, PCR analysis using external primers saw a reversion to wild-type for 8/10 of the found colonies as shown in figure 4.12. It is unknown by which mechanism this dual genotype occurred but time limitations prevented further experimentation surrounding these phenomena and further development of the *ppa* knock-out.

4.10 Discussion

None of the plasmid designs constructed achieved hemi-autotrophic growth using the established methods, where the parental strains re-transformed with its pCBB as control did. In all designed synthetic CBB cycle constructs, except the first, the CBB operon was regulated by the temperature-sensitive *cI857* system. Despite this commonality, the used pR promoter was proven active in LB medium and translates to an enzymatic activity level similar to that of the unrepressed pTet when cultured at 36.8°C. The small difference between test conditions and the calculated equivalence temperature are unlikely to be solely attributable to the total failure of growth observed in the synthetic CBB strains. The promoter regulating expression of the repressor is constitutively transcribed and not documented as affected by any of the molecular changes that could result from a switch to hemi-autotrophic medium [137].

Hemi-autotrophic control constructs for the in vivo activity of PFP were not observed to grow while the pCBB control was. The enzyme has been characterised in *E. coli* and shown able to perform both the *prkA* and *fbp* reactions when provided with the necessary metabolites [102]. There are no predicted expression changes resulting from the difference between LB to M9 that wouldn't equally affect the pCBB. The constructs that incorporat both *pfp* and *prkA* in the Δ*fbp* strain should have been able to grow hemi-autotrophically as its unique Pi-requirement is assumably met but this even failed in a fbp^+ background. Similarly, expression of the synthetic construct replacing *prkA* transformed in the fbp^+ strain should only fail hemi-autotrophic growth in the absence of sufficient PPi, though the presence of HPP is predicted counteract this. These combined results suggest wrongful assumptions were made about the catalytic properties of PFP and HPP or the CBB operon's regulation in this plasmid.

As *cI857* was shown to temperature-dependently repress pR to the equivalent transcription level of pTet at 36.8°C and PFP is expected to largely function when expressed [102] , the pCBBs_cI-E_prkA construct should have resulted in hemi-autotrophic growth. These pre-cultures were grown at 30°C as a precaution against mutations in the operon. Its hemi-autotrophic overexpression at 39°C should have been viable as CBB activity protects against

PRK toxicity. By the same logic, the construct with *pfp* inserted between *cbbM* and *prkA* should have grown too as it didn't interrupt any regulatory sequences. The observed reduction in *lacZ* activity at 39°C is likely due to the strain's altered metabolism reducing cell viability [125] below the level its increased translation rate can compensate for. Though growth of the hemi-autotrophic strain was demonstrated at 30 and 39°C in medium and on plate, no growth rates were measured for these. Considering the use of a temperature-sensitive promoter, its growth at non-standard temperatures should have been investigated further. A multitude of alternatives is available but few offer the uncomplicated modulation a temperature-sensitive promoter with a relatively high dynamic expression range offers [125].

The experiments conducted here provided no clear evidence for why the pCBB operon activity wasn't successfully replicated in any of the pSCBB constructs, under control of the *cI857* system. Results indicate it is possibly related to the deletion of the *tetR* fragments in the pCBBs plasmid these constructs are derived from. The decision for subsequent engineering with this plasmid was based on a single successful pCBBs culture that wasn't replicated until a second pCBBs was established as an intermediate of the dual operon constructs. The severely stunted growth in these cultures could have explained the failure of its derivatives if they had utilized pTet. This suggests the deleted region regulates another interaction or the operon requires a specific regulation on pCBB.

Though expression of the synthetic operon was controlled by the temperature-dependent *cI857*, relative expression of its composing genes wasn't variable. The duality of PFP in the CBB cycle limits the concentration range of PPi required for activity in both reactions. Their unique kinetic properties, activity of the cycle, availability of metabolites and heterologous expression of the gene make it unlikely minimal plasmid modifications combine to reproduce native-like activity. The introduced C and D RBSs were reported to induce higher translation rates than the E RBS present upstream of *prkA* on pCBB [135]. Considering the catalytic properties of PFP in comparison to *prkA*, a higher expression level is needed to support the same effective catalysis rate for the FBPase reaction. Lack of hemi-autotrophic growth in these strains signifies that either an equivalent level of *pfp* expression or a catalytic balance between its FBPase and PRK reactions, or its metabolites

was not achieved. No tools or datasets usable for the estimation of native gene expression in *M. capsulatus* bath were found.

The three different RBS sequences covered a relatively small fraction of potential expression space. A larger range of RBS sequences should have been tested for the heterologous expression of *pfp-hpp*, though determining which range and resolution would need coverage in their constructs is difficult to estimate. Compared to its baseline at 36.8°C, a maximum of 27-fold induction at the cost of a 6-fold reduction in cell viability was reported in *E. coli*, assuming an unmodified *E. coli* metabolism with a typical doubling time [125]. This induction range is relatively low for the overexpression of a gene that is completely uncharacterised. Use of a different inducible promoter, sensitive to either a non-metabolized chemical or physical characteristics [138] as trigger signal, could prove more compatible with a high culture doubling time.

The promoter assay strongly indicated a residual interaction between pTet missing its tet operator and the concatenated *tetR* fragment (as shown in section 4.8). Its DNA-binding domain was indeed found annotated to amino acid 25 to 45 (UniProt accession P0ACT4), well before its transposase insertion at nucleotide 382. The effect of the transposon-concatenated *tetR* on pTet remains uncharacterised and makes it impossible to accurately compare these promoter strengths. Why the tetO deletion only occurred upon introduction of a second pTet on the same plasmid remains ambiguous without specific analysis but it suggests a remnant regulatory mechanism with an expression fine-tuned to a single tet operator. The *tetR*-tetO interaction was likely too titre-sensitive to sustain hemi-autotrophic growth after introduction of a second binding site.

Mutations were found in nearly every construct design employing the full or synthetic CBB cycles, exclusively in functional regions of the (synthetic) CBB operon. A culturing temperature of 30°C greatly reduced the amount of found mutations in the temperature-sensitive operons but the those found were most prevalent in *prkA*, although they were also found in *cbbM*, *pfp* and *hpp*. As *prkA* produces the heterotrophically toxic RuBP [27] it would logically mutate the most readily. The multiple unique *pfp-hpp* mutations found signify its toxicity and subsequent temperature-sensitive regula-

tion of the operon demonstrates its effective modulation, even in rich media. The effort required to correct just one mutation in only one of the dual operon constructs demonstrated the selective pressure acting against the functional CBB cycle. Cloning and verification of constructs is already time-consuming, fixing mutations only compounds this and until a method of transformation that allows for maintaining hemi-autotrophic pressure is developed, mutations are likely to remain status quo for this operon. Growth at a lower temperature might provide some improvement but as pCBB doesn't contain a temperature-sensitive operon, its impact will be limited.

The genomic replacement of *fbp* with the kan^R knock-out cassette was confirmed in three clones by PCR using external primers after selection for the introduced resistance marker. It was confirmed by its failure to grow when transformed with pCBB and sub-cultured to hemi-autotrophic medium with the successful growth of its *fbp*$^+$ parental strain as control. This also demonstrated the paralogs *glpX* and *YggF* were not expressed sufficiently to functionally compensate for it, and therefore interfere with the replacing PFP reaction. Despite the development of five different constructs containing *fbp*, no complementation was achieved under hemi-autotrophic conditions. The three constructs expressing the gene independently from the CBB operon were proven to contain a mutation in *prkA* that couldn't be corrected timely and made hemi-autotrophic complementation impossible. The pUC19-fbp construct failed confirmation by both hemi-autotrophic growth and sequencing but had its insert confirmed by PCR. Finally, the extensively tested pCBB-fbp plasmid never supported hemi-autotrophic growth despite successful pCBB controls and confirmation of its sequence. The background strain contains the correct genotype and has been confirmed by failure of hemi-autotrophic growth. Though the *fbp* knock-out has been genotypically confirmed, a positive confirmation by complementation remains elusive, likely due to the regulation of pCBB in which it was integrated is the only validated *fbp*-complementation construct.

Assuming the functional expression of *pfp*, it was expected at least one of the tested RBS variants of the various CBB control constructs would have been able to catalyse either of its reactions at the tested temperatures of 34, 36, 37 or 39°C. It produces less pyrophosphate in its FBP reaction than its PRK reaction consumes but HPP is predicted to natively compensate for

this difference by maintaining cytoplasmic PPi. Assuming its functionality in *E. coli* from its inferred toxicity on the unmodulated construct, the most likely source of disruption to the delicate PPi balance it should maintain is inorganic phosphatase. Encoded by the essential *ppa*, gene it maintains a low concentration of PPi as it isn't used as energy carrier in the native *E. coli* metabolism. Its highly and constitutive expression [118] quickly catalyses the breakdown of pyrophosphate into orthophosphate. No successful knock-out of *ppa* was achieved and though dual genotypes were observed in several colonies, purification remained unsuccessful. Which mechanism the wild-type reverting colonies employed to retain antibiotic resistance without the knock-out marker is unknown.

Compensating for the essential functionality of PPA by use of the uncharacterised heterologous *hpp* gene was a risk that hasn't yielded a reward. Complementing the double genotype strain by expressing the *ppa* gene from a plasmid proven effective in *C. necator* was unsuccessful too. This is likely for the same reason as it has a maximum identity of 69% over a coverage of 88% to the *E. coli* variant, though it could also be caused by its induced expression. Limited time prohibited the implementation of a constitutive promoter or the expression of native *ppa* on a plasmid. Until either construct is tested, the possibility of a knock-out can not be dismissed. Functional complementation of inorganic pyrophosphatase by a H^+-PPase has been observed in *Saccharomyces Cerevisiae* [139]. Of the gene analogues used here, *E. coli* and *C. necator ppa* have a significantly greater sequence identity but as neither has been kinetically characterised, their functional similarities can't be speculated upon. *M. capsulatus* expresses both the *pfp* gene used here and a *ppa* gene with 72% identity over 66% coverage with the *E. coli* variant, suggesting a complimentary function between its native variants.

If more time had been available, a construct expressing the native ortholog should have been assembled but it is still disputable whether *ppa* is responsible for the unsuccessful functionality of the *pfp-hpp* operon. The putative functionality of the membraneous HPP involves proton transport as well, likely using the proton gradient to store generated energy in pyrophosphatase. Though it is likely it maintains PPi at a higher cytoplasmic concentration than PPA and whether similar enough conditions for its activity are present in *E. coli* is unknown. Regardless, it is possible HPP could inter-

fere with native mechanisms such as respiration or phosphate transport and even create a futile cycle wasting membrane potential. Unless HPP is functionally characterized (in *E. coli*), exploring alternative methods to maintain a moderate supply of cytoplasmic PPi might be more likely to yield timely results.

The restoring mutations found in a leave-one-out experiment performed on five key deletions from the strain [95] could also provide valuable improvements. Most of the mutations that didn't restore the knock-out gene were related to transporters, carbon or phosphate metabolism. The most occurring gene in these new mutations was *pitA*, coding for a phosphate transporter in 7 of the 31 variants found in *prs* and *crp*. This is of particular interest in relation to the synthetic CBB cycle as it is likely to severely impact cytoplasmic (pyro)phosphate concentrations. This knock-out wasn't tested because PPA was more likely to interfere with HPP because it has the same net reaction. Regardless, *pitA* potentially affects hemi-autotrophic growth and as phosphate regulator could interfere with optimal HPP function, making it a potential target for investigation. Mutations in the pyruvate transporter *yjiY* were found 7 times as well as in the two high density-growing strains of the initially evolved strains. A high priority could also be assigned to *cyaA*, with 3 of its 5 mutations co-occurring with *pitA* it is also closely related to phosphate and energy management.

In conclusion, no functional adaptations to the hemi-autotrophic CBB cycle were successfully cultured under its restrictive conditions. The omission of a reportedly non-functional region of the plasmid was shown to have a significant impact on the cultures harbouring this shortened pCBB plasmid. This by itself is not sufficient to explain the failure of the synthetic CBB cycle as it is predicted to increase the organisms energy and carbon efficiency. The FBPase gene *fbp* was successfully knocked out from the hemi-autotrophic strain but was not functionally complemented. The reliance on cytoplasmic availability of PPi by the introduced PFP of the synthetic CBB is the most likely cause of its dysfunction. Temperature-dependent regulation of the CBB genes reduced mutations in the operon but neither it nor RBS variants of its consecutive genes enabled hemi-autotrophic growth. As PPi is not natively used as a source of reductive potential in *E. coli* and is accordingly oxidized to orthoposphates by the abundant inorganic pyrophosphatase, it

is another potential source of interference with the synthetic CBB cycle. It is encoded by the essential *ppa* gene and should have a similar function to *hpp*, which is encoded on the same operon as PFP but its function is predicted as coupled to trans-membrane proton-transport. A knock-out of the native gene while expressing the uncharacterised HPP resulted in two obtained cultures displaying both the wild-type and knock-out genotype when PCR amplified by external primers but proved impossible to purify by consecutive generations of single-colony, selective cultures.

4.11 Acknowledgments

Additional consideration is owed to Professor Philippe Soucaille for envisioning and pursuing this bold pathway design, keeping the project aligned with its initial objective and obtaining external genetic materials relevant to the project. This includes the CBB strain, *M. capsulatus* Bath strain, the temperature-sensitive *cI857* promoter-repressor system on plasmid pIB4, pKD3, pKD4 and pKD46. Dr. Minyeong Yoo cloned the *pfp-hpp* operon from the *M. capsulatus* Bath genome onto a TOPO vector and designed its initial integration into pCBB before handing over the project in good order.

Chapter 5

Analysis and modelling of the (synthetic) CBB cycle

"Working on Collector data. Have ruled out artificially intelligent virus. Unless it's very intelligent and toying with me! Hmm, tests."

Mass Effect 2
BioWare

5.1 Introduction

The experiments conducted on the hemi-autotrophic CBB4 strain, obtained in the preceding chapters, are contrary to previous publications [94, 95] and conventional molecular biology. The demonstrated reduction in culture viability following the removal of plasmid regions documented as non-functional was unexpected and detrimental to the designs achieved in the work presented so far. The unexpected behaviour of the CBB operon and the failure of modifications made to it also supports the presence of additional regulatory mechanisms. Furthermore, the question of whether or not the *pfp* gene is a viable replacement for both *fbp* and *prkA* to form the proposed synthetic CBB cycle remains to be answered.

The objective for this chapter is to use in silico tools in the analysis of the metabolism and functional genetic components of the established hemi-auto-trophic CBB cycle, and its hypothesized synthetic derivative. It was attempted to reduce the amount of assumptions made regarding the CBB4 strain by using available kinetic and sequential data to infer their impact on the hemi-autotrophic strain and its derivatives. The *in silico* tools used in this chapter were controlled against established data wherever possible. Adaptations of established, peer-reviewed models and predictive tools were kept to a minimum. When multiple quantitative inputs were supported by the available literature for these methods, they were implemented at the disadvantage of this work's hypothesis. Comparisons to the proven growth of the hemi-autotrophic strain harbouring pCBB remains the comparative standard for its derivatives.

5.2 Sequence analyses

The hemi-autotrophic CBB4 strain of *E. coli* used in this work is reported to harbour a plasmid containing an operon that consists of *prkA*, *cbbM* and *CA*, each with an RBS to control expression of the downstream ORF [94]. After cultures of engineered RBS variants expressed under a temperature-sensitive promoter returned results that weren't explainable by an operon expressed by a single promoter, (in section 4.5) the published plasmid sequence (KX077536) was analysed. Promoters, RBSs and terminators on pCBB along with any other genes relevant to the project were investigated.

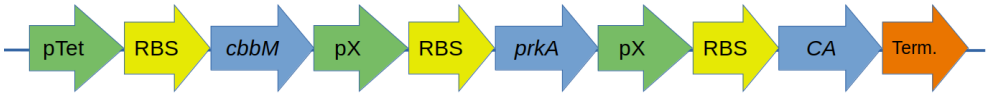

Figure 5.1: Schematic representation of the pCBB structure established following the analysis of its sequence. Red genetic elements signify a terminator (Term.); blue, genes; green, a promoter (X signifies an unknown name); yellow, an RBS; and line, an arbitrary nucleotide sequence. Displayed genetic element sizes don't correspond to their nucleotide lengths.

As shown in figure 5.1, individual promoters were found upstream of each gene in the region documented as the CBB operon, as opposed to the structure deduced from available documentation [94] and visualized in 4.4. These promoters exist in the ¡60 nucleotide 5′ untranslated region (UTR) that precedes the genes in this operon. When analysed in different constructs, these reliably produced the same transcription rate scores (as shown in table 5.1a), showing their expression rate is independent of downstream sequence. Performing similar analyses for RBSs demonstrated the translation rate downstream from the synthetic C, D and E variants [135] predicted translation rates that vary significantly depending on their downstream sequence. As shown in table 5.1b their relative translation rates didn't correlate proportionally, as was intended. An analysis of the *prkA*-variant plasmid under the B RBS returned in a score of 1096, demonstrating that the published proportionality largely holds true in this context and this outlier is likely dependent on a specific interaction between the C RBS and its downstream *prkA* sequence.

The only terminators found on pCBB were surrounding its origin of replication, just upstream of the *cbbM* gene and just downstream of the *cat* gene, without any between the three genes making up the purported CBB

operon. The chromosomal *fbp* gene is counter-orientated to its neighbouring genes and surrounded by bidirectional terminators, which were included in all clones incorporating this gene. The native *lacZ* gene is immediately followed by a terminator that was similarly included in derived constructs. The *cI857* module contains a bi-directional terminator downstream of the repressor gene. No terminating sequences were found within the recombined region of the *pfp-hpp* operon.

(a)

Context	Promoter	LDF
pCBB	ORI	4.13
	BLA	2.00
	transposase	2.17
	cbbM	4.28
	prkA	3.85
	CA	4.32
	cat	5.33
Thermo-sensitive	*cI857*	3.95
	pR	4.83
Synthetic	*pfp-hpp*	2.34
	$spec^R$	2.70
Chromosome	*fbp*	3.31
	ppa	4.05
	lacZ	1.94

(b)

Context	Promoter	Translation rate
pCBB	ORI	457
	BLA	1641
	transposase	40
	tet^R	556
	C:*cbbM*	7474
	E:*prkA*	267
	D:*CA*	842
	cat	11432
Thermo-sensitive	*cI857*	146
Synthetic	E:*pfp*	241
	N:*hpp*	3164
	N:$spec^R$	160
	C:*prkA*	10947
	D:*prkA*	558
	E:*prkA*	442
	C:*hpp*	172
	D:*hpp*	105
	E:*hpp*	132
Chromosome	*fbp*	124
	ppa	3584
	lacZ	415

Table 5.1: Calculated values for the promoters **(a)** and ribosome binding sites **(b)** relevant in this work. Linear Discriminant Function (LDF) is a score used to indicate promoter strength. The RBSs scored precede the gene stated, if available they are defined by their published descriptor [135] preceding the gene's name where 'N' is used to denote its native RBS.

5.3 Thermodynamic model

Thermodynamic analyses of the CBB cycle, independent of the kinetic properties, catalytic efficiencies, or translation rates of the various enzymes involved, were performed by use of the eQuilibrator web-tool [116]. The Gibbs free energy values of each reaction's substrate and product, provided by the software, were compared and the total feasibility of the pathway determined for a range of reactant concentrations. For the hemi-autotrophic CBB cycle the stoichiometric net reaction is "9 ATP + 6 NADH + 3 CO_2 + 5 H_2O = 6 NAD^+ + 9 ADP + 8 Orthophosphate + D-Glyceraldehyde 3-phosphate" and resulted in a ΔrG'm of -230.6 kJ/mol. This is a measure of the predicted Gibbs free energy difference in the metabolite conversions in the pathway's combined reactions, under assumed physiological conditions of 1 mM reactants. The driving force was calculated between 80.9 and 174.9 kJ/mol, depending on metabolite availabilities, and resulted in a Min-max Driving Force using optimized metabolite concentrations (MDF) of 5.6 kJ/mol for its bottleneck reactions under physiological conditions (fig. 5.2a). It also shows in fig. 5.2b that most secondary metabolites pose bottlenecks to the pathway under physiologically accurate settings, along with the Ru5P and 3PG that both interact with the heterologous reactions. The indication that the enzymes of *E. coli* are bottlenecks to the CBB cycle and the introduced PRK and RuBisCO aren't (as shown by the green and blue lines, respectively, in figure 5.2a) indicates the native conditions are not optimized to support this carbon fixation.

Replacing FBPase and PRK with their equivalent reactions as catalyzed by *pfp* in the synthetic CBB cycle changed the net reaction to "6 ATP + 6 NADH + 3 CO_2 + Diphosphate + 3 H2O = 6 NAD^+ + 6 ADP + 7 Orthophosphate + D-Glyceraldehyde 3-Phosphate" and resulted in a ΔrG'm of -133.0 kJ/mol. The driving force was calculated between 25.5 and 103.5 kJ/mol, resulting in an MDF of 2.1 kJ/mol as all reactions where bottlenecked by the metabolites at physiological concentration ranges. Fig. 5.2c also shows a shift the reaction profile of the pathway. As the overall energy difference decreases, the reactions of glyceraldehyde 3-phosphate dehydrogenase, triose phosphate isomerase, fructose 1,6-bisphosphatase and transketolase now perform below the bottleneck criterion. Under the optimized conditions for the synthetic CBB cycle, similar limiting metabolites were found as for the

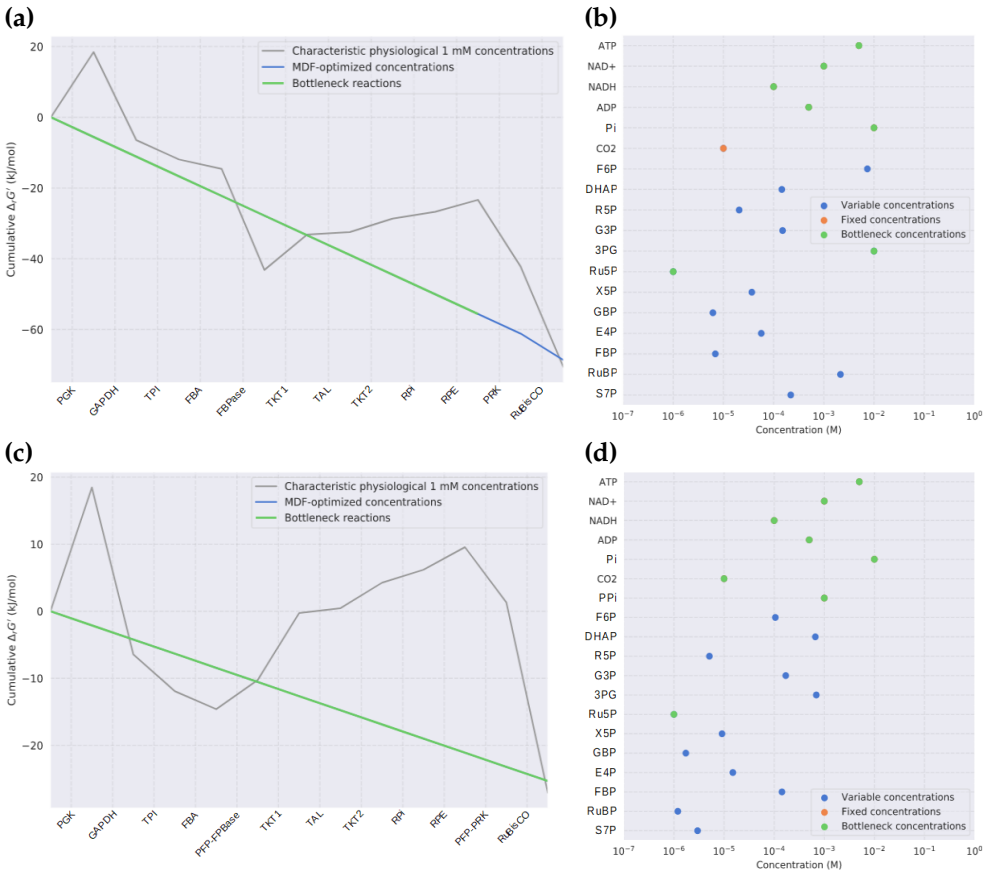

Figure 5.2: Results of the thermodynamic analysis using eQuilibrator [116], of the hemi-autotrophic (**a** & **b**) and synthetic (**c** & **d**) CBB pathways. Each reaction's ΔrG'm is sequentially displayed compared to its overall ΔrG', indicating bottlenecked reactions, when assuming physiological metabolite ranges, with the green line averaging at the determined MDF (**a** & **c**) and specifiying which metabolites are causing those bottlenecks (**b** & **d**).

hemi-autotrophic variant except for the addition of PPi and CO_2, as shown in figure 5.2d. 3PG was also no longer limiting which indicates RuBisCO would here operate at a higher rate in comparison to the rest of the pathway, both Pi and PPi are predicted as limiting under physiological concentrations for the PFP reactions. This indicates the wild-type CBB cycle is both energetically more favourable than the synthetic and has fewer metabolite-adjusted bottleneck reactions. The synthetic cycle has a significantly lower demand for energy and saw the capacity for CO_2-assimilation increase as the 3PG bottleneck disappeared with its changed thermodynamic profile.

5.4 Kinetic model of the (synthetic) CBB cycle

The model used for thermodynamic analysis could have been expanded using its enzyme cost minimization function but as it requires detailed parameters from enzymes that aren't all (well-)characterized, an alternative was used. To complement the thermodynamic model, the CBB cycle was analysed using the kinetic enzyme parameters and metabolite concentrations found in *E. coli*, or best available representable parameters for each of the involved reactions (of relatively closely analogs). The found constants were used to build a kinetic model that omitted secondary metabolites as their cytoplasmic concentrations are largely dependent on external factors and their kinetic parameters are often unreported. The cycle was stoichiometrically balanced by omitting the CO_2 from RuBisCO and generating 5 molecules of G3P from 3 RuBP molecules for mathematical stability.

The k_{eq} used in the model was either derived from the enzyme's kinetic parameters, if available, or the Gibbs free energy difference between its metabolites. From the available data, the lowest calculated equilibrium constant in the direction of the CBB cycle was used to prevent overestimation and was considered irreversible if above 10^4 (or below its inverse). These parameters are shown in table 5.2 with those used for this model in bold. The reaction's calculated Gibbs free energy (ΔG) is expressed in kJ/mol at 37°C. Reaction types used in the limited CBB model are either expressed in a Michaelis-Menten (MM) or mass action (MA) equation, depending on the available enzyme kinetics (V_{max} and K_M or k_{kat} and k_{eq}, respectively) and obtained as described in materials and methods section 2.4.4. For mass action equations, the V_{max} of the forward reaction in the CBB cycle was converted to k_{cat} if it was lower than the thermodynamically derived k_{eq}. MM equations with more than one substrate or product required a calculated k_{eq} component too.

Cytosolic start concentrations for the metabolites considered in this model were obtained from cytoplasmic measurements [140] and the reaction rates were multiplied by protein abundance [118, 141, 142], making reaction rates more accurate in the model but its time arbitrary. Abundances for PRK and RuBisCO as expressed from the low copy-number pCBB plasmid, were derived from their predicted promoter strengths, then normalized to that of

wild-type *fbp*, and multiplied by its abundance and p15a copy number [143].

	Forward			Reverse										
	V_{max}	K_M1	K_M2	V_{max}	K_M1	K_M2	K_M3	$k_{eq}1$	ΔG	$k_{eq}2$	Type	Rev.	Abundance	Ref.
PGK	**233**	0.17	0.125						-19.5	**1931**	MA	Y	6137.39	[144]
GAPDH	**1520**	**0.015**	0.007	**1783**	**1.5**	0.042	22	85.250	-1.2	**1.593**	MM	Y	18219.8	[145, 146]
TPI	**20**	**0.3**		**7.4**	**2.8**			25.225	-5.6	**8.783**	MM	Y	2924.99	[147, 148]
TAL	**4.8**	**0.038**	**0.285**	**60**	**1.2**	**0.09**		2.526	-0.9	**1.418**	MM	Y	5689.69	[149]
FBA	**4.3**	0.03		16	0.17			**1.523**	-23.2	8115.5	MM	Y	3038.54	[150, 151]
FBP	**24**	0.015							-8.6	28.128	MA	Y	1042.49	[152]
TKT1	62	**1.1**	**2.1**	**110**	**0.16**	**0.09**		0.082	10.3	**0.018**	MM	Y	1837.41	[153]
TKT2	100	**4**	**2.1**	**50.4**	**1.4**	**0.16**		0.694	3.8	**0.229**	MM	Y	1837.41	[153]
RPE	**2987**	**4.8**		**1149**	**3.2**			**1.733**	-3.4	3.740	MM	Y	666.32	[154, 155]
RPI	**5512**	0.2							2.1	**0.443**	MA	Y	1308.17	[156]
PRK	**300**	**0.27**							-28.6	**65954**	MM	N	42.6748	[133]
RuBisCO	**6.1**	**0.0128**	0.446						-32	**246693**	MM	N		[157]
PFP1	9	**0.328**	**8.69**	7.6	**2.27**	**0.027**		**0.122**	4.6	0.168	MM	Y		[102]
PFP2	**3**		**0.027**	**4.2**		**8.69**		**229.894**	-15.4	393.53	MM	Y		[102]
PPA	472	**0.0035**							16.5	**0.002**	MA	N	3130.36	[136]
HPP									13.3	**0.006**	MA	Y		

Table 5.2: The reaction parameters used for the composition of the limited CBB model. The used K_M values correspond to the order used in table 1.2. The shown equilibrium constant is derived from kinetic parameters (keq 1), or lacking those from the reactions thermodynamic properties (keq 2). Kinetic parameters are displayed with the accuracy reported and the values used in the model are displayed in bold. Reaction type is determined by the available kinetic data and mathematically expressed in a Michaelis-Menten (MM) or mass action (MA) equation. The protein abundance in *E. coli* is derived from published data [118].

As shown in figure 5.3b, the metabolite steady-states resulting from running the simulation resulted in a 5-fold greater concentration of DHAP than for the next highest metabolite, indicative of an FBA bottleneck as G3P is the 4th most abundant metabolite. The intervening F6P, E4P and following FBP abundances point to transaldolase and FBP as further limiting reactions. Figure 5.3a shows fluxes stabilize at levels proportional to each of the reaction stoichiometries as PGK and GDH; RPI, PRK and RuBisCO; TPI, FBA, RPE and aldolase; and transaldolase and TKT fluxes converge.

The FBP and PRK reactions from the canonical CBB cycle were replaced by their *pfp* analogues, including their (pyro)phosphate kinetics and estimated as they're expressed from plasmid. Running the simulation with

these conditions reduced flux to an insignificant fraction of the hemi-autotrophic model while accumulating S7P, F6P and Pi. This was resolved by fixing the concentrations of Pi to 0.374 mM [36] and estimating PPi to 0.001 mM, as approximated from its thermodynamically derived k_{eq} with the abundant and efficient PPA [118, 134].

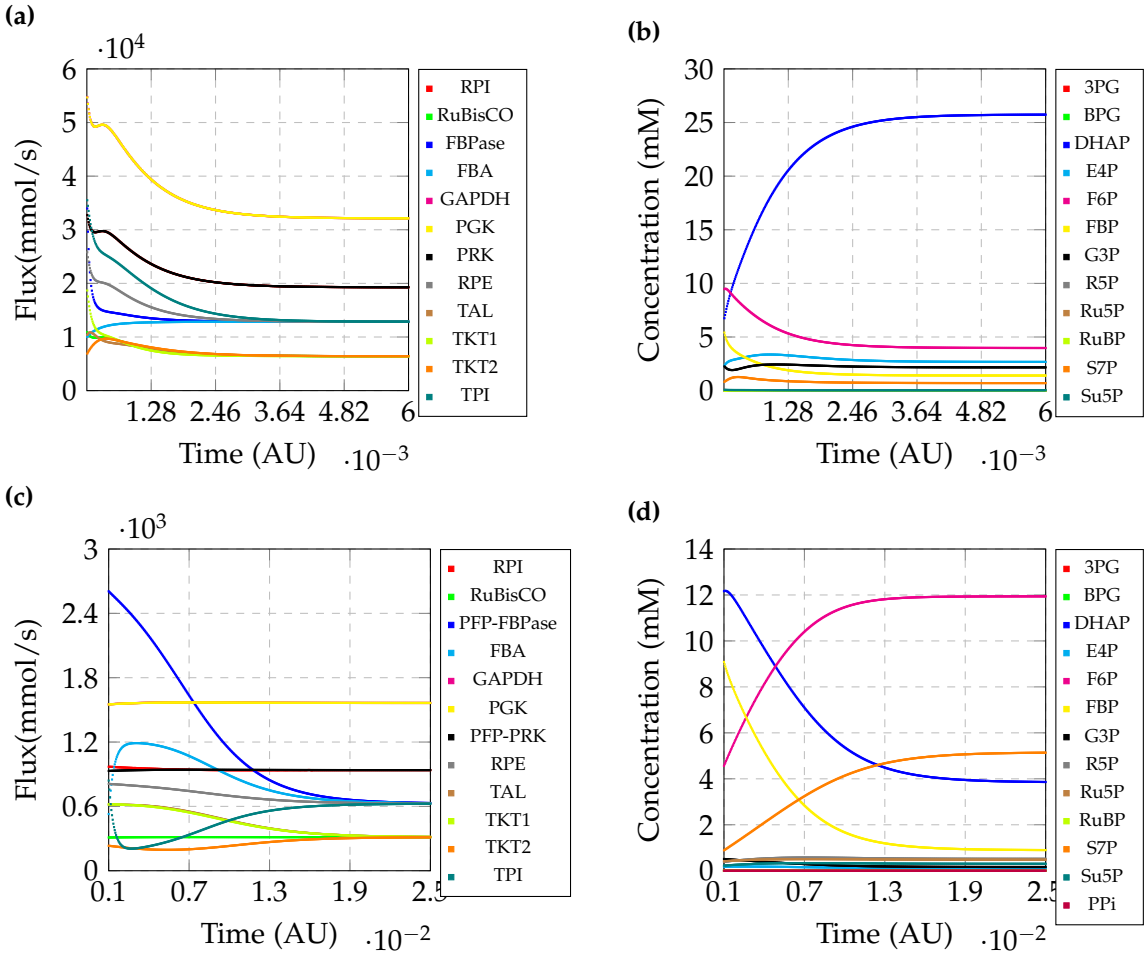

Figure 5.3: Early to steady-state results of simulations of the limited CBB pathway models. Fluxes **(a)** and metabolite concentrations **(b)** of the hemi-autotrophic model, as compared to fluxes **(c)** and metabolite concentrations **(d)** of the synthetic model. Fluxes are mmol/s where concentrations are displayed in mmol/mL, time was modelled in arbitrary units.

The synthetic CBB model shows the same relative stoichiometries but at approximately 20-fold lowered levels, as shown in figure 5.3c. The significant change in metabolite abundancies is shown in figure 5.3d, with F6P and S7P now as the most abundant metabolites after stabilization, points to TKT

or transaldolase as its majorly limiting reactions. DHAP is the remaining metabolite to balance above 1 mM and is indicative of FBA still remaining as a lesser bottleneck reaction. All these reactions require G3P, which is now maintained at a lower concentration than for the hemi-autotrophic model. Increasing the total catalysis rate of both *pfp* reactions by 20-fold saw an approximate restoration of the hemi-autotrophic flux as observed in the previous model 5.3a as shown in table 5.3. Increasing FBP and PRK further in this model didn't sustain this effect. Adjusting only the PPi concentration, peaking at approximately 10-fold higher concentration, saw overall fluxes increase by 4.4-fold. Omitting phosphates from the two reactions, like the other energy carriers in this model, resulted in greatly decreased reaction rates overall as it made the *pfp* reactions more reversible. It had no notable impact on steady-state metabolite distributions but demonstrates the importance of energy carriers to drive pathways.

Model	RuBisCO flux (AU)
Hemi-autotrophic	6423.92
Synthetic	312.85
Synthetic (PFP-FBP x 20)	311.01
Synthetic (PFP-PRK x 20)	4230.98
Synthetic (PFP x 20)	6034.31
Synthetic (PPi x 10)	1378.52

Table 5.3: Obtained values for RuBisCO flux for the constructed hemi-autotrophic model and the synthetic model, replacing the canonical FBP and PRK with their PFP-catalyzed reactions. The reaction rate of the two PFP reactions was increased by 20-fold, separately and combined.

5.5 Integrated kinetic model of the (synthetic) CBB cycle

To compare the synthetic pathway to the hemi-autotrophic CBB cycle as more than its thermodynamic, stoichiometric and catalytic properties, it was simulated as part of its integrated metabolic environment. To implement the hemi-autotrophic CBB cycle [94] and the synthetic CBB cycle suggested in this work, several published COPASI and MATLAB-based models of the *E. coli* (central carbon) [158] or C3 plant CBB cycle [159] metabolism were considered for adaptation. A suitable model [119] was selected for not only describing the glycolysis, pentose phosphate pathway and TCA cycle but also for its limited inclusion of aerobic respiration, energy carriers and phosphate

metabolism. Its central enzyme reactions are based on published kinetic parameters, and where available on chemostat data, making it more relevant and applicable model than the considered genome-scale models [160]. This model is completed by an overarching reaction of key metabolites and energy carriers converting to growth, defined as a fraction of its own cell volume to represent the adjacent reactions not included in the model. The only input of the model is glucose and its outputs are acetate and growth.

Using the COPASI software, all 15 parameters of the growth function were optimized using growth flux as its objective function. This was done to optimize the published model and make it comparable to all following, similarly optimized modifications as displayed in table 5.4. This saw growth flux increase by 54% as the published model was adjusted to reproduce existing data (version 1) where model version 2 was mathematically adjusted to maximize theoretical growth instead. Next, carbon influx of the model was equalized to that documented for the CBB4 strain used in this project. A pyruvate concentration of 5 g/L maintained at a dilution rate of 0.035/h, with an optimistically assumed OD_{600} of 1.0 at 0.39 g/L/OD_{600} and using the model's 1.77×10^{-3} Lcyt/gDW, an input flux of 0.122×10^{-4} mol/second was calculated for the model. For adjustment to the 6-carbon glucose in model 3 this flux was halved.

Like all newly introduced reactions to the model, RuBisCO and PRK were added in model 4 as irreversible Michaelis-Menten equations in accordance to the kinetic parameters presented in table 5.2. A mass exchange reaction to supply intracellular CO_2 from an infinite extracellular supply was also implemented. As a constitutively expressed protein in *E. coli*, PPA was included using the thermodynamically derived k_{eq} of 0.001658 and its reaction rate was included with a wide range in the optimization, as documented in supplementary table E.3. Its expression level and specific activity were assumed as high [136, 142] and could be the reason phosphate and pyrophosphate were used interchangeably in the unmodified model. The reaction maintained a negligible flux until the synthetic variant of the CBB cycle was included in the later iterations of this model.

The model's nutrient feed was adjusted from glucose to pyruvate while keeping total carbon input equal in model 5, resulting in the 550-fold drop in

growth flux shown in table 5.4. To allow for the transcriptional flexibility required for the substrate change, the K_M values of 22 reactions were included in the optimization within the ranges previously described [161], as shown in supplementary tables E.1 and E.2. These minimal justifiable modifications didn't allow for growth until it was found that PGK and GDH were the bottlenecks. Changing their k_{eq} values from 100 and 20 to their thermodynamically predicted values of 5.179×10^{-4} and 0.6, respectively, didn't result in feasible growth flux either, therefore the k_{eq} for PGK was included in the optimization too. Its range was set from the model's published value as maximum to the thermodynamically calculated value as minimum and its inclusion in the optimization function resulted in a restoration of growth flux in model 6. The assumed amount of CO_2 dissolved in the medium was 22 mM, calculated via Henry's law from the 10% atmospheric CO_2 used in incubators during this project. Using the maximum solubility in water at 37°C of 33 mM significantly inhibited TCA cycle flux by competing with the model's unmodified acetyl-coA forming reaction of pyruvate dehydrogenase. Finding an optimum could increase performance of the model but the incubator-accurate amount was applied to all iterations of the model, supporting both TCA and RuBisCO activity.

To adapt the model to the published conditions of the hemi-autotrophic

V.	PGK flux	Growth flux	Ace_out flux	RuBisCO flux	CO_2 flux	PPA flux	HPP flux
1	3.883E-01	2.752E-05	1.036E-06	n/a	n/a	n/a	n/a
2	3.693E-01	4.251E-05	3.538E-06	n/a	n/a	n/a	n/a
3	1.043E-01	9.101E-06	2.441E-06	n/a	n/a	n/a	n/a
4	6.922E-02	6.239E-06	3.832E-07	2.390E-02	-2.151E-01	7.329E-16	n/a
5	-1.718E-04	1.114E-08	6.178E-07	9.199E-05	-3.649E-01	0	n/a
6	-1.479E-02	8.878E-06	7.561E-08	1.330E-04	-1.517E-01	-3.489E-14	n/a
7	-1.208E-01	8.948E-06	1.221E-06	6.420E-02	-1.500E-01	0	n/a
8	-1.061E-01	8.020E-06	3.329E-06	5.643E-02	-1.723E-01	5.329E-01	1.655E-02
9	-1.212E-01	7.902E-06	2.049E-06	6.393E-02	-1.751E-01	3.185E-02	4.596E-02
10	-8.981E-02	6.751E-06	5.548E-06	4.776E-02	-2.028E-01	n/a	1.205E-02
11	4.121E-09	8.347E-11	1.595E-06	2.648E-08	-3.652E-01	1.396E-07	n/a
12	-1.404E-01	1.016E-05	1.470E-30	7.449E-02	-1.209E-01	1.852E-02	3.665E-02

Table 5.4: The steady-state fluxes in critical reactions of the developed models. As described in the text, version (V.) 1 is unedited as published [119]; version 7 represents the hemi-autotrophic model; version 9 is the synthetic CBB model from this work; version 10 is the *ppa* knock-out model; 11: *hpp* knock-out; and 12: *pta* knock-out variant of model 9. PRK flux was found to be equal to RuBisCO flux in all models, to at least within the accuracy shown.

E. coli strain, the (representative) reactions for *gpm(A/M)*, *zwf*, *pfk(A/B)*, malate synthase A (*aceA*) and isocitrate lyase (*aceB*) were removed in version 7. PPS was given a wide range in its optimization to account for the unquantifiable effect of *ppsR* deletion. FBP was changed from the model's Monod-Wyman-Changeux equation to reversible mass action as it represents the directionality more accurately after the removal of PFK. Lacking published kinetics for its reverse reaction, a thermodynamically derived k_{eq} of 28.13 was used. The knock-outs left PGL, PGI and ENO as futile reactions but after initial balancing, their fluxes stabilized to zero, without causing excessive build-ups of metabolites. Variations in the developed hemi-autotrophic model showed that only knocking down the PTA reaction could yield a growth flux of 1.02×10^{-5} mmol/s in model 12, which is slightly higher than the calculated flux of the hemi-autotrophic strain. As only the pyruvate-induced and published changes in protein expression were considered for the model, this omission of phosphate acetyl transferase wasn't explored further. The best growth achieved with these plausibly modified reaction rates in model 7 was 92% of the 9.722×10^{-6} mmol/s calculated (as the strain's growth rate per second [94]) and therefore regarded as representative enough for evaluation of the genetic changes relevant to the CBB4 strain.

To evaluate the synthetic CBB cycle's feasibility and compare its performance to the hemi-autotrophic variant, the reactions of *pfp* were added to the hemi-autotrophic model in iteration 8, using the kinetic parameters described in table 5.4. The proton-translocating pyrophosphatase *hpp* was adapted from the equation of ATP synthase using two electrons instead of four with a k_{eq} of 6.94 that was calculated from its net reaction and the *E. coli* membrane potential of 200 mV [162]. The V_{max} and k_{cat} parameters of these reactions were included in the optimization with a great freedom of range due to their plasmid-expression and/or uncharacterised kinetic properties (as shown in appendix E.3). These additions decreased growth by 10% as a the two FBP and PRK reactions counteract each other. Deletion of these pre-existing reactions in model 9, confirmed that RuBisCO and growth fluxes were successfully maintained by the *pfp* isozyme. The synthetic CBB cycle and its supporting pathways as simulated in model 9 are visualized in figure 5.4. The synthetic cycle further decreased growth by 1.5% compared to its antecedent and 12% below that of hemi-autotrophic model 7. The further removal of either *ppa* or *hpp* from synthetic CBB model 9 was also simulated,

as the reactions were suspected of competing with each other and could consequently make more PPi available to PFP. In model 10, *ppa* was removed and saw growth flux decrease by another 15%. Deletion of the HPP reaction in model 11 resulted in an infeasible growth as it optimized to a flux nearly five orders of magnitude below that of the synthetic model.

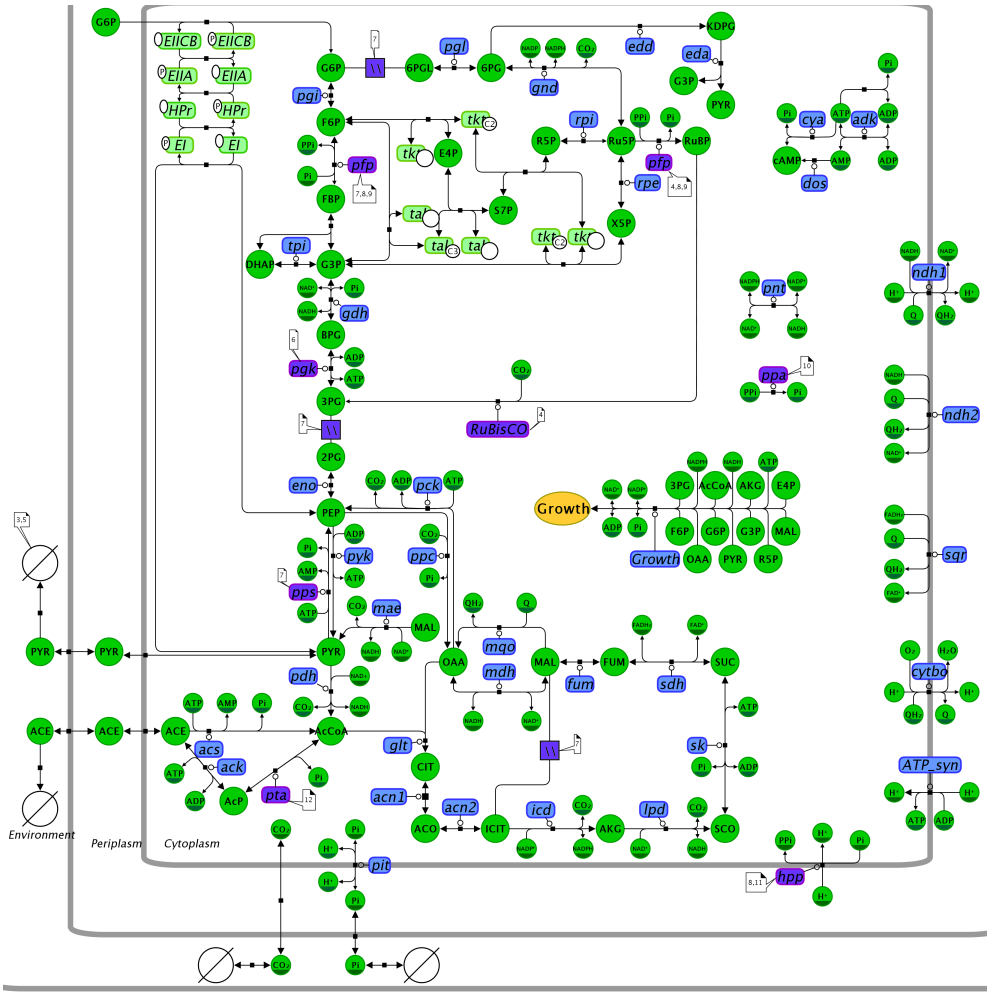

Figure 5.4: Schematic visualization of the synthetic CBB pathway, as simulated in model 9, adapted from the published model [119]. Enzymes are shown as blue rectangles, with those affected by changes in any of the specified models shown in purple, metabolites as green circles, half-reactions and the phosphotransferase system are denoted with green rectangles, transport reactions are omitted and knock-outs are represented by dashed squares.

For the finished and balanced model (7) of the hemi-autotrophic strain PGA3, PYR and BPG were the only variable metabolites with concentrations greater than 10 mM, at 14.5, 11.1 and 10.5 mM, resp. For the synthetic CBB

model (9) PGA3, G6P, S7P and F6P stabilized at 2.6 M, 928, 758 and 335 mM resp. with another six CBB cycle metabolites and pyruvate remaining between 100 and 10 mM. Additional removal of PPA resulted in pyruvate as the only variable metabolite remaining above 10 mM at 53.8 mM in model 11. The post-optimization parameters found for the 4 final models are included in supplementary table E.4. They show an increased tendency to reach the maximum values allowed within the optimization process. Hemi-autotrophic model 7 saw two values reach either limit with one optimizing opposite of its substrate-change transcription range whereas *ppa* knock-out model 11 saw 14 of its 45 variable parameters within 10% of either limit with four of those on the opposite end of its pyruvate-dependent expectation in the allowed range.

The successful replacement of PRK and FBP by the PFP reactions reduced ATP flux in the synthetic CBB cycle by 34% per RuBisCO flux, creating the model as visualized in figure 5.4. Normalizing this relative flux to the reduced growth flux brings ATP use to 74% of the ATP/RuBisCO/Growth-flux found in the hemi-autotrophic model. The same calculation applied to the PPA knock-out model resulted in a reduction of 14% from the synthetic CBB model, decreasing it's ATP efficiency compared to the synthetic model by 16%. Although the synthetic construct conserves ATP in the CBB cycle, it results in a decreased growth rate as the gained ATP is spent elsewhere to compensate for the metabolic changes made to the model.

Comparing the hemi-autotrophic to the synthetic model's fluxes showed the major indicative fluxes changed by less than 2-fold and maintained the same directionality. Fluxes through the TCA cycle increased by 13%, acetate metabolism by 46% and acetate output by 68%. PNT flux decreased by 6-fold, and CYA and DOS decreased by 10-fold in the synthetic model as the likely result of their PPi-dependency. A 46% decreased MAE flux with a 7-fold increase through the Entner-Doudoroff shunt as the membrane proteins had resultant flux changes of less than 25%.

As shown in table 5.5, the largest changes in ATP-reactions (5.5a) were a 9-fold decrease in CYA activity, a 5.5-fold increase in PYK and a 4-fold decrease in PCK with a 19% decrease in PPS flux. These changes show a concentration of flux in the TCA cycle and acetate metabolism(fig. 5.5b),

reducing ATP wastage in reactions included in the models optimization. Performing the same comparison for CBB cycle reactions saw a negligible decrease in the RuBisCO and PRK flux but an 8% increase in RPI and 4% in the TAL and TKT reactions, indicating reversibility of their reactions has increased with their dependency on PFP for completion of the CBB cycle.

(a)

	Model 7	Model 9	Fold change
FBA	-0.04822	-0.04983	1.03320
TPI	-0.04822	-0.04983	1.03320
GDH	-0.12083	-0.12183	1.00825
PGK	-0.12083	-0.12183	1.00825
RPE	-0.04683	-0.04859	1.03767
RPI	-0.01796	-0.01948	1.08458
TKT_Xu5P	-0.04683	-0.04859	1.03767
TKT_F6P	-0.02433	-0.02510	1.03187
TKT_S7P	-0.02250	-0.02349	1.04394
TAL_F6P	-0.02250	-0.02349	1.04394
TAL_S7P	-0.02250	-0.02349	1.04394
FBP	0.04822	0.04983	1.03320
PRK	0.06420	0.06393	0.99584
RUBISCO	0.06420	0.06393	0.99584

(b)

	Model 7	Model 9	Fold change
ACEK1	0.00165	0.00167	1.01205
ACK	-0.06538	-0.09553	1.46119
ACS	0.06537	0.09552	1.46118
ADK	0.18334	0.19058	1.03947
ATP MAIN-TENANCE	-0.36179	-0.26226	0.72489
ATP SYN	-0.21792	-0.25335	1.16259
CYA	0.00043	0.00005	0.10620
GROWTH	0.27299	0.24107	0.88305
PCK	-0.10970	-0.02911	0.26541
PGK	0.12083	0.12118	1.00290
PPS	0.11754	0.09501	0.80834
PRK	0.06420	0.00000	0.00000
PYK	-0.00512	-0.02829	5.52488
SK	-0.06646	-0.07654	1.15170

Table 5.5: Comparison of fluxes in **(a)** the CBB cycle and **(b)** ATP-involving reactions of the hemi-autotrophic (7) and synthetic (9) models, with its fold change in the synthetic model.

Evaluation of the Entner-Doudoroff pathway showed a relatively low flux moving towards pyruvate from CO_2-derived metabolites generated in the otherwise isolated CBB pathway, where removing the EDD reaction lowered growth of the model by half. A PTA knock-out (model 12 in table 5.4) increased synthetic growth by 29%, indicating that conservation of ATP otherwise futilely cycled through PTA, ACK and ACS resulted in increased growth. Calculating the ATP/CBB/Growth-flux ratio results in an ATP usage 58% that of the hemi-autotrophic model, performing 29% better than the synthetic model. Its highest metabolite concentrations are PGA3, BPG, ACCOA and G6P at 734, 24.1, 22.1 and 12.8 mM resp.

As displayed in supplementary table E.4, six of the parameter values included in model optimization reach or approach the allowed limits of the final models, with half of them opposed to their expected direction [161].

Most of these parameters stay relatively close to their initial values despite the range of optimization allowed for them. Sensitivity analysis of all respective 762 and 752 parameters in model 7 and 9 confirmed only 4 and 6 of 45 and 48 respective variables optimized for were in the top 10% of their positive and negative correlations, with none in either top ten. The highest sensitivity values were at least 25-fold higher than those optimized for. The top ten of limiting parameters consisted of Michaelis constants, mainly those of MDH, PPS, PPC and GLT.

The models developed here are based on limited and estimated kinetic data, and were optimized using only growth as its objective function. Regardless, the published model of the *E. coli* central carbon metabolism [119] was adapted to encompass the changes required for hemi-autotrophic CBB cycle [94]. The hemi-autotrophic model (7) replicated the flux calculated from its carbon input with only a 8% theoretical deficiency and further adaptation to the synthetic CBB cycle (of model 9) resulted in a 19% deficiency. Though the PFP-utilizing cycle did consume less ATP, this gain in energy efficiency was not reflected in its growth rate or CO_2 assimilation rate as the supporting metabolism is likely compensating for the changes made. The removal of reactions potentially interfering with the cytoplasmic PPi availability catalysed by HPP and PPA were not predicted as beneficial to the synthetic CBB model. The removal of the PTA reaction was indicated as potentially beneficial, likely as it prevents a futile, ATP-consuming loop.

5.6 Discussion

Sequence analysis of pCBB showed the CBB genes are not expressed on a single operon, as is described in its publication, but with each comprised gene preceded by a corresponding promoter. And without the lack of intervening terminators this operon structure leads to the additive transcription of its constituent genes, initiated by the promoter preceding *cbbM* and not ended until the terminator following *CA*. This means *cbbM* is transcribed only by its directly preceding promoter, *prkA* by both the preceding and that of *cbbM*, and *CA* is transcribed by all three. Considering the possible inaccuracies in current prediction tools for promoters and ribosome binding sites, this additive promoter design make estimation of effective expression levels in this plasmid complex and unreliable. The sequence analysis also revealed flaws in the design of *fbp* complementation constructs. The initial construct, inserting *fbp* just upstream of the *prkA* RBS, separated the coding sequence from its promoter. Most consecutive complementation constructs relied on the cloning of its native promoter that was located in the coding sequence of the upstream *mpl* gene, PCR amplification for these constructs correctly ended at the start of this ORF.

The calculated translation rates of each gene's RBS found a non-linearity in the used sequences [135] as sequence C upstream of *prkA* results in a 20-fold increase from D, where B, D and E were predicted linearly. For *hpp*, these same RBSs result in significantly lower translation rate predictions. Whether or not this is a realistic prediction for *prkA* and *hpp* was not determined in this work but despite demonstrating linearity, the same research also showed that translation levels are dependent on both the gene and its RBS. This discovery likely contributes to the lack of hemi-autotrophic growth achieved with the RBS variants of the synthetic CBB plasmid. The expression range provided by these synthetic RBSs is not as great as intended when compared to the found values for some of the other genes, making expression calibration by mutation extremely unlikely for these genes as down-regulating random mutations are more likely than up-regulation.

The thermodynamic analysis of the metabolites involved in the CBB cycle conclude the proposed synthetic alternative is viable. It also indicated that the functional hemi-autotrophic cycle is likely to be limited by the

metabolites available to it and the synthetic alternative didn't improve this. Replacing the dependency on ATP with PPi for the PRK reaction results in a ΔG lowered by 49% in comparison to the hemi-autotrophic pathway. This reduced energy differential decreases the burden the cycle imposes on overall metabolism but also makes it less likely to move forward efficiently as the involved metabolites are more easily diverted by more favourable reactions competing for the same substrate. The synthetic cycle has a lower driving force and an increase in bottleneck reactions indicates the metabolite availability of the cycle changes to lowered turnover rates of the CBB cycle.

The independent kinetic simulation of the CBB cycle demonstrated that the synthetic pathway is feasible when estimating enzyme kinetics and metabolite availabilities. It also showed that fluxes similar to the canonical pathway, expressed in *E. coli*, are achievable through this adaptation. Although the copy number calculation for the enzymes expressed from plasmids is a greatly simplified estimation, it shows that expressing *pfp* under the same estimated regulation as *prkA* will lead to reduced flux through the cycle, by approximately 20-fold according to the simulation. It also further reinforces the critical importance of energy-carriers and cofactors in driving the pathway forward. The use of PPi as an energy carrier in *E. coli* to drive the synthetic CBB cycle requires careful regulation of catalyzing enzymes and optimal PFP function might require an increase of its cytoplasmic concentration relative to FBP and PRK. The decreased efficiency of the synthetic PRK reaction could even be advantageous as it reduces the concentration of the RuBP that is a source of toxicity to *E. coli*.

The development of the kinetic model of the central carbon metabolism has led to the hemi-autotrophic model (7) maintaining a growth rate of 9.72×10^{-6} mmol/s; 92% of the flux calculated from the chemostat culture's pyruvate concentration, dilution rate and the model's cytosolic conversion rate. Replacement of the FBP and PRK reactions with their PFP isozyme and its supporting HPP reaction in the synthetic CBB cycle resulted in a stable model, simulating growth successfully under the same conditions as the hemi-autotrophic strain. With a growth rate 12% lower than that of the hemi-autotrophic model its metabolite concentrations were severely increased and a larger amount of its optimization limits were approached. Removal of PPA, as it potentially interfered with the newly required phosphate equi-

librium, decreased growth flux by another 14% while resulting in the only modification with realistic concentrations for all metabolites. Removal of the competing HPP instead, reduced growth to inviable levels.

The decreasing growth of modifications to the hemi-autotrophic metabolism was mirrored by an increase in reaction parameters that approached or reached their defined limits, and metabolite concentrations that grew beyond the physiological range. This is indicative of metabolic instabilities, despite their mathematical stability supporting reasonable growth, or the model's inherent properties, such as its unconsidered changes to transcription or allosteric interactions. The synthetic cycle resulted in an ATP/RuBisco/growth-flux ratio increased by the stoichiometrically predicted 33% compared to the hemi-autotrophic version. In spite of the growth-dependent function of RuBisCO, all hemi-autotrophic models produced more CO_2 (in pyruvate and TCA metabolism) than the CBB cycle assimilated. This is indicative of its neighbouring pathways compensating for the decreased ATP consumption in the cycle, determining whether or not this is due to the limitations of the model or a genuine consequence of the modifications made will require experimental support.

The removal of the PTA reaction resulted in the improved growth rate of 1.02×10^{-5} mM/s, 5% above that theoretically calculated for the strains growth in steady-state chemostat. This is due to the futile cycle in its acetate metabolism where ACK generates ATP converting to acetate while ACS expends ATP to regenerate acetyl-CoA. The latter consumes more energy as it reduces two phosphate groups whereas ACK phosphorylates only one. This acetate cycle is present in the unmodified model (1) but was calculated at a flux/growth ratio 18-fold lower, indicating the futility of this loop is exacerbated (and therefore its ATP expenditure) by the modifications made to support hemi-autotrophic growth. No literature was found to support a significant difference in regulation of these three reactions by the conditions simulated here, or the genotype of the CBB4 strain. In combination with the removal of PPA, the model resulted in a growth rate 3% increased and only two (growth) parameters approaching their limit but with 12 metabolites above 10 mM. It is unknown whether or not this is simulated accurately and warrants further investigation as the involved genes are non-essential. Knocking out one of these three genes from the parental strain should be a

relatively simple experiment and would be well-worth potentially obtaining these potentially critical results *in vivo*.

The 8% difference from the rate calculated from the strains published chemostat growth is explainable by the various limitations to the modelling of the pathway. One source of the discrepancy between the calculated growth and the model is the growth function. It is defined as a reaction of various metabolites and cofactors recycled to their counterparts, and represents biomass synthesis. The somewhat flexible and interchangeable reality of biological building blocks is translated to this model as a fixed stoichiometry and was determined as too fundamental to the model for adjustment in this work. Its comprising Michaelis constants were included in the optimization function though. As they aren't based on measured characteristics, these variables are best suited for adaptive optimization without need for reconstruction of the model. The *crp* mutation is cause for the similar inclusion of another four reactions [163] in the model, but their effect ranges aren't similarly documented so they were not thoroughly investigated. The model has approached replication of the hemi-autotrophic growth rate with minimal changes based on the inclusion of supported expression data in its optimization, to minimize the amount of made assumptions and increase its reliability. Newly introduced reactions were given a wide range of optimization (as shown in supplementary tables E.1, E.2 and E.3) as most are relatively variably expressed from plasmid, and *ppa* is known to be highly expressed and catalytically efficient [136, 142].

Without accounting for the substrate-change of glucose to pyruvate, growth was severely restricted. Including documented expression changes as a result thereof in the optimization process increased simulated growth but the parameters generally lingered closer to their initial values than those expected as a result of pyruvate-dependent growth, including the equilibrium constant of PGK. The seperated reactions of TKT and TAL and their variable parameters couldn't be coupled in the used software, though the Xu5P and F6P reaction parameterss of transketolase maintained proportionality throughout the models. Sensitivity analysis showed the largest impacts on growth of the final models were by parameters unchanged in this work, inherent to the model. This was interpreted as an artefact or limitation of the glucose-calibrated model or the intrinsic limits of the involved enzymes

in the pathway as they are known. Though it should be considered that the protein expression changes used to optimize for the media switch were obtained from batch cultures in early exponential phase, instead of continuous chemostat cultures where both the model and hemi-autotrophic growth data are derived from. Although the adaptations to the model were kept to a minimum, they were substantial considering *E. coli*'s change in trophic mode and shouldn't be interpreted as more than best-case scenario's until they can be verified by accurate culture data of any of the involved strains. Using thermodynamically derived equilibrium constants for reactions in the model is a reasonable representation but does not take sensitivities to individual metabolites into consideration and these reaction rates were to be completely estimated, both can be critical factors in the directionality of a reactions flux.

Replacement of its carbon source is a major change the model wasn't designed for [119] as it's largest flux route (the glycolysis) reversed. It is likely to have an impact on the model as it doesn't account for transcriptional changes [164], thought it wouldn't disrupt the inherent kinetics of the simulation. This was reflected in the gluconeogenesis II and pentose phosphate pathway shutting down and the subsequent massive accumulation of metabolites in model 5. Resolving this by including enzymes documented as affected by the applied carbon-source change and the PGK equilibrium constant in the optimization function and restored growth almost completely to rates approaching data published for the hemi-autotrophic strain, supporting its validity as no other relevant but minimal changes achieved a similar effect. As this reaction is not considered limiting in the literature, it could imply a significant inaccuracy in the model's energy metabolism, the same assumption could apply to the ATP wasted in the futile loop found in the model's acetate metabolism discovered in model 12. Inclusion of the ATP synthase and maintenance reactions in the optimization didn't intrinsically resolve this, further re-evaluations of the model's energy metabolism were beyond the scope and resources available for the project.

Considering the large amount of adaptations tested, the used semi-stochastic optimization method resulted in time-efficient generation of results. The obtained solutions approached theoretical maximum of the objective function, though they cannot be considered an absolute optimum solution of

the tested conditions. Differences of decreasing size were observed between iterative solutions for all adaptations, such that the recorded outcomes can be considered an approximation of each model's maximum objective function. Theoretical maxima are calculable by use of less practicable calculation methods than those used here but pose a significant increase in the computing resources required for their completion.

The energetic cofactors in the model are limited by their initial concentrations and their re-oxidation is simulated as a function of the *E. coli* proton-motive force. The energy metabolism of the model employs its main relevant reaction mechanisms but their mathematical parameters were calibrated to support the central carbon metabolism instead of fitting to the sparse data available on these complex processes. The model's inclusion of the transhydrogenase and NADH dehydrogenase reactions allows for the interconversion of energy carriers also including parameters representing the membrane potential. These fluxes changed when less NADPH was used as a result of the removal of ZWF in the hemi-autotrophic model (7). So although energy carriers are limited, increased efficiency in their use is reflected in the growth function as its largest consuming reaction. Growth is an appropriate measure of ATP availability in the model as an increased energy efficiency in the CBB cycle will be reflected in a greater growth flux per ATP spent in it, certainly in an otherwise unchanged background model. Inclusion of supporting energy reactions in the optimization resulted in increased obtained efficiencies but as they didn't yield principal effects otherwise unobserved or significant efficiency increases, they weren't considered as minimal modifications of relevance to the model comparisons.

All three models indicated that the synthetic CBB cycle works at reduced efficiency when compared to the validated hemi-autotrophic variant, but all utilized methods suggest its viability. The thermodynamic and independent kinetic CBB model both saw a great impact resulting from the introduction of PFP but this wasn't similarly reflected in the integrated CBB model. This implies that the cycle's efficiency isn't mainly dictated by energetics or catalytic efficiency of its main metabolites. The bottlenecks observed in the CBB-only kinetic model weren't found in the simulations that included the major adjacent pathways. This is likely due to its secondary metabolites or adjacent pathways driving the cycle forward. The extreme build-up of

PGA3 is attributed to the incorrect functioning of PGK in the model. As its equilibrium constant was given a wide range of variability, the involved cofactors must be responsible for this bottleneck. No literature to support this deviation was found but as no simple or efficient alternatives to enable growth in the hemi-autotrophic model were found, it was kept in the model for reproducibility of hemi-autotrophy.

5.7 Acknowledgments

Eglantine Boudignon performed the initial thermodynamic (and largely unchanged) analysis of the CBB cycles as part of a research project for her Master's degree at INSA Toulouse. Pierre Millard shared enlightening insights in his supportive correspondence regarding the kinetic of model of *E. coli* he worked on, during the early stages of its adaptation to the hemi-autotrophic CBB cycle.

Chapter 6

General discussion

"Evolution is a tinkerer, not an engineer."

François Jacob
Evolution and Tinkering

6.1 In summation

No substantial improvements to the methods of hemi-autotrophic culturing methods were achieved. Transformations were proven as a source of mutations for pCBB-derived plasmids as it requires a recovery in non-restrictive medium but no viable alternative methods were found. The reliability of these transformations and sub-culturing to hemi-autotrophic media was improved and tested with similar results to the parental strain and plasmid [94]. A method of culturing small volumes without the need for a CO_2 incubator greatly increased experimental throughput and reliability. A strain with somewhat improved growth characteristics was evolved but it showed no change to its transformation efficiency. It harboured at least 16 newly mutated genes and a curious undocumented knock-out was found, leaving an FRT scar in place of the *ychH* gene. This gene had no inferred relation to the hemi-autotrophic phenotype but most of the high-frequency mutations did. Though there was little overlap in the found polymorphisms with those previously found in derived strains, no other large or unexpected changes were found. A deletion of the transposase-interrupted *tetR*, and blastomycin, exhibited a suspected reduction in cell viability, the halving of its CBB-dependent growth rate and leads to the interpretation of its regulatory influence on this region of the CBB operon.

No hemi-autotrophic growth was detected using any construct expressing *pfp*. Initial attempts were most likely impeded by its expression under the same regulation as *prkA*, which catalyses its reaction with greater efficiency. Functionality of *pfp* in *E. coli* has been demonstrated [102] and was indirectly confirmed. Expression of the temperature-sensitive replacement promoter of the CBB operon was proven to be equally active to its predecessor near the standard culturing temperature of 37°C. The RBS modulation

applied to the operon still provided a limited expression space and is unlikely to have reached the range required for functioning of the synthetic CBB cycle. Completion of the dual operator constructs would largely allow functional confirmation of the theorized synthetic CBB cycle. Despite lacking complementation results, the *fbp* knock-out strain likely functions as intended. Though the functional expression of *hpp* is likely, it requires confirmation and characterisation in *E. coli* to investigate its exact functionality. Removing the essential *ppa* is unlikely to succeed and an alternative mechanism of regulating cytoplasmic PPi might be more productive. Culturing methods of the hemi-autotrohpic CBB strain should be improved to reduce the occurrence of mutations as they interfere with the efficient progression of cloning *E. coli* traditionally allows for. It was demonstrated that the tetO site of pTet is not inert, contrary to previous reports [94]. Its mechanism should be investigated to determine its effect on any pCBB modifications.

According to the performed predictions, expression of *pfp* is lower than that of *prkA* using the same RBSs and separate promoters were found for the CBB genes, described as a single operon in the publications describing pCBB. The synthetic CBB cycle proposed and investigated in this thesis is thermodynamically feasible but is calculated to have a lowered turnover rate. The kinetic CBB model indicated feasibility of *pfp* as replacement for *fbp* and *prkA* but the lowered specificity and higher reversibility require an $\sim$20-fold higher expression for the replacing enzyme. Its integration in an established kinetic model of the central carbon metabolism showed metabolite fluxes are disrupted by the synthetic cycle to the degree it lowers its overall growth rate. Allowing only for minimal, justifiable modifications the achieved simulated hemi-autotrophic growth rate was only 8% lower than published chemostat data of the strain. Adaptations to simulate the envisioned synthetic CBB cycle sustained growth in the model but at a 12% decrease from the theoretical growth rate. While the ATP requirement of the CBB cycle was decreased by the predicted amount, the neighbouring pathways compensated for the metabolic change to maintain overall functionality. A further knock-out of *ppa* didn't result in higher growth rates for the synthetic strain, it had the lowest steady-state metabolite concentrations of all developments modifications made to the parental hemi-autotrophic model. Adjustments to the acetate and energy metabolism could significantly improve these models but no reported evidence sufficiently in support of those modifications in this

strain has been found yet and warrants further investigation.

6.2 Concluding discussion

The objective of this work, to functionally replace *fbp* and *prkA* with *pfp* in the hemi-autotrophic CBB cycle, was not achieved. This can, in part, be attributed to the discovery that the project was devised on the premise of incorrect assumptions. The CBB operon, as it was documented [94] was analysed to reveal anything but a conventional operon structure. The transcription of *cbbM*, *prkA* and *CA* was correctly reported as collectively terminated but each gene was found to have a small, unique promoter preceding it, additively expressing each sequential gene in the operon. The temperature-dependent (cI857) replacement of the first promoter (pTet) of the operon therefore most significantly regulated the expression of RuBisCO and explains the failure of all constructs designed to employ this regulation. It was further reported that the concatenated *tetR* sequence was inactive, two separate experiments proved this assumption to be false. Deriving the main testing constructs on a plasmid lacking the *tetR* sequence (pCBBs) based on a single unrepeated and unquantified observation of growth, undermined its success. Finally, the obtained CCB strain's stability was overestimated. It was presumed as adaptable as any *E. coli* strain but its unique metabolism and reduced growth rate make it phenotypically more akin to a new species, as it was shown to have barely adapted to its artificial metabolism.

The unreported additive expression of the CBB operon makes it difficult to isolate any effects of its regulation. This is contrary to conventional practises in molecular engineering and complicates the prediction of its expression levels and allow for reliable modifications to it. Why or how this unusual regulation went unnoticed or unreported is unknown. This question similarly applies to the remaining interaction of the first *tetR* fragment with the pTet promoter of the operon, and the discovery of an undocumented knock-out of the seemingly unrelated *ychH* raises the same question once more.

The replacement of pTet with the *cI857* temperature-sensitive expression system negated any residual impact of the partial *tetR* regulation on the construct. Though its activity has been equated to the unregulated pTet

at ∼36.8°C it was shown that pCBB lacking the concatenated *tetR* grows at approximately half the rate, making this equation invalid. Regardless, the construct that only had its pTet replaced by *cI857* and pR should have supported hemi-autotrophic growth, albeit at this reduced rate, and why it hasn't remains elusive. It might be explained by the variant with the specific deletion of tetO showing activity where its full sequence returned no activity in the second pTet on the plasmid. More experiments are required to explain this behaviour as a single pTet did successfully express the CBB operon. Repeating the same promoter assay while expressing *lacZ* under the temperature-dependent system on pCBBs should provide more insight. Insertion of a separate temperature-dependent operon to express any further additions would have isolated the synthetic from the functional hemi-autotrophic regulation. The persistent *prkA* mutation in these constructs prevented this promising approach from timely progressing towards the project's objective.

The strain's inability to transformation directly to hemi-autotrophic media is responsible for the many mutations found to hinder plasmid construction as the CBB operon imposes a metabolic burden that is easily mutated out under heterotrophic growth. This is associated with its decreased cell viability under these selective conditions and as the strain demonstrated a ∼100-fold increased chromosomal mutation rate, adapting it further is likely to at least partly remedy this. This is indicative of the strain requiring further adaptation to stabilize in its dependency on this artificial metabolism, as it was interrupted almost immediately after CO_2 utilization started in the strain. The selection of strain 4 [94] for this project is also controversial. Strains 1 and 3 showed superior growth rates and densities compared to this strain posessing only the minimal modifications required to support the hemi-autotrophic phenotype. Batch evolution proved effective in obtaining mutations that moderately improved some growth characteristics, though further development is required. Until the strain significantly increases in viability, transforming it without intermediate rich culturing will be impossible and mutations unavoidable. An advantage of CBB strain 4 was its limited amount of mutations, making the model signficantly more adaptable to it as the other mutants would require the introduction of greater assumptions.

Despite mention of its functionality, the kinetic characteristics of the PFP

reaction phosphorylating Ru5P remain unreported [102]. Therefore its (P)Pi dependent parameters were used in the kinetic models for their highest possible accuracy. Its FBP dephosphorylating activity is documented and otherwise regarded as functional as its secondary metabolite in the CBB cycle is abundantly available. Its FBPase-replacing reaction was likely insuffiently upregulated to compensate for its decreased reaction rate and reversibility. For the PRK-replacing reaction of PFP, its functionality is still uncharacterized. Simulations indicated the synthetic CBB cycle is thermodynamically and kinetically feasible but its metabolic viability still remains to be investigated as it was demonstrated likely PFP has an insufficient availability of PPi to drive its PRK-replacing reaction forward. Reliable cytoplasmic measurements of the metabolite in *E. coli* are absent in literature and the model instead used a concentration calculated from its thermodynamically derived equilibrium with Pi, equilibrated by the abundant and efficient *ppa* gene. Its accuracy is therefore debatable but its cytoplasmic availability is certainly (undetectably) low in native *E. coli* [140]. Presumably, HPP regulates this in its native context but until this functionality is confirmed in *E. coli*, alternative methods of cytoplasmic PPi regulation might prove worth pursuing.

Sequence analysis indicated that the experimental expression ranges covered by the combination of the *cI857* system and the used RBSs might not have been sufficient. The 3 RBSs for their downstream sequences were retroactively analysed to indicate a high similarity in translation rate with one relatively extreme outlier. This makes the used sequences unsuitable for any indication regarding their intended downstream gene modulation. The temperature-dependent modulation of genes in this strain is likely unsuitable too as its reduced viability during hemi-autotrophic growth makes it undesirable to significantly deviate from the optimum temperature of 37°C. Though these cultures were typically grown at a maximum of 39°C but were shown viable up to at least 40°C. Further and quantitative research should be conducted to determine the suitability of the promoter in this strain as this temperature allows for a maximum expression increase of 11-fold while cell viability drops by 40% [125]. A 20-fold increased expression was indicated as required for the functional replacement of *prkA* by *pfp* in a kinetic simulation of only the CBB cycle. Though a similar simulation integrated with neighbouring pathways brought this down to 10-fold.

6.3 Project limitations and concluding remarks

As a PhD project, the four-year limitation of a study could leave plenty of time for down-stream applications of an ambitious project but unforeseen developments and unprecedented times can similarly shrink its relative space. Adjustments made during its development were prioritized in accordance with the objective. When time was found for *in silico* analyses, several observations became explainable but not in time for their successful practical implementation. At the project's conception the largest difficulties were expected from the extraordinary, multifunctional enzyme but were actually found in a reliable model organism. Its adapted metabolism and growth profile make it phenotypically more akin to a new organism, and it wasn't appropriately regarded as such. This work has identified several challenges in working with this unique strain and addressed issues unconsidered at its conception while attempts to invalidate its hypothesis remain unsuccessful. This work has pursued a rational deduction of the objective's obstacles, starting from a naïve assumption, and has largely cleared a navigable path for anyone following.

6.4 Future perspective

For the continuation of any project utilizing the same hemi-autotrophic strain as this project it is strongly recommended to further develop its genome to improve its growth characteristics and potentially reduce mutability of its plasmid. This could be done by continuing the employed batch evolution strategy but preferably in a substantially larger culture volumes than in this work, or in a chemostat if available for an extended period of time. Though obtaining the other CBB strains would be more advantageous, the mutations documented for CBB strains 1 and 3 [94] can also be replicated to achieve the increased growth characteristics reported for them. Knock-outs in the acetate metabolism were suggested as beneficial by modelling, and as non-essential genes they could prove relatively easy to implement by further P1 transduction. Careful variation of the used hemi-autotrophic medium could signify a modest improvement in the strain's general growth conditions. A quick transformation on competent cells stored in DMSO with a direct inoculation to hemi-autotrophic medium (after recovery) would exclude glycerol as the source of its failure.

To achieve the objective as defined for this work it is recommendable to confirm the function of the Ru5P-phosphorylating reaction of *pfp*. It could prove beneficial to replace the *ppa* native to *E. coli* with its *M. capsulatus* bath ortholog (and its *pitA/B* phosphate transporters), in an attempt to cultivate the phosphate concentrations in *E. coli*. If the dual operon design should prove unsuccessful, the CBB operon and its regulatory elements would need either quantification or substitution and it might prove beneficial to reconstruct pCBB completely using characterised genetic elements. Removing the interference from the interrupted *tetR* gene on pTet and replacing it with a constitutive promoter would be a prerequisite, and terminators would have to be introduced downstream of *cbbM* and *prkA*. As a consequence, the promoters regulating *prkA* and *CA* expression would require adaptation to this singular expression. Reconstructing and recalibrating the CBB operon could be a substantial amount of work and could explain its current unconventional design.

The successful continuation of construct design dependent on RBS variations would require the prediction of its ORF-dependent translation rate. Considering the reaction rate difference of *pfp* with its native isozymes, covering just below 2 orders of magnitude would likely be sufficient. Overexpression of its genes and forcing their modulation by mutation in chemostat evolution might prove a pragmatic approach for this scenario. Appropriate selection of RBSs could complement the used thermo-sensitive promoter but if the maximum targeted expression severely reduces culturability, alternatives must be used. For this slow-growing strain, a non-metabolized induction signal [138] is critical to its success.

The unique hemi-autotrophic metabolism inspires ambitions of full synthetic photo-autotrophy. Though an interesting area of research, the inclusion of CCMs or photorespiratory pathways would be of little benefit to this model organism growing only at increased CO_2 concentrations. The inclusion of proteorhodopsin encoded from a single plasmid, was proven functional [165] and shown to increase culture viability [166] in *E. coli*, could form the basis for its achievement. Alternative sources of energy entering the metabolism outside of the CBB cycle are possible too; the reductive glycine pathway [86] could be implemented in this strain to create a secondary CO_2-assimilating pathway to generate energy from formate. Re-

cently, a soluble methane monooxygenase has been engineered and functionally expressed in *E. coli* [167]. These substrates would all feed into the energy-generating TCA cycle and with some further pathway engineering they could potentially push this inspiring hemi-autotrophic *E. coli* to support a net negative carbon metabolism.

Bibliography

1. Rylott, E. L. & Bruce, N. C. How synthetic biology can help bioremediation. *Current Opinion in Chemical Biology* **58,** 86–95 (2020).
2. Von Hertzen, L., Hanski, I. & Haahtela, T. Natural immunity: biodiversity loss and inflammatory diseases are two global megatrends that might be related. *EMBO reports* **12,** 1089–1093 (2011).
3. Lv, X. *et al.* Synthetic biology for future food: Research progress and future directions. *Future Foods* **3,** 100025 (2021).
4. Yim, H. *et al.* Metabolic engineering of *Escherichia coli* for direct production of 1, 4-butanediol. *Nature chemical biology* **7,** 445–452 (2011).
5. Ongey, E. & Neubauer, P. Lanthipeptides: chemical synthesis versus in vivo biosynthesis as tools for pharmaceutical production. *Microbial cell factories* **15,** 97 (2016).
6. Gong, F., Cai, Z. & Li, Y. Synthetic biology for CO_2 fixation. *Science China Life Sciences* **59,** 1106–1114 (2016).
7. Zhang, R., Li, C., Wang, J., Yang, Y. & Yan, Y. Microbial production of small medicinal molecules and biologics: from nature to synthetic pathways. *Biotechnology advances* **36,** 2219–2231 (2018).
8. Bar-Even, A., Noor, E., Lewis, N. E. & Milo, R. Design and analysis of synthetic carbon fixation pathways. *Proceedings of the National Academy of Sciences* **107,** 8889–8894 (2010).
9. Rastogi, R. P. & Sinha., R. P. Biotechnological and industrial significance of cyanobacterial secondary metabolites. *Biotechnology advances* **27,** 521 (2009).
10. Qian, Z.-G., Xia, X.-X. & Lee, S. Y. Metabolic engineering of Escherichia coli for the production of cadaverine: a five carbon diamine. *Biotechnology and bioengineering* **108,** 93–103 (2011).
11. Ducat, D. C., Way, J. C. & Silver, P. A. Engineering cyanobacteria to generate high-value products. *Trends in biotechnology* **29,** 95–103 (2011).
12. Von Hertzen, L., Hanski, I. & Haahtela., T. Natural immunity: biodiversity loss and inflammatory diseases are two global megatrends that might be related. *EMBO reports* **12,** 1089 (2011).
13. WMO. *Greenhouse gas levels in atmosphere reach new high* `https://public.wmo.int/en/resources/library/wmo-greenhouse-gas-bulletin` (2018).
14. Bar-Even, A., Flamholz, A., Noor, E. & Milo, R. Thermodynamic constraints shape the structure of carbon fixation pathways. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1817,** 1646–1659 (2012).

15. Huber, H. *et al.* A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*. *Proceedings of the National Academy of Sciences* **105,** 7851–7856 (2008).

16. Kono, T. *et al.* A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nature communications* **8,** 14007 (2017).

17. Orita, I. *et al.* The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon *Thermococcus kodakaraensis*. *Journal of bacteriology* **188,** 4698–4704 (2006).

18. Sánchez-Andrea, I. *et al.* The reductive glycine pathway allows autotrophic growth of *Desulfovibrio desulfuricans*. *Nature communications* **11,** 1–12 (2020).

19. Berg, I. A. Ecological aspects of distribution of different autotrophic CO_2 fixation pathways. *Applied and Environmental Microbiology* (2011).

20. Yishai, O., Bouzon, M., Döring, V. & Bar-Even, A. In vivo assimilation of one-carbon via a synthetic reductive glycine pathway in *Escherichia coli*. *ACS synthetic biology* (2018).

21. Bassham, J. A. & Calvin, M. in *Die CO_2-Assimilation/The Assimilation of Carbon Dioxide* 884–922 (Springer, 1960).

22. Braakman, R. & Smith, E. The emergence and early evolution of biological carbon-fixation. *PLoS computational biology* **8,** e1002455 (2012).

23. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences* **115,** 6506–6511 (2018).

24. Raven, J. A. Rubisco: still the most abundant protein of Earth? *New Phytologist* **198,** 1–3 (2013).

25. Bar-Even, A. *et al.* The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50,** 4402–4410 (2011).

26. Orr, D. J. *et al.* Engineering photosynthesis: progress and perspectives. *F1000Research* **6** (2017).

27. Wilson, R. H., Martin-Avila, E., Conlan, C. & Whitney, S. M. An improved *Escherichia coli* screen for Rubisco identifies a protein–protein interface that can enhance CO_2-fixation kinetics. *Journal of Biological Chemistry* **293,** 18–27 (2018).

28. Erb, T. J. & Zarzycki, J. A short history of RubisCO: the rise and fall (?) of Nature's predominant CO_2 fixing enzyme. *Current opinion in biotechnology* **49,** 100–107 (2018).

29. Nisbet, E. *et al.* The age of Rubisco: the evolution of oxygenic photosynthesis. *Geobiology* **5,** 311–335 (2007).

30. Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E. & Scott, S. S. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *Journal of experimental botany* **59,** 1515–1524 (2008).

31. Cegelski, L. & Schaefer, J. NMR determination of photorespiration in intact leaves using in vivo $13CO_2$ labeling. *Journal of Magnetic Resonance* **178,** 1–10 (2006).

32. Young, J. D., Shastri, A. A., Stephanopoulos, G. & Morgan, J. A. Mapping photoautotrophic metabolism with isotopically nonstationary 13C flux analysis. *Metabolic engineering* **13,** 656–665 (2011).

33. Peterhansel, C. & Maurino, V. G. Photorespiration redesigned. *Plant physiology* **155,** 49–55 (2011).

34. Shi, X. & Bloom, A. *Photorespiration: The Futile Cycle?* 2021.

35. Raines, C. A. The Calvin cycle revisited. *Photosynthesis research* **75,** 1–10 (2003).

36. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature chemical biology* **5,** 593–599 (2009).

37. Nakahigashi, K. *et al.* Systematic phenome analysis of Escherichia coli multiple-knockout mutants reveals hidden reactions in central carbon metabolism. *Molecular systems biology* **5,** 306 (2009).

38. Tamoi, M., Miyazaki, T., Fukamizo, T. & Shigeoka, S. The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD (H)/NADP (H) ratio under light/dark conditions. *The Plant Journal* **42,** 504–513 (2005).

39. Long, C. P., Gonzalez, J. E., Sandoval, N. R. & Antoniewicz, M. R. Characterization of physiological responses to 22 gene knockouts in *Escherichia coli* central carbon metabolism. *Metabolic engineering* **37,** 102–113 (2016).

40. Michelet, L. *et al.* Redox regulation of the Calvin–Benson cycle: something old, something new. *Frontiers in plant science* **4,** 470 (2013).

41. Gontero, B. & Maberly, S. C. An intrinsically disordered protein, CP12: jack of all trades and master of the Calvin cycle. *Biochemical Society Transactions* **40,** 995–999 (2012).

42. Plaxton, W. & McManus, M. T. *Annual plant reviews, control of primary metabolism in plants* (John Wiley & Sons, 2008).

43. Bowien, B. & Kusian, B. Genetics and control of CO_2 assimilation in the chemoautotroph *Ralstonia eutropha*. *Archives of microbiology* **178,** 85–93 (2002).

44. Shimada, T., Fujita, N., Yamamoto, K. & Ishihama, A. Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. *PloS one* **6,** e20081 (2011).

45. Ramseier, T. *Cra* and the control of carbon flux via metabolic pathways. *Research in microbiology* **147,** 489–493 (1996).

46. Pandi, K., Chauhan, A. S., Khan, W. H. & Rathore, A. S. Phosphate starvation controls lactose metabolism to produce recombinant protein in *Escherichia coli*. *Applied Microbiology and Biotechnology* **104,** 9707–9718 (2020).

47. Shimada, T., Fujita, N., Maeda, M. & Ishihama, A. Systematic search for the Cra-binding promoters using genomic SELEX system. *Genes to Cells* **10,** 907–918 (2005).

48. Harinarayanan, R., Murphy, H. & Cashel, M. Synthetic growth phenotypes of Escherichia coli lacking ppGpp and transketolase A (*tktA*) are due to ppGpp-mediated transcriptional regulation of *tktB*. *Molecular microbiology* **69,** 882–894 (2008).

49. Avison, M. B., Horton, R. E., Walsh, T. R. & Bennett, P. M. *Escherichia coli* CreBC is a global regulator of gene expression that responds to growth in minimal media. *Journal of Biological Chemistry* (2001).

50. Sørensen, K. I. & Hove-Jensen, B. Ribose catabolism of *Escherichia coli*: characterization of the *rpiB* gene encoding ribose phosphate isomerase B and of the rpiR gene, which is involved in regulation of *rpiB* expression. *Journal of bacteriology* **178,** 1003–1011 (1996).

51. Rosgaard, L., de Porcellinis, A. J., Jacobsen, J. H., Frigaard, N.-U. & Sakuragi, Y. Bioengineering of carbon fixation, biofuels, and biochemicals in cyanobacteria and plants. *Journal of biotechnology* **162,** 134–147 (2012).

52. Bogorad, I. W., Lin, T.-S. & Liao, J. C. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature* **502,** 693 (2013).

53. Zabaleta, E., Martin, M. V. & Braun, H.-P. A basal carbon concentrating mechanism in plants? *Plant Science* **187,** 97–104 (2012).

54. Raven, J. A., Cockell, C. S. & De La Rocha, C. L. The evolution of inorganic carbon concentrating mechanisms in photosynthesis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363,** 2641–2650 (2008).

55. Mackinder, L. C. *et al.* A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proceedings of the National Academy of Sciences* **113,** 5958–5963 (2016).

56. Osborne, C. P. & Beerling, D. J. Nature's green revolution: the remarkable evolutionary rise of C4 plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361,** 173–194 (2006).

57. Kellogg, E. A. C4 photosynthesis. *Current Biology* **23,** R594–R599 (2013).

58. Lefebvre, S. *et al.* Increased sedoheptulose-1, 7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. *Plant Physiology* **138,** 451–460 (2005).

59. Uematsu, K., Suzuki, N., Iwamae, T., Inui, M. & Yukawa, H. Increased fructose 1, 6-bisphosphate aldolase in plastids enhances growth and photosynthesis of tobacco plants. *Journal of Experimental Botany* **63,** 3001–3009 (2012).

60. Liang, F. & Lindblad, P. Effects of overexpressing photosynthetic carbon flux control enzymes in the cyanobacterium *Synechocystis PCC 6803*. *Metabolic engineering* **38,** 56–64 (2016).

61. Miyagawa, Y., Tamoi, M. & Shigeoka, S. Overexpression of a cyanobacterial fructose-1, 6-/sedoheptulose-1, 7-bisphosphatase in tobacco enhances photosynthesis and growth. *Nature biotechnology* **19,** 965–969 (2001).

62. Ducat, D. C. & Silver, P. A. Improving carbon fixation pathways. *Current opinion in chemical biology* **16,** 337–344 (2012).

63. Kanno, M., Carroll, A. L. & Atsumi, S. Global metabolic rewiring for improved CO_2 fixation and chemical production in cyanobacteria. *Nature communications* **8,** 14724 (2017).

64. Tcherkez, G. G., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences* **103,** 7246–7251 (2006).

65. Savir, Y., Noor, E., Milo, R. & Tlusty, T. Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proceedings of the National Academy of Sciences* **107,** 3475–3480 (2010).

66. Kumar, A., Li, C. & Portis, A. R. *Arabidopsis thaliana* expressing a thermostable chimeric Rubisco activase exhibits enhanced growth and higher rates of photosynthesis at moderately high temperatures. *Photosynthesis research* **100,** 143–153 (2009).

67. Shih, P. M. *et al.* Biochemical characterization of predicted Precambrian RuBisCO. *Nature communications* **7,** 10382 (2016).

68. Satagopan, S., Sun, Y., Parquette, J. R. & Tabita, F. R. Synthetic CO_2-fixation enzyme cascades immobilized on self-assembled nanostructures that enhance CO_2/O_2 selectivity of RubisCO. *Biotechnology for biofuels* **10,** 175 (2017).

69. Yang, B. *et al.* Genetic engineering of the Calvin cycle toward enhanced photosynthetic CO_2 fixation in microalgae. *Biotechnology for biofuels* **10,** 1–13 (2017).

70. Occhialini, A., Lin, M. T., Andralojc, P. J., Hanson, M. R. & Parry, M. A. Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO_2. *The Plant Journal* **85,** 148–160 (2016).

71. Li, Z. *et al.* Engineering the Calvin-Benson-Bassham cycle and hydrogen utilization pathway of *Ralstonia eutropha* for improved autotrophic growth and polyhydroxybutyrate production. *Microbial cell factories* **19,** 1–9 (2020).

72. Gong, F. *et al.* Quantitative analysis of an engineered CO_2-fixing Escherichia coli reveals great potential of heterotrophic CO_2 fixation. *Biotechnology for biofuels* **8,** 86 (2015).

73. Bonacci, W. *et al.* Modularity of a carbon-fixing protein organelle. *Proceedings of the National Academy of Sciences* **109,** 478–483 (2012).

74. Flamholz, A. I. *et al.* Functional reconstitution of a bacterial CO_2 concentrating mechanism in *Escherichia coli*. *Elife* **9,** e59882 (2020).

75. Hanson, M. R., Lin, M. T., Carmo-Silva, A. E. & Parry, M. A. Towards engineering carboxysomes into C3 plants. *The Plant Journal* **87,** 38–50 (2016).

76. Lieman-Hurwitz, J., Rachmilevitch, S., Mittler, R., Marcus, Y. & Kaplan, A. Enhanced photosynthesis and growth of transgenic plants that express *ictB*, a gene involved in HCO3- accumulation in cyanobacteria. *Plant Biotechnology Journal* **1,** 43–50 (2003).

77. Atkinson, N. *et al.* Introducing an algal carbon-concentrating mechanism into higher plants: location and incorporation of key components. *Plant biotechnology journal* **14,** 1302–1315 (2016).

78. Long, B. M. *et al.* Carboxysome encapsulation of the CO_2-fixing enzyme Rubisco in tobacco chloroplasts. *Nature communications* **9,** 1–14 (2018).

79. Sage, R. F., Sage, T. L., Kocacinar, F., *et al.* Photorespiration and the evolution of C4 photosynthesis. *Annual review of plant biology* **63,** 19–47 (2012).

80. Schlüter, U. & Weber, A. P. The road to C4 photosynthesis: evolution of a complex trait via intermediary states. *Plant and Cell Physiology* **57,** 881–889 (2016).

81. Yang, X. *et al.* A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytologist* **207,** 491–504 (2015).

82. Ort, D. R. *et al.* Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the national academy of sciences* **112,** 8529–8536 (2015).

83. Claassens, N. J., Sousa, D. Z., dos Santos, V. A. M., de Vos, W. M. & van der Oost, J. Harnessing the power of microbial autotrophy. *Nature Reviews Microbiology* **14,** 692 (2016).

84. Schwander, T., von Borzyskowski, L. S., Burgener, S., Cortina, N. S. & Erb, T. J. A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* **354,** 900–904 (2016).

85. Satanowski, A. *et al.* Awakening a latent carbon fixation cycle in *Escherichia coli*. *Nature communications* **11,** 1–14 (2020).

86. Kim, S. *et al*. Growth of E. coli on formate and methanol via the reductive glycine pathway. *Nature chemical biology* **16,** 538–545 (2020).

87. D Mattozzi, M., Ziesack, M., Voges, M. J., Silver, P. A. & Way, J. C. Expression of the sub-pathways of the *Chloroflexus aurantiacus* 3-hydroxypropionate carbon fixation bicycle in *E. coli*: Toward horizontal transfer of autotrophic growth. *Metabolic engineering* **16,** 130–139 (2013).

88. Weber, A. P. & Bar-Even, A. Update: Improving the efficiency of photosynthetic carbon reactions. *Plant Physiology* **179,** 803–812 (2019).

89. Kebeish, R. *et al.* Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. *Nature biotechnology* **25,** 593 (2007).

90. Maier, A. *et al.* Transgenic introduction of a glycolate oxidative cycle into *A. thaliana* chloroplasts leads to growth improvement. *Frontiers in Plant Science* **3,** 38 (2012).

91. Scheffen, M. *et al.* A new-to-nature carboxylation module to improve natural and synthetic CO_2 fixation. *Nature Catalysis* **4,** 105–115 (2021).

92. Trudeau, D. L. *et al.* Design and in vitro realization of carbon-conserving photorespiration. *Proceedings of the National Academy of Sciences* **115,** E11455–E11464 (2018).

93. Shih, P. M., Zarzycki, J., Niyogi, K. K. & Kerfeld, C. A. Introduction of a synthetic CO_2-fixing photorespiratory bypass into a cyanobacterium. *Journal of Biological Chemistry,* jbc–C113 (2014).

94. Antonovsky, N. *et al.* Sugar synthesis from CO_2 in Escherichia coli. *Cell* **166,** 115–125 (2016).

95. Herz, E. *et al.* The genetic basis for the adaptation of *E. coli* to sugar synthesis from CO_2. *Nature communications* **8,** 1705 (2017).

96. Gleizer, S. *et al.* Conversion of *Escherichia coli* to generate all biomass carbon from CO_2. *Cell* **179,** 1255–1263 (2019).

97. Liu, M. *et al.* Global Transcriptional Programs Reveal a Carbon Source Foraging Strategy by *Escherichia coli*. *Journal of Biological Chemistry* **280,** 15921–15927 (2005).

98. Von Borzyskowski, L. S. *et al.* An engineered Calvin-Benson-Bassham cycle for carbon dioxide fixation in *Methylobacterium extorquens AM1*. *Metabolic engineering* **47,** 423–433 (2018).

99. Gassler, T. *et al.* The industrial yeast Pichia pastoris is converted from a heterotroph into an autotroph capable of growth on CO_2. *Nature biotechnology* **38,** 210–216 (2020).

100. Chi, A. & Kemp, R. G. The primordial high energy compound: ATP or inorganic pyrophosphate? *Journal of Biological Chemistry* **275,** 35677–35679 (2000).

101. Bielen, A. A. *et al.* Pyrophosphate as a central energy carrier in the hydrogen-producing extremely thermophilic *Caldicellulosiruptor saccharolyticus*. *FEMS microbiology letters* **307,** 48–54 (2010).

102. Reshetnikov, A. S. *et al.* Characterization of the pyrophosphate-dependent 6-phosphofructokinase from *Methylococcus capsulatus Bath*. *FEMS microbiology letters* **288,** 202–210 (2008).

103. Ward, N. *et al.* Genomic insights into methanotrophy: the complete genome sequence of *Methylococcus capsulatus (Bath)*. *PLoS biology* **2,** e303 (2004).

104. Coté, R. J. Aseptic technique for cell culture. *Current protocols in cell biology*, 1–3 (1998).

105. Guyer, M., Reed, R., Steitz, J. & Low, K. *Identification of a sex-factor-affinity site in E. coli as $\gamma\delta$* in *Cold Spring Harbor symposia on quantitative biology* **45** (1981), 135–140.

106. Baba, T. *et al.* Construction of *Escherichia coli K-12* in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology* **2,** 2006–0008 (2006).

107. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli K-12* using PCR products. *Proceedings of the National Academy of Sciences* **97,** 6640–6645 (2000).

108. Baba, T. *et al.* Construction of *Escherichia coli K-12* in-frame, single-gene knockout mutants: the Keio collection. *molecular systems biology* **2** (2006).

109. Salamov, V. S. A. & Solovyevand, A. Automatic annotation of microbial genomes and metagenomic sequences. *Metagenomics and its applications in agriculture, biomedicine and environmental studies,* 61–78 (2011).

110. Lesnik, E. A. *et al.* Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic acids research* **29,** 3583–3594 (2001).

111. Salis, H. M. in *Methods in enzymology* 19–42 (Elsevier, 2011).

112. Boratyn, G. M. *et al.* BLAST: a more efficient report with usability improvements. *Nucleic acids research* **41,** W29–W33 (2013).

113. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Science* **27,** 135–145 (2018).

114. Okonechnikov, K., Golosova, O., Fursov, M. & Team, U. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28,** 1166–1167 (2012).

115. Benchling, I. *Benchling* June 5, 2022. https://benchling.com.

116. Flamholz, A., Noor, E., Bar-Even, A. & Milo, R. eQuilibrator - the biochemical thermodynamics calculator. *Nucleic Acids Res* (2012).

117. Hoops, S. *et al.* COPASI-a complex pathway simulator. *Bioinformatics* **22,** 3067–3074 (2006).

118. Wiśniewski, J. R. & Rakus, D. Quantitative analysis of the Escherichia coli proteome. *Data in brief* **1,** 7–11 (2014).

119. Millard, P., Smallbone, K. & Mendes, P. Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in *Escherichia coli*. *PLoS computational biology* **13,** e1005396 (2017).

120. Wiese, R. & Mazein, A. *Krayon4SBGN* version 1.0.1. Sept. 13, 2018. https://github.com/wiese42/krayon4sbgn.

121. Foundation, T. D. *LibreOffice Calc* version 6.2.3.2(x64). Apr. 12, 2019. https://www.libreoffice.org.

122. The GIMP Development Team. *GIMP* version 2.10.12. June 12, 2019. https://www.gimp.org.

123. Lamport, L. *LaTeX, a document preparation system* (1986).

124. Van der Zander, B. *TexStudio* version 4.2.3. Apr. 17, 2022. https://www.texstudio.org/.

125. Villaverde, A., Benito, A., Viaplana, E. & Cubarsi, R. Fine regulation of cI857-controlled gene expression in continuous culture of recombinant *Escherichia coli* by temperature. *Applied and environmental microbiology* **59,** 3485–3487 (1993).

126. Zhu, X.-D. & Sadowski, P. D. Cleavage-dependent Ligation by the FLP Recombinase: Characterization of a mutant FLP protein with an alteration in a catalytic amino acid. *Journal of Biological Chemistry* **270,** 23044–23054 (1995).

127. Lee, J., Hiibel, S., Reardon, K. & Wood, T. Identification of stress-related proteins in *Escherichia coli* using the pollutant cis-dichloroethylene. *Journal of applied microbiology* **108,** 2088–2102 (2010).

128. Wielgoss, S. *et al.* Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3: Genes— Genomes— Genetics* **1,** 183–186 (2011).

129. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proceedings of the National Academy of Sciences* **109,** E2774–E2783 (2012).

130. Cheah, U. E., Weigand, W. A. & Stark, B. C. Effects of recombinant plasmid size on cellular processes in *Escherichia coli*. *Plasmid* **18,** 127–134 (1987).

131. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *Journal of molecular biology* **166,** 557–580 (1983).

132. Goodall, E. C. *et al.* The essential genome of *Escherichia coli K-12*. *MBio* **9,** e02096–17 (2018).

133. Wadano, A., Nishikawa, K., Hirahashi, T., Satoh, R. & Iwaki., T. Reaction mechanism of phosphoribulokinase from a cyanobacterium, *Synechococcus PCC7942. Photosynthesis research* **56,** 27–33 (1998).

134. Vainonen, Y. P. *et al.* Active dimeric form of inorganic pyrophosphatase from *Escherichia coli*. *Biochemistry (Moscow)* **68,** 1195–1199 (2003).

135. Zelcbuch, L. *et al.* Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic acids research* **41,** e98–e98 (2013).

136. Kapyla, J. *et al.* Effect of D97E substitution on the kinetic and thermodynamic properties of *Escherichia coli* inorganic pyrophosphatase. *Biochemistry* **34,** 792–800 (1995).

137. Valdez-Cruz, N. A., Caspeta, L., Pérez, N. O., Ramırez, O. T. & Trujillo-Roldán, M. A. Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and/or pR promoters. *Microbial cell factories* **9,** 1–16 (2010).

138. Lindner, F. & Diepold, A. Optogenetics in bacteria-applications and opportunities. *FEMS Microbiology Reviews* **46,** fuab055 (2022).

139. Pérez-Castiñeira, J. R., López-Marqués, R. L., Villalba, J. M., Losada, M. & Serrano, A. Functional complementation of yeast cytosolic pyrophosphatase by bacterial and plant H+-translocating pyrophosphatases. *Proceedings of the National Academy of Sciences* **99,** 15914–15919 (2002).

140. Park, J. O. *et al.* Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nature chemical biology* **12,** 482–489 (2016).

141. Ishii, N. *et al.* Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* **316,** 593–597 (2007).

142. Ishihama, Y. *et al.* Protein abundance of the *Escherichia coli cytosol*. *BMC genomics* **9,** 102 (2008).

143. Held, D., Yaeger, K. & Novy, R. New coexpression vectors for expanded compatibilities in E. coli. *InNovations* **18** (2003).

144. Schmidt, P. P., Travers, F. & Barman, T. Transient and equilibrium kinetic studies on yeast 3-phosphoglycerate kinase. Evidence that an intermediate containing 1, 3-bisphosphoglycerate accumulates in the steady state. *Biochemistry* **34,** 824–832 (1995).

145. Soukri, A. *et al.* role of the histidine 176 residue in glyceraldehyde-3-phosphate dehydrogenase as probed by site-directed mutagenesis. *Biochemistry* **28,** 2586–2592 (1989).

146. Lambeir, A.-M. *et al.* The cytosolic and glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma brucei*: Kinetic properties and comparison with homologous enzymes. *European journal of biochemistry* **198,** 429–435 (1991).

147. Babul, J., Clifton, D., Kretschmer, M. & Fraenkel, D. G. Glucose metabolism in *Escherichia coli* and the effect of increased amount of aldolase. *Biochemistry* **32,** 4685–4692 (1993).

148. Alvarez, M. *et al.* Triose-phosphate isomerase (tim) of the psychrophilic bacteriumvibrio marinus: Kinetic and structural properties. *Journal of Biological Chemistry* **273,** 2199–2206 (1998).

149. Sprenger, G. A., Schörken, U., Sprenger, G. & Sahm., H. Transaldolase B of *Escherichia coli K-12*: Cloning of Its Gene, *talB*, and Characterization of the Enzyme from Recombinant Strains. *Journal of bacteriology* **177,** 5930–5936 (1995).

150. Zgiby, S., Plater, A., Bates, M., Thomson, G. & Berry, A. A functional role for a flexible loop containing Glu182 in the class II fructose-1, 6-bisphosphate aldolase from *Escherichia coli*. *Journal of molecular biology* **315,** 131–140 (2002).

151. Zgiby, S. M., Thomson, G. J., Qamar, S. & Berry, A. exploring substrate binding and discrimination in fructose 1,6-bisphosphate and tagatose 1,6 bisphosphate aldolases. *European journal of biochemistry* **267,** 1858–1868 (2000).

152. Kelley-Loughnane, N. *et al.* Purification, kinetic studies, and homology model of *Escherichia coli* fructose-1,6-bisphosphatase. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **1594,** 6–16 (2002).

153. Sprenger, G. A., Schörken, U., Sprenger, G. & Sahm, H. Transketolase A of *Escherichia coli K12*. Purification and properties of the enzyme from recombinant strains. *European journal of biochemistry* **230,** 525–532 (1995).

154. Sobota, J. M. & Imlay, J. A. Iron enzyme ribulose-5-phosphate 3-epimerase in *Escherichia coli* is rapidly damaged by hydrogen peroxide but can be protected by manganese. *Proceeding of the National Academy of Sciences* **108,** 5402–5407 (2011).

155. Davis, L., Lee, N. & Glaser., L. On the mechanism of the pentose 5 phosphate epimerases. *Journal of Biological Chemistry* **247,** 5862–5866 (1972).

156. Zhang, R. *et al.* Structure of *Escherichia coli* Ribose-5-Phosphate Isomerase: A Ubiquitous Enzyme of the Pentose Phosphate Pathway and the Calvin Cycle. *Structure* **11,** 31–42 (2003).

157. Liggins, J. R. & Gready, J. E. Putative functional role for the invariant aspartate 263 residue of *Rhodospirillum rubrum* Rubisco. *Biochemistry* **48,** 2226–2236 (2009).

158. Shepelin, D., Machado, D., Nielsen, L. K. & Herrgård, M. J. Benchmarking kinetic models of *Escherichia coli* metabolism. *bioRxiv* (2020).

159. Asplund-Samuelsson, J., Janasch, M. & Hudson, E. P. Thermodynamic analysis of computed pathways integrated into the metabolic networks of *E. coli* and Synechocystis reveals contrasting expansion potential. *Metabolic engineering* **45,** 223–236 (2018).

160. Hädicke, O. & Klamt, S. EColiCore2: a reference network model of the central metabolism of *Escherichia coli* and relationships to its genome-scale parent model. *Scientific Reports* **7,** 1–15 (2017).

161. Kaberdina, A. C., Ruiz-Larrabeiti, O., Lin-Chao, S. & Kaberdin, V. R. Reprogramming of gene expression in *Escherichia coli* cultured on pyruvate versus glucose. *Molecular Genetics and Genomics* **294,** 1359–1371 (2019).

162. Zajdel, T. J., TerAvest, M. A., Rad, B., Ajo-Franklin, C. M. & Maharbiz, M. M. *Probing the dynamics of the proton-motive force in E. coli* in *SENSORS, 2014 IEEE* (2014), 1764–1767.

163. Grainger, D. C., Hurd, D., Harrison, M., Holdstock, J. & Busby, S. J. Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proceedings of the National Academy of Sciences* **102,** 17693–17698 (2005).

164. Zheng, D., Constantinidou, C., Hobman, J. L. & Minchin, S. D. Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. *Nucleic acids research* **32,** 5874–5893 (2004).

165. Walter, J. M., Greenfield, D., Bustamante, C. & Liphardt, J. Light-powering *Escherichia coli* with proteorhodopsin. *Proceedings of the National Academy of Sciences* **104,** 2408–2412 (2007).

166. Song, Y. *et al.* Proteorhodopsin overproduction enhances the long-term viability of *Escherichia coli*. *Applied and environmental microbiology* **86,** e02087–19 (2019).

167. Kim, H. J. *et al.* Biological conversion of methane to methanol through genetic reassembly of native catalytic domains. *Nature Catalysis* **2,** 342–353 (2019).

Appendix A: Covid-19 disclaimer

In an attempt to mitigate the initial Covid-19 outbreak in the UK, the University of Nottingham shut down all on-site research activity on 18/3/2020 and recommended to work from home where possible. As part of a staggered return to work I restarted lab activities at 14/7/2020 in a, due to a maximum capacity of about a quarter of normal, alternating factory shift system for only essential practical work. These shifts allowed 7.5 hours of work per day starting either at 6:30h or ending at 22:00h with half an hour buffer for changes. Initially, 2 persons were allowed per bay (with a pre-covid capacity of 6) and this was increased to 3 at 28/9 to allow working at full daily capacity in over both shifts. After initial returns, the parts of the assigned shifts furthest removed from reasonable office hours were largely unused, other work was heavily minimized as lab usability remained largely disrupted during this time and various benches were still assigned to senior staff without lab work or former co-workers. From 23/2/2021, lab management moved to a system where assigned but unused thirds of the shift could be given up to those with use for them as technicians already had office hours and workspaces overlapping with the shifts since initial reopening. On 27/9/2021, the full pre-Covid working hours and capacity were restored by lab management, any and all raised concerns were put to personal responsibility while management continued to work from home for at least another half year.

The early shift had most of its hours overlapping with regular working hours but in the late shift the majority of work was slotted in what was previously considered as personal time. Because most of the building's services operated on reduced office hours, the late shift had limited access to these. Goods in was only open until 16:00h and closed entire Tuesday afternoon, and highly booked in what little time remained for this shift. The same applied to reception services, completely closed on Thursday and Fridays. Autoclaves were only run once a day, in the morning shift (and finished a maximum of 20 minutes before the end of the shift), and sequencing samples were collected at the end of the morning shift. Returning home at the end of the late shift, carrying the valuable laptop and notes required in the home office through the streets of Nottingham is even counter-advised by the uni-

versity itself. No equipment was supplied to keep these safe during the strongly recommended pedestrian/cycling commute. Weekend work was strictly reduced to activities vital to laboratory operation and pre-booked before Thursday, making it inaccessible for incidental work that couldn't be completed on Friday. It was also divided into time slots, effectively collapsing what little flexible personal time remained of the week if needed. Combined, this made clonings, consecutive incubations and culture preparations planned after Wednesdays impractical for the entire year since reopening the lab.

Because the lab was only accessible for half of the working day, practical work was condensed as much as possible by multi-tasking of pre-planned activities. The 7.5h absolute time limit on any lab work severely hindered certain activities. Reaching sufficient densities of competent cells with slow-growing strains was unreliable and anything involving long or consecutive incubations, such as ligation-transformations on the same day as the DNA isolation, were unavoidably interrupted by freezer storage, further reducing reliability. The other half of the working day was used for planning, calculations, meetings, literature review, writing and other theoretical work. Together they made for intenser and longer working days, nonetheless plagued by a severe drop in productivity compared to pre-Covid work. Interruption of work by shift changes, and the travel between the home office and lab was a consistent disruption to work flow. The difficulty of a reasonably timed lunch time, and the social interactions with peers and colleagues were undoubtedly the greatest loss during this time as it is the irreplaceable source of stress relief, inspiration and information/experience exchange at the core of academic life. The disfigured time schedule around which the working day formed was completely incompatible with any healthy pattern of diet, rest and the few precious social engagements that remained.

When the management of School of Life Science and Synthetic Biology Research Centre showed an incredible eagerness to responsibly resume research during the first lock-down, it was a light through the dark clouds overshadowing this project. Unfortunately the passion for the projects in the lab ended at its reopening for as many users as possible as bench-time that many effective users could have employed was left unutilized for nearly half a year. Though reports varied wildly, I personally haven't been able to

book any bench-time other than that allocated by default since the introduction of a booking system intended for yielding unused time to others. On the busiest days I saw slightly more than half the socially distanced benches available used in a fully booked lab, with occupations further declining with time and increase of outside (and inside!) temperatures. Use of these benches and accompanying desk space was still severely impeded by the large amount of belongings of the pre-covid users of this spaces during this entire time, no effort was made or organised to clear or reallocate the few heavily used spaces. And while lab benches might have been socially distant, nearly all equipment remained placed in narrow hallways or in/around (theoretically) occupied bays, where (fleeting) passes were the rule instead of exemption. The responsible management style was also exemplified by a mandatory logging system to monitor building access for previously unnecessary safety purposes (despite strict access planning) and failing to use it for the monitoring and optimizing of available space.

Personally, I've worked full-time during the first lock-down on *in silico* work, supplemented with literature review, future planning and updating documentation on the project. The reduction in possible lab-productivity was overcompensated for, leading to 10 to 13 hour workdays and included most weekends, for at least half of the first year since UK government recognition of a pandemic. The objectives for the project presented here were never adjusted, were expected to proceed as previous and were even added to with the modelling work started during lock-down.

While regularly receiving reassuring emails from the various levels of bureaucracy that PhD students should firstly take care of themselves and only perform necessary work they were Covid-comfortable with, the practical reality was very different. While having had work-priority over technicians, no easing of responsibilities or expectations, compensation for home office work (electricity, equipment, space, ergonomics) or protection from health related effects of the Covid working conditions was ever seen in reality. Despite spending years of study to achieve highly valued levels of skill and knowledge, PhD candidates are still considered students, and employer responsibilities such duty of care in the UK are practically voluntary, and where at all possible, ignored. PGR management's suggested solution to the continuous loss of research time was to halt my PhD project until the inde-

terminable time it could resume at a pace realistic to satisfy the unchanged expectations from the thesis project, completely ignorant of their own responsibilities in this regard or any logistic and economical implications of this absurd suggestion.

A maximum of 3 months costed extension (a contingency that pre-dates the Covid pandemic) was allowable as compensation for the 4 months lost during university lock-down and the subsequent 13 months of practical work at <50% previous efficiency due to the university's own policies. Eight weeks of funded extension were eventually granted upon appeal of its initial rejection, less than 6 weeks before the original thesis submission date. This time was entirely used to fill gaps in experimental work, as was outlined in its application, now possible due to partially lifted lab restrictions. These required a rewrite of the thesis, still insufficiently finished due to the need to acquire sufficient final experimental data. Despite completing and submitting all forms for an unfunded term extension four weeks prior, the primary supervisor only completed his part on the last possible day. The university's requirement to leave these matters to the last month(s) is devastating in already stressful times and its convoluted bureaucracy detracted even more valuable time from actually progressing thesis work.

Appendix B: pCBB plasmid map

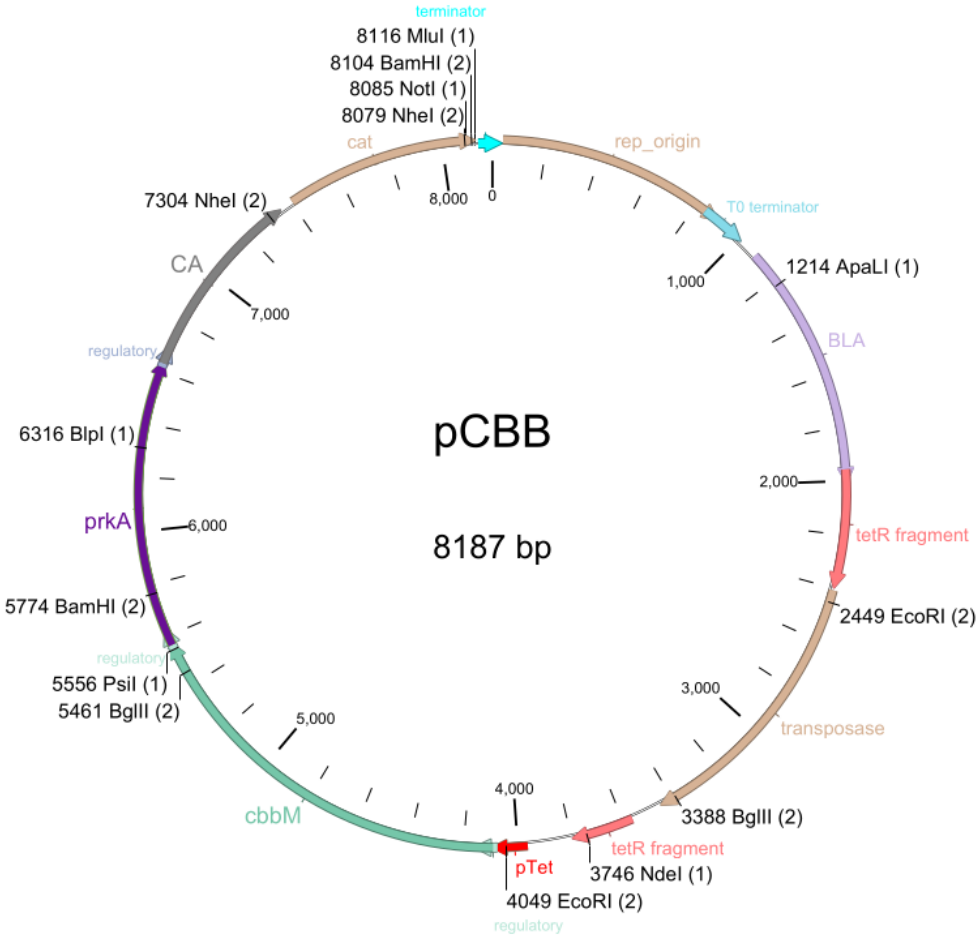

Figure B.1: The pCBB plasmid as obtained in the CBB4 strain, confirmed by sequencing. Annotations obtained from NCBI accession KX077536.1 and added to as appropriate.

Appendix C: Chapter 2, supplementary data

Gen.	Date	week	Shift	Inoculation time offset	Growth time	Evolution time (days)	Sub-cultured uL	OD_{600}	Inoculation OD_{600}	Specific growth rate	Average doubling time	cell divisions	Comments
0	26/02/21	8	late			0							
1	01/03/21	9	early	-5	67	3	1000						
2	03/03/21	9	early	0	48	5	100						
3	08/03/21	10	late	5	125	10	100						
4	12/03/21	10	late	0	96	14	100						
5	15/03/21	11	early	-5	67	17	80						-CO2 control
6	18/03/21	11	early	0	72	20	50						Glycerol stock
7	22/03/21	12	late	5	101	24	50	0.981					
8	25/03/21	12	late	0	72	27	30	1.095	0.0098	0.0655	10.5844	7.2222	-CO2 control
9	29/03/21	13	early	-5	91	31	20	1.2	0.0066	0.0572	12.1125	7.8337	Glycerol stock
10	01/04/21	13	early	0	72	34	5	1.226	0.0048	0.0770	9.0037	9.9349	
11	07/04/21	14	late	5	149	40	10	1.113	0.0012	0.0457	15.1634	9.1053	-CO2 control; large culture for midiprep, sequenced pCBB without mutations
12	12/04/21	15	early	-5	115	45	10	0.96	0.0022	0.0528	13.1392	9.1791	
13	15/04/21	15	early	0	72	48	5	1.249	0.0019	0.0900	7.7043	9.5861	
14	19/04/21	16	late	5	101	52	10	0.875	0.0012	0.0649	10.6852	9.4792	
15	23/04/21	16	late	0	96	56	10	0.801	0.0018	0.0638	10.8618	9.0933	Long weekend, partially grown at 30 degrees
16	04/05/21	18	late	0	264	67	50	0.618	0.0016	0.0226	30.7277	7.0181	
17	07/05/21	18	late	0	72	70	100	0.801	0.0062	0.0676	10.2593	5.2697	Glycerol stock
18	10/05/21	19	early	-5	67	73	50	0.716	0.0160	0.0567	12.2218	6.8057	

19	12/05/21	19	early	0	48	75	25	1.011	0.0072	0.1031	6.7212	7.1461	Glycerol and DMSO stock; used for direct cbb-spec 5 electroporation
20	14/05/21	19	early	0	48	77	10	1.324	0.0051	0.1160	5.9754	8.5767	-CO2 control
21	17/05/21	20	late	5	77	80	10	1.256	0.0026	0.0800	8.6617	9.0419	-CO2 control
22	20/05/21	20	late	0	72	83	5	1.246	0.0025	0.0862	8.0409	9.9773	
23	24/05/21	21	early	-5	91	87	5	1.098	0.0012	0.0745	9.3015	10.1482	
24	27/05/21	21	early	0	72	90	4	1.42	0.0011	0.0995	6.9654	9.9167	
25	01/06/21	22	late	5	125	95	5	1.183	0.0011	0.0556	12.4697	10.2292	
26	04/06/21	22	late	0	72	98	5	0.969	0.0012	0.0932	7.4396	10.2537	Used for direct cbb-spec 2 electroporation
27	07/06/21	23	early	-5	67	101	4	1.105	0.0010	0.1051	6.5976	10.0982	Glycerol and DMSO stock; cultured for plasmid loss and made competent cells
28	11/06/21	23	early	0	96	105	5	0.793	0.0009	0.0708	9.7869	10.4444	
29	14/06/21	24	late	5	77	108	4	1.32	0.0008	0.0963	7.1956	9.5526	
30	17/06/21	24	late	0	72	111	3	1.105	0.0011	0.0966	7.1776	10.9592	
31	21/06/21	25	late	0	96	115	3	1.343	0.0007	0.0793	8.7399	10.4213	used for direct cbb-spec 4 electroporation
32	23/06/21	25	late	0	48	117	1.5	1.286	0.0008	0.1537	4.5112	11.7653	
33	28/06/21	26	early	-5	115	122	1	1.187	0.0004	0.0698	9.9248	12.4033	
34	02/07/21	26	early	0	96	126	1	1.352	0.0002	0.0901	7.6951	12.0999	
35	05/07/21	27	late	5	77	129	1	1.278	0.0003	0.1099	6.3081	12.3689	
36	09/07/21	27	late	0	96	133	1	1.249	0.0003	0.0885	7.8338	12.3208	
37	12/07/21	28	early	-5	67	136	STOP	1.318	0.0002	0.1279	5.4184		DMSO stock

Table C.1: Overview of the available data of the batch-evolution cultures. Due to restricted lab access and an alternating shift system, measurement times could not be consistent and 'inoculation time offset' compensates for this difference in the already unrecorded time of measurement and subsequent inoculation of its sub-culture. Though fundamentally inaccurate, over 90% of the measurements should have fallen within a 1.5-hour window of this estimation. 'Inoculation density' was calculated by dividing its sub-culture volume by the total media volume (5 mL) and multiplying it by the inoculum OD_{600}. Specific growth rate was calculated by subtracting the logarithm of the final culture density from the logarithm of the inoculating density and dividing it by its growth time. The doubling time was obtained by dividing the natural logarithm of 2 by the specific growth rate. Cell divisions were calculated by dividing the culture volume times density by its inoculation volume and density, and taking the base-2 logarithm of it.

Variant Frequency	Type	From	To	CDS	Length	# of intervals	Effect	Position
27.7%	SNP (transversion)	104033	104033		1	1		
47.8%	SNP (transversion)	104041	104041		1	1		
42.2%	SNP (transversion)	104042	104042		1	1		
30.0%	SNP (transversion)	104048	104048		1	1		
34.8%	SNP (transversion)	161051	161051		1	1		
33.3%	SNP (transversion)	161056	161056		1	1		
27.90%	Deletion	244029	244123		95	1		
20.20%	Deletion	244124	244522	*rayT*	399	1	Start Codon Loss	1
20.20%	Deletion	244571	244586	*rayT*	16	1	Frame Shift	448
23.9%	SNP (transversion)	357022	357022	*lacA*	1	1	truncation	550
35.20%	substitution	360176	360179		4	1		
27.50%	substitution	360279	360281		3	1		
30.0%	SNP (transversion)	360283	360283		1	1		
40.30%	substitution	569968	569969		2	1		
62.40%	substitution	569971	569972		2	1		

74.2%	SNP (transversion)	569974	569974		1	1		
74.90%	substitution	569976	569984		9	1		
85.60%	substitution	569986	569988		3	1		
86.10%	substitution	569990	569992		3	1		
88.10%	substitution	569994	569995		2	1		
88.90%	substitution	569997	570000		4	1		
90.20%	substitution	570002	570006		5	1		
91.90%	substitution	570008	570009		2	1		
93.80%	substitution	570012	570013		2	1		
93.60%	substitution	570015	570016		2	1		
94.60%	substitution	570019	570021		3	1		
95.6%	SNP (transversion)	570023	570023		1	1		
96.1%	SNP (transversion)	570029	570029		1	1		
96.4%	SNP (transition)	570032	570032		1	1		
95.7%	substitution	570034	570035		2	1		
96.5%	substitution	570037	570038		2	1		
96.4%	SNP (transversion)	570040	570040		1	1		
94.7%	SNP (transition)	571242	571242	*ompC*	1	1	substitution	1040

93.60%	substitution	571250	571254	*ompC*	5	1	substitution	1028
90.1%	Insertion	571257	571256	*ompC*	0	1	Frame Shift	1026
94.3%	SNP (transversion)	571260	571260	*ompC*	1	1	substitution	1022
92.20%	substitution	571263	571274	*ompC*	11	2	truncation	1008
89.6%	Deletion	571278	571284	*ompC*	6	2	Frame Shift	998
94.0%	SNP (transition)	571288	571288	*ompC*	1	1	substitution	994
92.80%	Deletion	571292	571294	*ompC*	3	1	Frame Shift	988
93.2%	SNP (transition)	571299	571299	*ompC*	1	1	substitution	983
93.0%	SNP (transversion)	571301	571301	*ompC*	1	1	substitution	981
92.9%	SNP (transversion)	571304	571304	*ompC*	1	1	substitution	978
91.40%	substitution	571310	571311	*ompC*	2	1	substitution	971
90.00%	substitution	571313	571315	*ompC*	2	2	substitution	967
86.6%	SNP (transversion)	571327	571327	*ompC*	1	1	substitution	955
82.30%	substitution	571330	571331	*ompC*	2	1	substitution	951
57.00%	substitution	571334	571336	*ompC*	2	2	substitution	946
19.80%	substitution	571338	571340	*ompC*	3	1	substitution	942
20.00%	Deletion	691965	691970	*tRNA-Gln* 5'UTR	6	1		

20.00%	Deletion	692076	692090	*tRNA-Gln* 5′UTR	15	1		
20.0%	Deletion	692092	692094	*tRNA-Gln* 5′UTR	3	1		
34.80%	Deletion	692120	692320	*tRNA-Met, Gln, Gln, Leu*	201	1		
33.30%	Deletion	692321	692395	*tRNA-Gln*	75	1		
20.70%	Deletion	692396	692512	*tRNA-Leu*	117	1		
26.8%	SNP (transversion)	703498	703498		1	1		
21.4%	SNP (transversion)	703519	703519		1	1		
20.00%	Deletion	733623	733632	*ybfD*	10	1	Frame Shift	76
20.00%	Deletion	733637	733683	*ybfD*	47	1	Frame Shift	90
35.70%	Deletion	776010	776176	*tRNA-Lys*	167	1		
36.30%	Deletion	776177	776298	*tRNA-Val*	122	1		
48.00%	Deletion	776299	776426	*tRNA-Lys*	128	1		
56.30%	Deletion	776427	776472		46	1		
54.70%	Deletion	776473	776602	*tRNA-Val*	130	1		
38.00%	Deletion	776603	776776	*tRNA-Lys*	174	1		
42.70%	Deletion	776777	776778		2	1		
55.20%	Deletion	776779	776823		45	1		
47.2%	Deletion	776824	776824		1	1		
34.90%	Deletion	776825	777032	*tRNA-Lys*	208	1		

38.30%	Deletion	777033	777218	*tRNA-Lys/ RtT* sRNA	186	1		
24.70%	Deletion	777219	777264	*RtT* sRNA	46	1		
20.0%	Deletion	777266	777266		1	1		
100.0%	SNP (transition)	781465	781465	*aroG*	1	1	substitution	377
39.1%	SNP (transversion)	910669	910669		1	1		
36.7%	SNP (transversion)	910671	910671		1	1		
29.4%	SNP (transversion)	910683	910683		1	1		
20.0%	Deletion	936871	936872	*dmsA*	2	1	Frame Shift	457
20.0%	Deletion	936892	936894	*dmsA*	3	1	Deletion	478
20.0%	Deletion	936910	936911	*dmsA*	2	1	Frame Shift	496
20.00%	Deletion	936913	936951	*dmsA*	39	1	Deletion	499
20.0%	Deletion	936957	936961	*dmsA*	5	1	Frame Shift	543
29.6%	SNP (transversion)	1092574	1092574		1	1		
20.0%	Deletion	1092658	1092658		1	1		
20.00%	Deletion	1092685	1092754	*RtT* sRNA	70	1		
100.0%	substitution	1254251	1254252	*ychH*	2	1	substitution	5
100.0%	substitution	1254254	1254258	*ychH*	5	1	substitution	8
100.0%	SNP (transversion)	1254260	1254260	*ychH*	1	1	substitution	14
100.0%	substitution	1254262	1254263	*ychH*	2	1	substitution	16

100.0%	substitution	1254265	1254273	*ychH*	9	1	substitution	19
100.0%	SNP (transversion)	1254275	1254275	*ychH*	1	1	substitution	29
100.0%	substitution	1254277	1254284	*ychH*	8	1	substitution	31
100.0%	substitution	1254286	1254294	*ychH*	8	2	substitution	40
100.0%	substitution	1254296	1254303	*ychH*	7	2	substitution	50
100.0%	substitution	1254308	1254310	*ychH*	3	1	substitution	62
100.0%	substitution	1254456	1254458	*ychH*	3	1	substitution	210
100.0%	substitution	1254460	1254461	*ychH*	2	1	substitution	214
100.0%	substitution	1254464	1254473	*ychH*	10	1	substitution	218
85.70%	substitution	1254475	1254482	*ychH*	8	1	substitution	229
100.0%	substitution	1254484	1254486	*ychH*	3	1	substitution	238
100.0%	SNP (transition)	1254489	1254489	*ychH*	1	1	None	243
100.0%	substitution	1254491	1254494	*ychH*	4	1	substitution	245
100.0%	substitution	1254497	1254498	*ychH*	2	1	substitution	251
100.0%	SNP (transversion)	1254500	1254500	*ychH*	1	1	substitution	254
100.0%	substitution	1254503	1254504	*ychH*	2	1	substitution	257
28.10%	Deletion	1282164	1282519	*RtT* sRNA/ *RtT* sRNA	356	1		
21.30%	Deletion	1282699	1282816	*tRNA-Tyr*	118	1		

22.4%	SNP (transversion)	1356089	1356089	*puuD*	1	1	substitution	713
21.7%	SNP (transition)	1390272	1390272		1	1		
21.3%	SNP (transversion)	1390280	1390280		1	1		
21.10%	Insertion	1390282	1390282		1	1		
23.5%	substitution	1390284	1390285		2	1		
23.30%	substitution	1390287	1390292		6	1		
23.0%	substitution	1390295	1390296		2	1		
23.3%	SNP (transition)	1390300	1390300		1	1		
37.80%	Deletion	1460387	1460524	pseudogene *ydbA*	138	1	Deletion	739
38.20%	Deletion	1460525	1460818	pseudogene *ydbA*	294	1	Deletion	877
31.50%	Deletion	1460819	1460974	pseudogene *ydbA*	156	1	Deletion	1171
22.0%	SNP (transition)	1460986	1460986	pseudogene *ydbA*	1	1	None	1338
20.00%	Deletion	1461055	1461089	pseudogene *ydbA*	34	2	Frame Shift	1407
20.00%	Deletion	1461099	1461106	pseudogene *ydbA*	8	1	Frame Shift	1451
20.0%	Deletion	1461110	1461111	pseudogene *ydbA*	2	1	Frame Shift	1462

20.1%	Deletion	1461115	1461115	pseudo-gene *ydbA*	1	1	Frame Shift	1467
20.00%	Deletion	1461117	1461122	pseudo-gene *ydbA*	6	1	Deletion	1469
100.0%	SNP (transition)	1552685	1552685	*ddpD*	1	1	substitution	590
24.2%	SNP (transition)	1662226	1662226	*mlc*	1	1	truncation	596
22.0%	SNP (transversion)	1828729	1828729	*ydjY*	1	1	substitution	357
63.9%	SNP (transition)	2096425	2096425		1	1		
62.2%	SNP (transition)	2096428	2096428	*wbbL*	1	1	substitution	443
61.50%	substitution	2096430	2096433	*wbbL*	4	1	substitution	438
62.3%	SNP (transversion)	2096437	2096437	*wbbL*	1	1	substitution	434
59.30%	substitution	2096439	2096441	*wbbL*	3	1	substitution	430
65.7%	SNP (transition)	2096444	2096444	*wbbL*	1	1	substitution	427
65.0%	SNP (transition)	2096447	2096447	*wbbL*	1	1	substitution	424
62.9%	SNP (transversion)	2096450	2096450	*wbbL*	1	1	substitution	421
62.50%	substitution	2096453	2096456	*wbbL*	4	1	substitution	415
54.00%	substitution	2096458	2096467	*wbbL*	10	1	substitution	404

58.1%	SNP (transition)	2096470	2096470	*wbbL*	1	1	substitution	401
58.5%	SNP (transversion)	2096472	2096472	*wbbL*	1	1	substitution	399
56.0%	SNP (transversion)	2096477	2096477	*wbbL*	1	1	truncation	394
40.4%	SNP (transversion)	2096481	2096481	*wbbL*	1	1	None	390
20.3%	Deletion	2112959	2112959	*wzxC*	1	1	Frame Shift	681
20.00%	Deletion	2112969	2113318	*wzxC*	350	1	Frame Shift	322
20.0%	Deletion	2113329	2113330	*wzxC*	2	1	Frame Shift	310
20.0%	Deletion	2113366	2113368	*wzxC*	3	1	Deletion	272
20.00%	Deletion	2113376	2113390	*wzxC*	15	1	Deletion	250
31.6%	SNP (transversion)	2178908	2178908		1	1		
24.4%	SNP (transversion)	2222669	2222669		1	1		
23.60%	Deletion	2297879	2297991	REP161	113	1		
94.7%	Insertion	2371140	2371139	*menD*	0	1	Frame Shift	1600
100.0%	Deletion	2400020	2400028	*lrhA*	9	1	Deletion	93
26.2%	SNP (transversion)	2463906	2463906	putative gene *gtrS*	1	1	substitution	1297
21.3%	SNP (transversion)	2463923	2463923	putative gene *gtrS*	1	1	substitution	1314
33.3%	SNP (transversion)	2470484	2470484	*dsdC*	1	1	substitution	625
26.80%	substitution	2470486	2470487	*dsdC*	2	1	substitution	622

100.0%	SNP (transversion)	2617489	2617489	*ppk1*	1	1	substitution	1087
29.70%	Deletion	2811142	2811220	*tRNA-Arg*	79	1		
44.60%	Deletion	2811221	2811247		27	1		
50.80%	Deletion	2811248	2811315		68	1		
39.50%	Deletion	2811316	2811411		96	1		
50.40%	Deletion	2811412	2811555	*tRNA-Arg*	144	1		
49.30%	Deletion	2811556	2811662	*tRNA-Arg*	107	1		
43.60%	Deletion	2811663	2811730		68	1		
32.80%	Deletion	2811731	2811825		95	1		
20.60%	Deletion	2811826	2811830		5	1		
20.00%	Deletion	2871152	2871314		163	1		
20.00%	Deletion	2871347	2871640		294	1		
20.0%	Deletion	2871642	2871647		6	1		
20.0%	Deletion	2871652	2871652		1	1		
100.0%	Deletion	2897433	2897493		61	1		
23.7%	SNP (transversion)	2958460	2958460	*thyA*	1	1	substitution	55
97.7%	SNP (transition)	3053538	3053538	*argP*	1	1	substitution	427
20.40%	Deletion	3187902	3188058		157	1		
31.70%	Deletion	3188059	3188103		45	1		
45.70%	Deletion	3188104	3188124		21	1		
40.20%	Deletion	3188125	3188184	*ibsD*	60	1	Start Codon Loss	1
38.90%	Deletion	3188185	3188190		6	1		
41.90%	Deletion	3188191	3188278		88	1		
26.50%	Deletion	3188279	3188435		157	1		
20.90%	Deletion	3188436	3188479		44	1		
20.00%	Deletion	3188526	3188529	*ibsE*	4	1	Frame Shift	27
20.0%	Deletion	3188531	3188532	*ibsE*	2	1	Frame Shift	32
20.0%	Deletion	3188536	3188536	*ibsE*	1	1	Frame Shift	37

60.70%	substitution	3221131	3221132		2	1		
58.00%	substitution	3221139	3221140		2	1		
31.6%	SNP (transversion)	3358884	3358884	*YhcE* family protein	1	1	substitution	341
32.8%	SNP (transversion)	3358889	3358889	*YhcE* family protein	1	1	substitution	346
32.2%	SNP (transition)	3358895	3358895	*YhcE* family protein	1	1	substitution	352
30.00%	substitution	3358898	3358900	*YhcE* family protein	3	1	substitution	355
41.7%	SNP (transversion)	3358903	3358903	*YhcE* family protein	1	1	None	360
48.70%	substitution	3358909	3358911	*YhcE* family protein	3	1	substitution	366
22.40%	Deletion	3416777	3417021	*rrfF, tRNA-Thr*	245	1		
20.3%	SNP (transition)	3417030	3417030	*rrfD*	1	1		
29.3%	SNP (transversion)	3453450	3453450	*gspF*	1	1	substitution	275
100.0%	SNP (transition)	3479942	3479942	*crp*	1	1	substitution	464
95.50%	Insertion	3547466	3547465	*malT*	0	1	Frame Shift	1023
100.0%	SNP (transversion)	3553566	3553566	*glpR*	1	1	substitution	400
20.80%	substitution	3645383	3645385	*yhiS*	2	2	substitution	733

32.60%	substitution	3645389	3645391	*yhiS*	3	1	substitution	739
30.9%	Insertion	3645393	3645392	*yhiS*	0	1	Frame Shift	743
82.4%	SNP (transition)	3646591	3646591	*yhiS*	1	1	None	4
81.8%	Deletion	3646595	3646595	*yhiS*	1	1	Frame Shift	8
82.1%	SNP (transversion)	3646599	3646599	*yhiS*	1	1	substitution	12
81.0%	substitution	3646601	3646604	*yhiS*	4	1	substitution	14
82.1%	SNP (transversion)	3646607	3646607	*yhiS*	1	1	substitution	20
81.5%	substitution	3646610	3646611	*yhiS*	2	1	substitution	23
81.0%	SNP (transition)	3646618	3646618	*yhiS*	1	1	substitution	31
72.8%	SNP (transversion)	3646622	3646622	*yhiS*	1	1	substitution	35
42.40%	substitution	3646627	3646628	*yhiS*	2	1	substitution	40
31.4%	SNP (transversion)	3774382	3774382		1	1		
29.8%	SNP (transversion)	3774516	3774516		1	1		
23.5%	SNP (transversion)	3774517	3774517		1	1		
45.9%	SNP (transversion)	3774518	3774518		1	1		
98.4%	SNP (transversion)	3903148	3903148	*pstC*	1	1	substitution	611

100.0%	SNP (transition)	3956192	3956192	*gppA*	1	1	truncation	1398
21.7%	SNP (transversion)	4008442	4008442	*metE*	1	1	substitution	2030
56.20%	Deletion	4020680	4020777		98	1		
83.3%	SNP (transition)	4020783	4020783		1	1		
20.00%	Deletion	4072886	4072910		25	1		
20.2%	Deletion	4072915	4072915		1	1		
100.0%	SNP (transversion)	4151176	4151176	*fabR*	1	1	substitution	128
100.0%	SNP (transversion)	4154051	4154051	*btuB*	1	1	substitution	485
21.40%	substitution	4161664	4161665	*rrfB*	2	1		
24.5%	SNP (transversion)	4161669	4161669	*rrfB*	1	1		
22.20%	substitution	4161671	4161678	*rrfB*	8	1		
20.00%	Deletion	4281692	4281834	*nrfE*	143	1	Frame Shift	313
20.00%	Deletion	4281860	4281863	*nrfE*	4	1	Frame Shift	481
20.00%	Deletion	4281869	4281890	*nrfE*	22	1	Frame Shift	490
20.0%	Deletion (tandem repeat)	4281901	4281901	*nrfE*	1	1	Frame Shift	522
20.00%	Deletion	4281906	4281917	*nrfE*	12	1	Deletion	527
20.7%	SNP (transversion)	4375498	4375498	*mscM*	1	1	truncation	289
21.9%	SNP (transversion)	4402596	4402596	*yjfM*	1	1	None	393

22.0%	SNP (transversion)	4586580	4586580	*yjjN*	1	1	None	774
20.00%	Deletion	4595895	4595959	*tRNA-Leu*	65	1		
20.00%	Deletion	4595986	4596031	*tRNA-Leu*	46	1		
23.00%	Deletion	4596032	4596128	*tRNA-Leu*	97	1		
20.00%	Deletion	4618950	4619068	*ettA*	119	1	Frame Shift	1272

Table C.2: All new chromosomal polymorphisms found in the 37-generation batch-evolved CBB4 strain, when compared to the BW25113 reference.

appendix E: Chapter 3, supplementary data

Figure D.1: The pMTL71301 plasmid as obtained from Dr. Minyeong Yoo, used for the complementation of *ppa* after appropriate modification.

Assay	Temperature (°C)	Prom.	Miller units	CI cal.	Nat cal.	Norm. CI	Norm. Nat	Norm. val.
3	30	CI	**0.1709**	39.1318	1.6731	0.1709		0.1709
	37	CI	39.1318	39.1318	1.6731	39.1318		39.1318
	39	CI	**19.1884**	39.1318	1.6731	19.1884		19.1884
	37	Nat	1.6731	39.1318	1.6731			
4	35	CI	**8.0357**	24.9960		12.5801		12.5801
	30	CI	**0.1367**	24.9960		0.2140		0.2140
	37	CI	24.9960	24.9960				
5	33	Nat	**0.6982**		1.9963		0.5852	0.5852
	37	Nat	1.9963		1.9963			
7	34	CI	**2.3348**	19.3226		4.7285		4.7285
	37	CI	19.3226	19.3226				

Table D.1: The Miller units calculated for each β-galactosidase assay and their subsequent normalized values against either the natural *lacZ* or *cI857* promoters cultured at 37°C. Bold values are those calibrated for in the assay's calibration to result in normalized values.

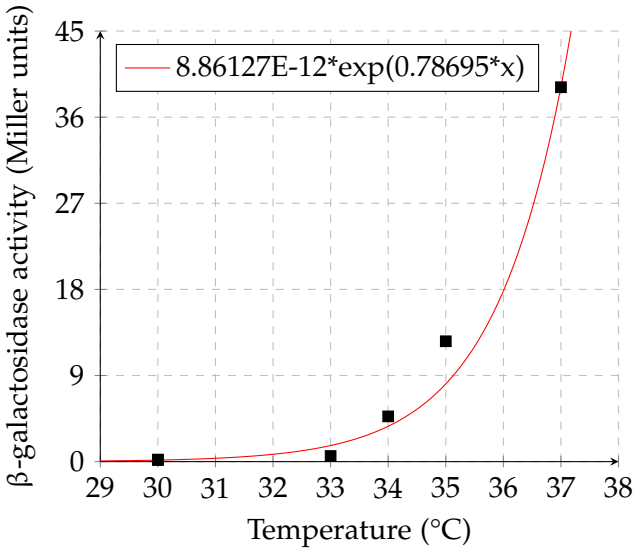

Figure D.2: Plot and trend line (r^2=0.943) of normalized β-galactosidase activity of pCBB-DO_lacZ, up to its observed peak at 37°C. Used for the interpolation of equivalent pCBB-DOt_lacZ activity.

Appendix E: Chapter 4, supplementary data

Reaction	From fold-	To fold-	Initial value	Max.	Min.
acn1			9.7241	11	2
acn2			9.8657	11	2
edd	3.3		0.1114	0.12	0.025
eno	2.5		11.718	12.5	3.5
fba	2	3	21.697	23	5
fum	2		53.341	55	20
gdh	4		8.6657	9.5	1.5
glt	4		57.058	62	10
lpd			0.0684	0.075	0.01
pgi	2		2.3246	2.5	1
mdh	2	3	6.1149	7	1.5
pdh	3		961.70	1050	250
pgi	2		2.3246	2.5	0.8
pyk	2	5	0.7472	0.9	0.1
tal F6P-GAP	4.2		119.99	130	20
tal S7P-E4P			99.999	130	15
tkt F6P-E4P	2.9		40.000	44	10
tkt S7P-R5P			199.97	220	50
tkt Xu5P-GAP			40	44	10

Table E.1: The fold expression range used in all pyruvate-fed models [161], with the model's unmodified V_{max} or k_{cat} value and the range relevant models were optimized with for these affected reactions.

Reaction	From fold-	To fold-	Initial value	Max.	Min.
ACK	2.2		7.23	20	6
ACS	4.3		7.3	33	7
FBP	2		5.7	13	5
ICD	2.6	5.1	2461.97	13500	2200
MAE	3		6.64269	20	6
MQO	5		4.62283	30	4
PPS	2	6	0.01638	0.11	0.014

Table E.2: The fold expression range used in all pyruvate-fed models [161], with the model's unmodified V_{max}or k_{cat} value and the range relevant models were optimized with for these affected reactions.

Reaction	Initial rate	Min.	Max.
RuBisCO	1	0.001	1000
PRK	1	0.001	1000
PPA	1	0.001	1000
PFP-FBP	1	0.001	1000
PFP-PRK	1	0.001	1000
HPP	1	0.001	1000

Table E.3: The range of optimization used in applicable models for genes of unknown kinetic properties, not in the published model, or expressed from plasmid. A great range was given to these enzymes as they are variably expressed or highly efficient and should be excluded as potential limiting factors.

Parameter	Reaction	Min.	Max.	Initial value	Model 7	Model 8	Model 9	Model 10	Model 12
Km ACCOA	Growth	1.00E-04	10	4.94E-02	2.79E-02	2.52E-01	6.77E-01	**1.00E-04**	5.45E-01
Km AKG	Growth	1.00E-03	1.00E+02	5.12E+00	7.76E+00	1.07E+01	8.83E+01	1.18E+00	1.79E+01
Km ATP	Growth	1.00E-06	1.00E+01	4.68E-02	8.54E-01	4.34E+00	5.45E-01	3.28E-04	**1.00E-06**
Km E4P	Growth	1.00E-04	1.00E+02	1.63E+00	5.57E-01	4.05E-01	4.00E+01	1.01E+01	2.23E+00
Km F6P	Growth	1.00E-04	1.00E+01	3.66E-01	1.81E+00	2.35E+00	3.33E+00	**1.00E-04**	3.08E+00
Km G6P	Growth	1.00E-04	1.00E+02	1.21E+00	5.82E+01	1.14E-02	7.38E-03	**1.00E-04**	7.56E+00
Km GAP	Growth	1.00E-04	1.00E+01	2.49E-02	4.23E+00	4.16E-01	4.25E-01	2.11E-01	3.11E-01
Km NAD	Growth	1.00E-04	1.00E+01	2.82E+00	3.12E-03	1.38E-03	4.73E+00	6.25E-01	3.47E+00
Km NADPH	Growth	1.00E-03	1.00E+02	3.60E+00	6.95E-02	3.67E+00	3.31E-01	1.23E-01	2.30E-01
Km OAA	Growth	1.00E-06	1.00E+01	2.48E-02	2.49E-04	**1.00E-06**	1.42E-02	9.97E-03	5.23E-04
Km PEP	Growth	1.00E-05	1.00E+01	4.58E-01	1.68E-02	1.61E-02	3.37E-03	9.34E-01	4.36E-02
Km PGA3	Growth	1.00E-04	1.00E+01	7.65E-02	2.70E+00	2.01E-01	2.64E-01	3.32E+00	2.88E-02
Km PYR	Growth	1.00E-05	1	4.64E-03	2.31E-01	5.43E-03	2.50E-05	**1.00E-05**	**1.00E-05**
Km R5P	Growth	1.00E-06	1.00E+01	2.12E-02	7.79E-01	4.35E-03	4.31E-02	**1.00E-06**	6.30E-03
Vmax	Growth	1.00E-03	1.00E+02	9.74E+00	4.19E+01	7.70E+01	2.83E+01	1.29E+01	3.35E+01
Vmax	PRK	1.00E-02	1.00E+02	1.00E+00	1.07E-01	1.10E+01	n/a	n/a	n/a
Vmax	RuBisCO	1.00E-02	1.00E+02	1.00E+00	4.49E+00	3.76E+01	1.25E+01	8.08E+00	2.71E+00
Vmax	CO_2	1.00E-02	1.00E+03	1.00E+00	4.83E+02	8.18E+01	1.96E+02	4.57E+01	8.71E+01
V	PPA	1.00E-01	3.00E+04	1.00E+00	7.79E+02	6.34E+03	1.00E+00	n/a	2.99E+00
Vmax	PPS	0.015	3	1.64E-02	2.47E+00	2.26E+00	1.65E+00	2.42E+00	1.85E+00
Vmax	ACK	6	20	7.23E+00	1.29E+01	8.30E+00	***6.53E+00***	1.68E+01	***6.00E+00***
Vmax	ACN1	1	10	9.72E+00	4.48E+00	5.96E+00	5.80E+00	1.63E+00	8.14E+00
Vmax	ACN2	1	10	9.87E+00	2.36E+00	3.41E+00	4.86E+00	***1.00E+01***	2.83E+00
Vmax	ACS	7	33	7.23E+00	2.48E+01	7.50E+00	2.08E+01	1.03E+01	1.89E+01
Vmax	EDD	0.025	0.12	1.11E-01	8.88E-02	6.17E-02	5.77E-02	6.97E-02	5.42E-02
Vmax	ENO	3.5	12.5	1.17E+01	1.04E+01	5.96E+00	4.05E+00	3.50E+00	5.56E+00
Kcat	TKT(f6p-e4p)	10	44	4.00E+01	3.44E+01	2.24E+01	1.37E+01	***4.40E+01***	1.20E+01
Kcat	TAL (f6p-gap)	20	140	1.20E+02	4.99E+01	8.19E+01	***1.40E+02***	1.01E+02	1.06E+02
Vmax	FBA	6	23	2.17E+01	6.00E+00	***2.22E+01***	1.29E+01	**6.08E+00**	1.37E+01
Kcat	FBP	5	13	5.70E+00	1.24E+01	*5.59E+00*	n/a	n/a	n/a
Vmax	FUMA	20	55	5.33E+01	2.71E+01	3.61E+01	**5.05E+01**	3.16E+01	***5.37E+01***
Vmax	GDH	1.5	9.5	8.67E+00	7.06E+00	6.62E+00	5.86E+00	5.31E+00	3.76E+00
Vmax	GLT	10	62	5.71E+01	5.80E+01	4.77E+01	5.17E+01	***6.20E+01***	***6.20E+01***
Kcat	ICD	2200	13000	2.46E+03	6.33E+03	4.71E+03	8.58E+03	3.82E+03	1.10E+04
Vmax	LPD	0.01	0.075	6.84E-02	2.68E-02	1.20E-02	2.50E-02	4.55E-02	3.47E-02

Vmax	MAE	6	20	6.64E+00	8.61E+00	9.82E+00	*6.40E+00*	***1.96E+0***1	1.36E+01
Vmax	MDH	1.5	7	6.11E+00	2.07E+00	3.34E+00	2.21E+00	2.42E+00	4.75E+00
Vmax	MQO	4	30	4.62E+00	**3.00E+01**	2.48E+01	1.08E+01	***4.00E+00***	*5.95E+00*
Vmax	PDH	250	1050	9.62E+02	4.41E+02	4.78E+02	4.58E+02	4.22E+02	8.16E+02
Vmax	PGI	0.8	2.5	2.32E+00	1.67E+00	1.75E+00	1.02E+00	**8.06E-01**	1.42E+00
Vmax	PYK	0.1	0.9	7.47E-01	2.93E-01	3.98E-01	3.11E-01	**1.00E-01**	2.66E-01
Kcat	TAL (S7P-E4P)	20	110	1.00E+02	9.39E+01	5.00E+01	7.58E+01	3.31E+01	**8.96E+01**
Kcat	TKT (S7P-R5P)	50	220	2.00E+02	1.72E+02	1.81E+02	1.12E+02	5.50E+01	7.05E+01
Kcat	TKT (Xu5P-GAP)	10	44	4.00E+01	*4.16E+01*	1.95E+01	1.60E+01	3.74E+01	1.97E+01
Vmax	HPP	1.00E-03	1.00E+03	1.00E+00	n/a	1.93E-01	5.37E-01	9.32E-02	8.11E-01
Vmax	PFP-FBP	1.00E-03	1.00E+03	1.00E+00	n/a	3.34E+01	6.08E-01	5.52E+01	3.79E-01
Vmax	PFP-PRK	1.00E-03	1.00E+03	1.00E+00	n/a	3.95E+01	1.51E+01	8.66E-02	1.02E+02
Keq	PGK	5.18E-04	100	1.00E+02	1.90E+00	3.44E+01	6.71E+01	1.77E-01	2.32E+01

Table E.4: Overview of the variables included in the optimization function and the values they had in the main model versions reported. Bold values approached either the minimum of maximum limit set for it within 10%, italicized values changed within the small tolerance allowed for the opposite direction than reported [161]. The final optimized values displayed were obtained for model 7: hemi-autotrophy, model 8: half synthetic hemi-autotrophy, model 9: synthetic hemi-autotrophy, model 10: synthetic hemi-autotrophy with *ppa* knock-out and model 12: synthetic hemi-autotrophy with *pta* knock-out.

COVID19 Impact Statement 2020
For use by doctoral PGR students registered prior to 1 March 2020 with end dates up to and including 30 September 2023

The University of Nottingham aims to support all PGRs to complete their degrees within their period of study, by meeting their Doctoral Outcomes. We recognise, and aim to take into account, personal circumstances that may affect a PGR's ability to achieve this.

This Impact Statement can be used to provide details and evidence of impact for:

- annual review and progression discussions, processes and decisions;
- confirmation of need for an extension to your registered period of study
- the thesis examination.

Please keep a completed copy of this form as you may want to refer to it as evidence of impact in your thesis examination.

We strongly encourage you to discuss the completion of this form with your supervisors. If there are confidential issues that you prefer not to share with your supervisors, you can alternatively discuss how best to complete this form with an appropriate member of PGR support staff such as your DTP/CDT Director or Manager, DTP/CDT Welfare Officer, School Postgraduate Student Advisor, School PGR Director or other member of the Welfare team.

To ensure that you cover the full impact of the COVID-19 pandemic on you and your research **since March 15th 2020**, we advise inclusion of information in all relevant sections of the form. You can be very brief, using note or bullet form. The information can be dynamic, showing how you have responded to the evolving situation, or a summary of outcomes. How you record this is entirely up to you.

We suggest that you could include the following information relating to the exceptional conditions of the ongoing COVID-19 pandemic but you can include any relevant information here:

- your ability to work effectively under Covid-secure conditions, or if you are not in your usual working environment particularly if your working environment is not entirely suitable for effective working;
- any change in ease or amount of access to research settings or facilities, such as archives, field-sites, laboratories, software, or databases;
- any changes in your personal circumstances or environment resulting from remote working, or national restrictions , including those related to:
    - additional caring responsibilities,
    - disability and/or being at higher risk from coronavirus
    - impacts on your supervisory team that have affected your research progress
    - your mental health, and whether you have access to mental health support if needed,
    - any financial impacts, either personal or on the research in progress or planned.
    - any other considerations that should be taken into account, whether these do or do not relate to any protected characteristics.

It is allowable, and advised that you include this form in your thesis for submission to examiners. If you do include this form in the thesis submitted for examination, you may remove or redact any information that you wish to remain confidential.

Please remember that if this form is included in the final completed approved thesis, you should ensure that all confidential or personal information is removed, as the thesis will be available to all through eTheses.

If you also use the information contained in this form for any purpose that requires further information sharing, such as in a case for an extension to time or funding, any information you share that may be used for other than the stated purposes will be anonymised, and all personal information through which anyone could be identified removed. Any information contained on this form that you share with the University that could be used to identify you will not be shared with anyone, including

supervisory teams, for other than the stated purposes, without your permission.

If the information included in this form is used for a case for an extension to study or thesis pending, other than for the confirmation of need for an exceptional Covid additional of period of study, you should also refer to the University's online Policy on Circumstances Affecting Students' Ability to Study and Complete Assessments (under Exceptional Guidance to Extenuating Circumstances Panels), section 16 of the PhD Regulations (see Appendix 2, section 1), relating to existing regulations on circumstances that may or may not be usual grounds for an extension, and any other regulations relating to exceptional Covid-19 extensions.

Background Information – your details			
Family Name:	Hendriks	First Name(s)	Rudolf Martinus Antonius
ID:	4321348	School:	of Life Sciences
Original date thesis due for submission	30/1/2022		

1. DESCRIBING THE IMPACT

(Please complete this section with more detailed information on the main impacts on your and your research)

For example you could write a short clear description of the nature of the impacts or problems that you face/have faced, make making this description as brief, and specific as possible. You could also give more detail on the nature of the impacts on your research progress. If you are in years 2 or 3 of your programme you can write this as either a chronological description, or alternatively you can update the information following discussions with your supervisors as you progress through the programme.

We understand that personal and research impacts will be related, so if it helps you should make that clear where appropriate.

Section 1, additional guidance

The impact **on you**:

Severely disrupted work-life balance due to factory shift system. Instead of decreasing expectations proportionally to availability of university resources, responsibilities were only increased. This led to a minimum average of 10+ daily working hours to compensate in at least half of the first year since first lock-down. Minimized social interactions with colleagues and almost no informal meetings. No compensation for the costs of working from home or provision of its required (ergonomic) equipment. Severely increased stress due to the uncertainties of the project and the nightmarish bureaucracy that came with Covid extensions.

The impact **on your research**:

Lost 4 months of practical work in first lock-down and effectively lost a lot more in the following time due to time restricted access to the lab (this doesn't agree with microbiology). These decreased accessibility hours and shared lab spaces required clearing and cleaning of work, further reducing effective experimental access. Increased bureaucracy. Decreased access to university resources/facilities (sending/receiving goods, autoclavation, inductions, real-time problem-solving, consumable shortages). Most meetings, analyses, calculations and planning had to be conducted outside of lab access hours. No weekend access, or only at pre-scheduled times decreasing the enjoyment of this vital relaxation-time. The primary investigator of this project is a part-time professor and wasn't able to visit Nottingham for nearly 2 years, the weekly meetings generally took away time from the already less productive remaining time.

2. ACTIONS TAKEN TO MINIMISE THE IMPACT

a) How have you tried to mitigate the risk to your project?

Briefly show how you have tried to minimise the impact of the situation on your research activities and progress.

For example,

- have you discussed how to do this with your supervisors? Include information of the people with whom you have discussed your plans for progression.
- have you considered different ways to get the research done, such as changing your research plans to alter the order in which you do different elements?
- have you altered your research design, for example to conduct research online, or using other digital resources?
- what constraints or barriers did you have to try to remove, modify or overcome?
- If you have not been able to alter your plans, why not? how are you managing your progress if this is the case?

Try to show how/whether your work to date is planned to, or already meets some of the University and QAA Doctoral Outcomes, clarify which doctoral outcomes are not currently met and how your plan will enable you to meet these.

Section 2 additional guidance

During first lock-down kinetic modelling was applied to the relevant pathways of the project, proving vital to its progression. To complete this effort, it required a substantial investment of time to complete since the return to lab.
The limited lab time available was maximised, with meetings, analyses, calculations and planning taking place outside of it at the severe detriment of my personal life and mental well-being.
(Un-)funded extensions were pursued. The funded one was entirely spent catching up with experiments after the shift system was abolished. The recording of growth curves for example, was previously impossible to conduct without relying on sheer luck of having access during the exponential phase of an unpredictable culture.
I convinced the head of lab management to adopt a scheduling system for hours unused of other shifts (as there was mixing staff, working normal hours the entire time), though reports varied its implementation didn't allow personal use of it.

b) List the aspects of your research plan that you have managed to achieve or progress during the period of impact.

Further develop and trouble-shoot constructs to test various aspects of the thesis; develop and evaluate control constructs and knock-outs, develop implement and assess new designs, characterize successful strains, conduct assays to confirm regulatory interactions.
Complete the considerable adaptation of the pathway model and its dependencies, then integrate the hypothesized changes with minimal disruption.
Monitor incubating cultures and perform maintenance tasks.

3. NEXT STEPS

Please **list** / bullet point what you plan to do from now, to both progress your research and in order to continue to lessen the impact on your research, now that you are able to access research facilities and can resume the specific activities listed in Section 1, or start to implement the changes to your research plans. The plan should show how you aim to complete your research and thesis as close to your original end date as possible.

For example, what plans do you have to make sure that elements of your research that you have altered can be completed effectively to meet the doctoral outcomes? What, if any, further changes do you need to make? What impact(s) cannot be effectively mitigated and how will you address those?

As and when your research plans change, it is useful to keep a record of such changes. If you have a research plan from before the Covid pandemic, it is useful to include it in the document, and compare to your revised plans are you move forwards in your programme.

Section 3 additional guidance

- I have made every effort to mitigate impact, no further solutions were available or remotely sustainable within the limited time and possibilities made available to me.
- Submit thesis
- Defend thesis

4. EVIDENCE

You do not need to provide any evidence supporting the information contained in the form but if you have any available evidence, then you can keep a list of it here as a reminder.

Section 4 additional guidance

The various communications from Nigel Minton, Alan Burbridge, heads of school/department, PGR and the attached thesis.

5. CONFIDENTIAL INFORMATION

Please use this section to provide any confidential information if you wish to record this as a record for you, and if appropriate your supervisors. You can redact or remove this information if you use this form as evidence of impacts for your examiners.

☒ **I have completed this form after/in discussion with**:
(indicate all those that apply, discussion is strongly encouraged with at least one person)

☐ Primary supervisor/other supervisor ☐ SPSA ☐ School PGR Director ☐ DTP/CDT Director
☐ DTP/CDT Manager ☐ DTP/CDT Welfare Officer ☐ other member of the Welfare Team
☐ Other (please make a note below)

Appendix 1.

University of Nottingham Criteria for award of PhD and other qualifications at Doctoral Level

(i) the creation and interpretation of new knowledge, through original research or other advanced scholarship, of a quality to satisfy peer review, extend the forefront of the discipline, and merit publication;
(ii) a systematic acquisition and understanding of a substantial body of knowledge which is at the forefront of an academic discipline or area of professional practice;
(iii) the general ability to conceptualise, design and implement a project for the generation of new knowledge, applications or understanding at the forefront of the discipline, and to adjust the project design in the light of unforeseen problems;
(iv) a detailed understanding of applicable techniques for research and advanced academic enquiry.

Typically, holders of the qualification will be able to:
(a) make informed judgements on complex issues in specialist fields, often in the absence of complete data, and be able to communicate their ideas and conclusions clearly and effectively to specialist and non-specialist audiences;
(b) continue to undertake pure and/or applied research and development at an advanced level, contributing substantially to the development of new techniques, ideas, or approaches; and will have:
(c) the qualities and transferable skills necessary for employment requiring the exercise of personal responsibility and largely autonomous initiative in complex and unpredictable situations, in professional or equivalent environments.

Appendix 2.

Additional Guidance notes.

What to include:

Section 1, Describing the impact. Please use this form to show how you have planned to adjust the project or use other means to mitigate the impacts of the Covid-19 pandemic, as far as is possible. You should say how you have changed or adjusted your research plans in response to the pandemic, what impacts (if any) you continue to deal with, and if any have been impossible to address.

Please include all research activities that have been impacted by the pandemic, but not those that are attributable to other events. For example, if you had planned a period of research activity at another organisation before or after the pandemic that has had to be cancelled, or postponed and cannot be rescheduled within your registered period of study, this should be recorded, and the way in which you have altered your approach to this part of the research recorded.

Please do not feel that you have to write a large amount in any of the sections of this form. Your record of impact can be brief and to the point but should show the full impact of the pandemic on you and your research for progression purposes and for confirmation of any need for an extension of time of your registered period of study at your final annual review.

This form will continue to be available on the R&I sharepoint and you can use it to record longer-term impacts of COVID-19 on your work over the coming months or years.

Section 2, Action taken. Please make a note the people with whom you have discussed your research plans and what advice and support you have had in adjusting your activities to mitigate any risk to the progress of your research. You are not obliged to consult or discuss the completion of this form with your supervisors, but we encourage you to do so. Include if and how your plans have changed as a result of either these discussions or your own planning.

Please also detail the aspects of your research plans that you have managed to achieve or move forwards under the pandemic conditions, even if you feel that you haven't managed to achieve as much as you originally planned. Please show how your achievements relate to your previous and future research plans.

Section 3, next steps. It's important to plan both how to deal with a current or any emergent situation that disrupts your research, and also how to get back into 'normal' working once you are able to do so. Back-up plans for how you might deal with further lockdowns or restrictions on access to research facilites are particularly important, so that you can quickly react to changing situations. For example, is there a piece of advanced scholarship, such as writing a narrative or systematic review, that you could move to. These plans should include how you will continue to progress your research, under Covid-secure conditions and taking into account any limitations or constraints that these will impose, and if/as things move back to a more 'normal' situation,including how you plan to complete your research and thesis. Your future/revised plans do not need to be complicated, nor in Gantt chart form unless this is a planning method that you already use. A simple table of milestones, deadlines, and outputs is more than sufficient.

If there is anything that is still presenting you with a problem, and that is likely to continue to do so, please record it here. Record information on why this might be an ongoing concern, giving brief information on discussions you have had to try and solve the problem. As you progress, you can update this section with any changes, either ways in which you have solved problems or ongoing difficulties, as appropriate.

Section 4 You do not need to provide evidence but if you have any, you can make a note here so that you are reminded if you subsequently use this form to support your thesis examination.

Privacy and confidentiality: We encourage everyone to discuss the information contained in the form, and its completion with a member of the PGR support staff in the University, particularly with your supervisors. We do however recognise that there may be aspects of this form that you might wish to keep confidential, and so you could alternatively discuss things with

your SPSA, your School PGR Director(s), your DTP/CDT Director, Manager or Welfare Officer, Senior Tutor, mentor or other supportive member of staff.

If the circumstances you describe have an impact on your final thesis assessment, it will probably be necessary to share key information with the examiners/internal assessor. If there is anything that you wish to keep confidential from the examiners please only include it in section 5 of the form. This will enable sharing of the rest of the information in a way that will let the assessors/examiners understand the impact on you and your research while enabling you to remove or redact this confidential information prior to inclusion of the form in your submitted thesis.

For use in progression and thesis assessment: We strongly suggest that you keep a copy of this form updated as you progress through your research and thesis writing. It should be used in annual progression discussions with your supervisors and internal assessor, and we strongly advise you to include a copy in your annual review documents as a record of your challenges and how/if you have managed to solve these. This is particularly important at your final annual review should you require to confirm your need for a Covid-19 extension to your period of registered study at this point. You can also use this form, with any confidential material redacted, for inclusion with your submitted thesis as a record of how you have managed the impact of the COVID pandemic on your achievements during your postgraduate research degree.