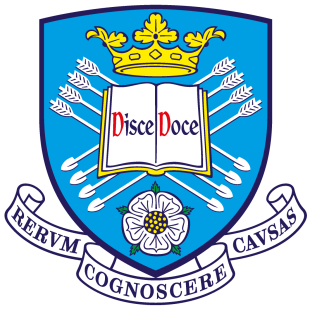


The University Of Sheffield.

Expanding engineering tools for *Cupriavidus necator* H16 to convert waste stream into useful chemicals *via* rational design and evolutionary engineering approaches

By:

Miriam González Villanueva

A thesis submitted for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Engineering

Department of Chemical and Biological Engineering

October 2018

(This page is left blank intentionally)

DECLARATION

This thesis and the work presented in it is a product of original research conducted by me as at the Department of Chemical and Biological Engineering, at The University of Sheffield. All sources of information presented in this work have been accordingly referenced. This thesis has never been previously submitted at this University or any other institution. Parts of this work will be presented elsewhere in the form of scientific publications.

This thesis is written in conformance to the rules of Alternative Format Thesis (Code Of Practice For Research Degree Program 2017-18) of the University of Sheffield.

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

A	Adenine
ACM	Acidomycin or actithiazic acid
ACP	Acyl carrier protein
Amp^r	Ampicilin resistance gene
APS	Ammonium persulphate
ATP	Adenosine Triphosphate
AZ	Azelaic acid
B. subtilis	*Bacillus subtilis*
B. sphaericus	*Bacillus sphaericus*
bp	base pairs
C	Cytosine
C. necator H16	*Cupriavidus necator* H16
Ca^{2+}	Calcium ion
$CaCl_2$	Calcium chloride
Cam^r	Chloramphenicol resistance gene
CBB	Calvin-Benson-Bassham cycle
CCU	Carbon capture utilisation
c.g.d. %	Cell growth decrease %
CDW	Cell dry weight
cm	centimetres
CO_2	Carbon dioxide
CoA	Coenzyme A
DAPA	7,8-diaminopelargonic acid
DD	*Desulfovibrio desulfuricans*
DDI	Distilled De-Ionized
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTB	Dethiobiotin
DV	*Desulfovibrio vulgaris*
E. coli	*Escherichia coli*
ED	Entner-Duodoroff
EDTA	Ethylenediaminetetraacetic acid
egfp	Enhanced green fluorescence protein gene
eGFP	Enhanced green fluorescence protein
Em	Emission coefficient
EMS	Ethyl methanesulfonate
Ex	Excitation coefficient
FRET	Fluorescence resonance energy transfer
Ft	Falcon tube
G	Guanine
Gen^r	Gentamicin resistance gene
glpD	glycerol-3-phophate dehydrogenase gene
GlpD	glycerol-3-phophate dehydrogenase protein
glpF	glycerol facilitator gene
GlpF	glycerol facilitator protein
glpK	glycerol kinase gene
GlpK	glycerol kinase protein

Gly	Glycerol
h	hours
H_2O	Water
Indels	Insertions and deletions
Kan^r	Kanamycin resistance gene
KAPA	7-keto-8-aminopelargonic acid (KAPA)
kb	kilobase
kDa	kilo Dalton
kg	kilogram
KH_2PO_4	Potassium phosphate monobasic
K_2HPO_4	Potassium phosphate dibasic
KOH	Potassium hydroxyde
kV	Kilovolt
L	Litre
lb	libra (pound)
lcl-PHAs	Long-chain-length PHAs
LPS	Lipopolysaccharide
m	mutant
M	Molar
mcl-PHAs	Medium-chain-length PHAs
MCS	Multiple Cloning Site
Mg^{2+}	Magnesium ion
$MgCl_2$	Magnesium chloride
MIC	Minimum inhibitory concentration
min	minutes
mL	millilitres
mm	millimetres
mM	milli-molar
Mp	microtitre plate
ms	milli seconds
MSM	Mineral salts medium
NA	Not applicable
NADH	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NaOH	Sodium hydroxide
NB	Nutrient Broth
N_L	Nitrogen-Limiting condition
nm	nano-meter
nM	nano-molar
NTG	*N*-methyl-*N'*-nitro-*N*-nitrosoguanidine
OD	Optical density
O.N.	Overnight
P(3HB-co- 3HHx)	Poly-3-hydroxybutyrate-co-3-hydroxyhexanoate
P(3HB-co-3HV)	Poly-3- hydroxybutyrate-co-3-hydroxyvalerate
PCR	Polymerase chain reaction
PHA	Polyhydroxyalcanoate
PHB	Poly[(R)-3-hydroxybutyrate] or polyhydroxybutyrate
RBS	Ribosome binding site
Rd	Round

ReA	Restriction enzyme analysis
ReD	Restriction enzyme digestion
rfp	Red fluorescence protein gene
RFP	Red fluorescence protein
RNA	Ribonucleic acid
rpm	revolution per minute
RuBisCO	Ribulose-1,5-bisphosphate carbozylase/oxygenase
s	seconds
SAM	S-adenosyl-L-methionine
scl-PHAs	short-chain length PHAs
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SEM	Standard error of mean
SNPs	Single-nucleotide polymorphism
Suc	Sucrose
t	time
T	Thymine
TCA	Tricarboxylic acid cycle
TEMED	Tetramethylethylenediamine
TVA	5-(2-thienyl)-valeric acid or 5-(2-thienyl)-pentanoic acid
UV	Ultraviolet
v	variant
Vol.	volume
v/v	volume per volume
WGS	Whole genome sequencing
WT	Wild-type
w/v	weight per volume
ºC	degree Celsius
α	alpha
β	beta
μ	Specific growth rate
μm	micrometer
μg	microgram
μL	microlitre
2MC	2-methyl citrate
3MB	3-methyl-1-butanol
(-)ve	Negative control
(+)ve	Positive control

ACKNOWLEDGEMENTS

Just before starting this research someone told me that pursuing a PhD could be a solitary experience; a journey in which most of the time I would be alone with my thoughts. To be honest, sometimes it has been a lonely walk. There have been many moments of solitude looking for answers and thinking about the questions. There have been many hours iterating processes once and again in solitaire at the laboratory and many hours writing on my own. However, I have not been all alone. Far from it.

Many people that have accompanied me throughout this journey. Their presence, their thinking, and their support have nurtured this research experience in one way or another and the time has come to say "thank you".

Firstly, I am very glad to come to the University of Sheffield and join the Chemical and Biological Engineering Department, where I found the perfect combination of PhD supervisor, colleagues, workplace, and environment. I would like to thank my esteemed supervisor Dr Tuck Seng Wong for his invaluable encouragement and guidance. He did not only enable me to carry out my research but also motivated me to keep seeking and challenging me to go further. I appreciate each one of the moments Dr Tuck opened the door to my questions.

Additionally, I would like to thank Dr Kang Lan Tee for her continuous support and influence in the construction of this research. Her acute thinking and timely observations played a great deal on this journey.

I would like to express special gratitude to my beloved friends, lab mates, colleagues, and fellow PhD students – Yomi, Abdul, Zaki, José, Pawel, and Inas – for sharing with me this experience and making the past four years one of the best periods in my life. I want to thank them for helping me to refine ideas and techniques. Their sole presence was, in more than one occasion, enough to spark an idea or simply, to bright the day. Also, I would like to thank my sponsor CONACYT of the Government of Mexico for funding my PhD.

Last but not least, I certainly want to thank my family, a cornerstone of my life, and mainly to my beloved husband Carlos, who always supported me, day and night during my PhD. Their encouragement and unconditional support in the most difficult times helped me to keep going through this academic journey full of learning and satisfaction. Thank you.

ABSTRACT

Introduction:

Cupriavidus necator H16 is a chemolithoautotrophic Gram-negative bacterium that belongs to the class of β-proteobacteria. In recent years, it has attracted the biotechnological attention as a cell factory to produce value-added chemicals. Thus, a need for expanding engineering tools for chemical production in *C. necator* H16 has emerged. In order to use *C. necator* H16 as a cell factory, mainly two different engineering strategies are reviewed in this work: rational design engineering and evolutionary engineering.

Aim:

The overall aim of the study is to engineer *C. necator* H16 with rational design and evolutionary engineering tools to explore the biotechnological potentials of the strain and fine-tune the properties of *C. necator* H16 for chemical production using waste stream as a feedstock.

Thesis content:

In **Chapter 1**, literature review about *C. necator* H16 was compiled to understand the natural metabolism of the strain and to analyse its different biomanufacturing potentials, as well as engineering strategies used for chemical production in *C. necator* H16 and other related microorganisms, afterwards, in **Chapter 2** the development of an optimised transformation method and a synthetic biology toolbox for *C. necator* H16 was developed to design robust strains of *C. necator* H16. Once the synthetic biology toolbox was tailored for *C. necator* H16, in **Chapter 3**, the synthetic biology toolbox was tested for the expression of two clusters of genes as well as possibly obtain bioproducts from the expression of these. Then, in **Chapter 4**, evolutionary engineering tools were optimised to engineer *C. necator* H16, the evolutionary engineering tool used was directed evolution *via* random mutagenesis, where an optimised protocol for random

mutagenesis with a chemical mutagen for *C. necator* H16 was optimised and used to try to understand the biotin biosynthesis pathway of *C. necator* H16, which could furtherly be used for chemical production. In **Chapter 5**, the knowledge and strategies studied in previous chapters and proven to work in *C. necator* H16 supported the approaches selected to construct recombinant and evolved strains of *C. necator* H16 for chemical production (bioplastics) using waste stream (crude glycerol) as a feedstock. **Chapter 6** highlights the general discussions, results, future work and perspectives of this PhD work.

CHAPTER 1

Literature Review:

Engineering *Cupriavidus necator* H16 to use it as a chassis cell for chemical production

Abstract

Cupriavidus necator H16 is a chemolithoautotrophic Gram-negative bacterium that belongs to the class of β-proteobacteria. In recent years, *C. necator* H16 has attracted the biotechnological attention as cell factory to produce value-added chemicals. Hence, there is a need for expanding engineering tools for chemical production in *C. necator* H16. In order to use *C. necator* H16 as a cell factory, mainly two different strategies are reviewed in this **Chapter 1**: Rational design and Evolutionary engineering, which are the most common engineering tools used for chemical production in industry. The current and potential bioproducts that can be biosynthesised in *C. necator* H16 and the different carbon sources -including sustainable and cheap feedstock- that this bacterium can utilise are also discussed in this Chapter.

1 Chapter 1. Literature Review: Engineering *Cupriavidus necator* H16 to use it as a chassis cells for chemical production

1.1 Introduction

Fossil hydrocarbons are used as building blocks to produce a wide range of organic carbon-based chemical products; nevertheless, it has been vastly discussed that we can no longer afford the dependency on fossil hydrocarbons, because they are not only finite but also due to environmental concerns. The biotechnological community has put more efforts to fine-tune strategies to minor environmental burdens. For instance, one of the strategies reinforced during the last decades is the production of biodiesel from vegetable or animal fats which was proposed as a more environmentally-friendly alternative to conventional petroleum diesel production. To make this process entirely environmentally-friendly, the waste that results from biodiesel production should be reused, as the waste stream cannot be disposed of in the environment due to contaminants present in it. One of the main byproducts of biodiesel production is crude glycerol, which has about 50 % purity once it has emerged from the biodiesel process. Although pure glycerol is a value-added chemical for the industry, only a small percentage of biodiesel producers have the equipment to refine crude glycerol to at least its 80 % purity in order to be sold afterwards to glycerol refiners. Since crude glycerol is contaminated with other components such as methanol, free fatty acids, and salts, it cannot be released into the environment; therefore, proper disposal strategies must be considered to make the process sustainable. As reviewed in Volodina *et al.* (2016), there are other industrial byproducts like residual glycerol that contain high carbon concentrations per weight, such is the case for corn, palm, and soybean oils, and other materials rich in sugar such as lignocellulose, molasses, and starch. These residues represent the most promising sustainable feedstocks, and have been extensively studied.

These sort of waste streams have directed scientists to design engineered cells that can function as chemical factories, factories that can contribute to the bio-based

economy by employing waste stream as biological feedstock. Some organisms can naturally use a wide range of carbon sources and waste stream as biological feedstock and convert them into value-added chemicals, although most of the times, their native pathways must be engineered, as the natural utilisation of the substrates and the bioproductivity yield of wild-type cells does not fulfil the expected market demand. Therefore, to design a robust strain capable of targeting the general necessities of the biobased economy, this strain must minimise overall productions costs and maximise its bioproduction to be economically competitive by converting cheap and already available waste stream into a sustainable biological feedstock that acts as a building block for chemical production.

As reviewed in Johnson *et al.*, 2015, the organisms that can use waste stream as biological feedstock range from plants, mammalian or microbial cells, being microbial cells the more widely used platforms due to their simple systems and their ability to use biomass feedstock that allows rapid growth and biosynthesis. Microbial cells have another advantage, which is their relatively easy genetic manipulation since the complete genomic sequence annotation of some bacteria is already available, there is also as well, the possibility of regulating the fermentation conditions to achieve high product titres. Some of the most well studied microbiological systems used worldwide for biomanufacturing are *Escherichia coli*, *Bacillus subtilis*, *Bacillus megaterium*, *Pseudomonas fluorescens*, *Corynebacterium glutamicum* (Witthoff *et al*., 2015), *Saccharomyces cerevisiae*, *Pischia pastoris* and in recent years *Cupriavidus necator* H16 (reviewed in Johnson *et al.*, 2015).

The work of this PhD study was centred on using *C. necator* H16 as a cell factory for biomanufacturing purposes, where rational design and evolutionary engineering strategies were studied and optimised for chemical production. In this **Chapter 1**, the rationales behind of choosing *C. necator* H16 as a candidate for biomanufacturing are given, as well as an overview of rational design and evolutionary engineering tools that have been used to engineer this strain.

1.2 *C. necator* H16: A candidate for chemical production

The soil bacterium *C. necator* H16 (also known as *Ralstonia eutropha* H16) is a Gram-negative bacterium isolated almost 60 years ago in Goettingen, Germany (Schlegel *et al*., 1961a, b; Wilde 1962), it is found in soil and aquatic environments under facultative anaerobic conditions. *C. necator* H16 belongs to the class of β-proteobacteria, and it can use either organic compounds (heterotrophically) or H_2 as sources of energy (autotrophically), it is as well capable of growing under mixotrophic conditions depending on the carbon source supplied (reviewed in Jajesniak *et al.*, 2014), which is of significant interest for biotechnological engineers. The strain can also fix CO_2 through the Calvin-Benson-Bassham (CBB) cycle, where hydrogen is used as the energy source. As reviewed in Reinecke *et al*. (2009), although *C. necator* H16 can grow either in oxygen-limited or anaerobic conditions, it relies primarily on aerobic conditions, where it uses oxygen, nitrate or nitrite as terminal electron acceptors. For heterotrophic growth, it can metabolise some organic compounds ranging from sugars, amino acids, fatty acids and citric acid cycle intermediates. Some of the different carbon fixation pathways where *C. necator* H16 can grow autotrophically and heterotrophically are shown in **Table 1.1**. Apart from its natural ability to use a variety this wide range of substrates, when *C. necator* H16 is grown under unbalanced nutrient conditions, this microorganism produces and accumulates polymers. These characteristics, as well as the possibility of re-routing the carbon accumulated to produce value-added chemicals, makes this strain the perfect candidate for chemical production.

Table 1.1 Carbon fixation pathways of *C. necator* H16

Growth	Carbon fixation pathway	Key fixation reaction
Autotrophic	Calvin-Benson-Bassham cycle	$CO_2 + H_2O \rightarrow$ 2x 3-Phosphoglycerate
Heterotrophic	ED-pathway	$C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ CO_2 + 6\ H_2O$

Adapted from Jajesniak *et al.*, 2014.

In the absence of organic substrates, *C. necator* H16 acts as a chemolithoautotrophic microorganism, using either H_2 or formate as an energy

source and fixes CO_2 *via* the CBB cycle. This fact tags *C. necator* H16 as an excellent platform for CO_2 assimilation with the subsequent production of value-added chemicals such as bioplastics, bioalcohols or biodiesel, although it must be considered that using CO_2 as a carbon substrate implies scientific and technological challenges at industrial-scale production. As reviewed in Jajesniak *et al.* (2014), carbon capture utilisation (CCU) using *C. necator* H16 as a biological system is regarded as a green platform, due to it could contribute to the reduction of CO_2 emissions resulting mainly from petroleum-based refinery processes.

Some approaches have been developed to improve the CO_2 assimilation in *C. necator* H16; strains 25-1 and HB1 were constructed by the mutation or deletion of the *can* gen, which encodes for one of the four genes for carbonic anhydrases β-Can (Ahrens *et al*., 1972; Kusian *et al*., 2002; Bowien *et al*., 2002; Gai *et al*., 2014), although still some challenges must be overcome, such as the constraint that the RuBisCO has. Even when RuBisCO is the key enzyme of the CBB cycle, it is characterised with low catalytic activity, due to it catalyses both carboxylation and oxygenation (reviewed in Tcherkez *et al*., 2006; in Schrag, 2007; in Chu, 2009). Possible options to solve these issues could include leveraging on CO_2-capture ability of carbonic anhydrases, introduce and engineer RuBisCO with an enhanced catalytic activity and specificity which can be achieved with evolutionary engineering, overexpress RuBisCO or apply carbon-capture enzymes designed computationally with the aid of metabolic engineering (Atsumi *et al*., 2009; reviewed in Ducat *et al*., 2011; reviewed in Drummond *et al*., 2012; reviewed in Johnson *et al*., 2015).

Apart from its attractive metabolism, another key milestone that converts *C. necator* H16 a perfect chassis for biomanufacturing purposes is its unravelled genome. The genome of *C. necator* H16 comprises three replicons: chromosome 1 (4.1Mb), chromosome 2 (2.9 Mb), and the megaplasmid pHG1 (452,156 bp). The sequence of the megaplasmid pHG1 was reported in 2003 (Schwartz *et al*., 2003), while its whole genome -chromosome 1 and chromosome 2- was published in 2006 (Pohlmann *et al*., 2006). Chromosome 1 has 3,651 coding sequences, which comprises essential genes and cell functions, such as DNA replication,

transcription, and translation, as well as ribosomal proteins. On the other hand, chromosome 2 and the megaplasmid pHG1 contain only a few essential genes. Chromosome 1 and 2 encode for 6,116 putative genes, where chromosome 2 is responsible for the central steps of the 2-keto-3-deoxy-6-phosphogluconate (KDPG) sugar and sugar acid catabolism, and for the utilisation of alternative nitrogen sources as well as for the decomposition of aromatic compounds (Pohlmann *et al*., 2006).

Figure 1.1 shows some of the key metabolic pathways of *C. necator* H16. The regulation of some metabolic genes from these pathways has enabled the enhancement of the production of different value-added chemicals using *C. necator* H16 as a cell factory when grown under specific environmental conditions (Park *et al*., 2011). For instance, when *C. necator* H16 is grown under chemolithoautotrophic conditions, the Calvin-Benson-Bassham (CBB) cycle genes –encoded by the *cbb* operon- are induced for its expression; the same case is observed when plant oils such as oleate are used as substrates, where when used as feedstock, the genes of the *β*-oxidation cycle are up-regulated (Brigham *et al*., 2010).

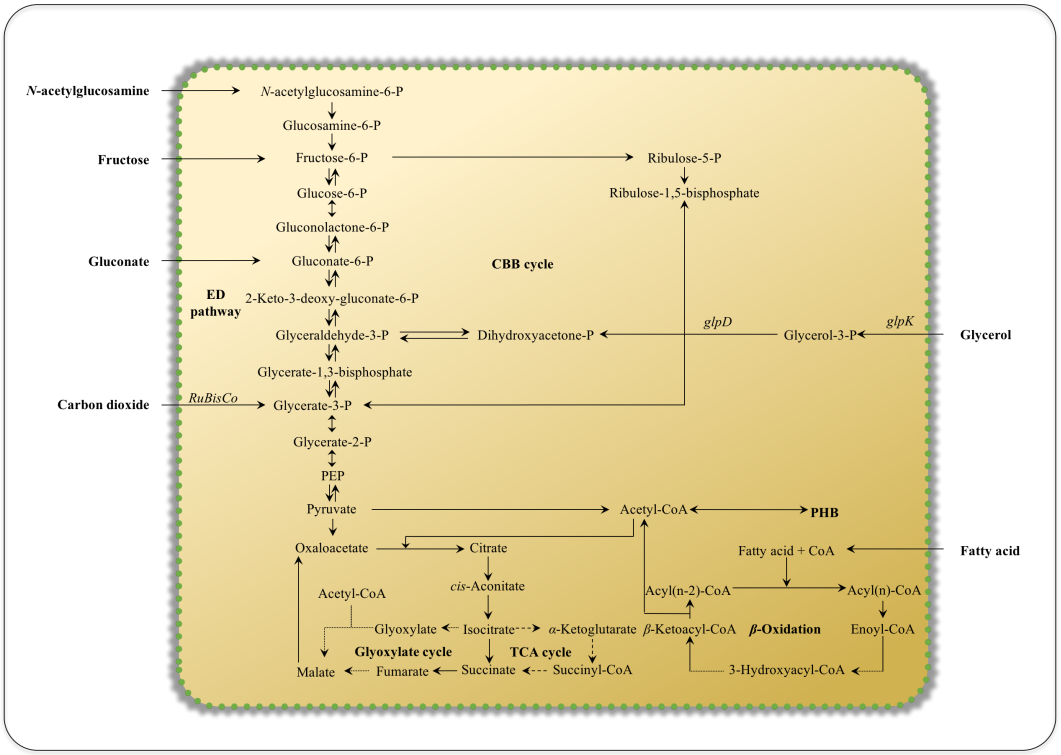


Figure 1.1 Key metabolic pathways in C. necator H16 under heterotrophic and autotrophic conditions. Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo) converts Ribulose-1,5-bisphosphate to Glycerol-3-P, which undergoes into the Entner-Duodoroff (ED) pathway for Pyruvate production, other carbon substrates such as N-Acetylglucosamine, Fructose and Gluconate are also channelled into Pyruvate through the same metabolic pathway. Pyruvate is catabolised by the pyruvate dehydrogenase, which converts Pyruvate into Acetyl-CoA, that can be converted into poly[(R)-3-hydroxybutyrate] (PHB) or can enter the Tricarboxylic acid cycle (TCA). Glyoxylate cycle, a variation of the TCA cycle, has many intermediates in common with the TCA cycle, although it differs in the conversion of Isocitrate, where Malonate is synthesised. Fatty acids and some organic acids are metabolised via β-oxidation, which also generates Acetyl-CoA. C. necator H16 can naturally metabolise glycerol with glycerol kinase (glpK) and glycerol-3-phosphate dehydrogenase (glpD) (Adapted from Johnson et al., 2015; Volodina et al., 2016).

One of the most attractive characteristics of *C. necator* H16 is its natural ability to produce poly[(R)-3-hydroxybutyrate] (PHB), a bio-plastic polymer known for its biodegradability. PHB is stored intracellularly when an excess of carbon or energy supply is present in the media; this excess is redirected from the central sink for accumulation in cellular granules. PHB has gained attention since it is a bioplastic that shares mechanical and thermoplastic properties with petroleum-based plastics, and it is the building block for the biosynthesis of other value-added compounds which converts this compound a key metabolite for the bio-based economy. *C. necator* H16 can accumulate PHB to over 80% (w/w) of its cell dry weight (CDW) (reviewed in Reinecke *et al*., 2009).

C. necator H16 represents a strain that covers the requirements as a chassis for biomanufacturing due to its characteristics such as 1) the elucidation of its complete genome, 2) the *in silico* analysis of metabolic networks, 3) genome-scale reconstruction and 4) various demonstrations of its genetic tractability via bacterial transformation by electroporation or bacterial conjugation (Park *et al.*, 2011; reviewed in Johnson *et al.*, 2015; Tee *et al.*, 2017); these characteristics have allowed focussing efforts to use this bacterium as a candidate for the production of value-added chemicals.

1.3 Engineering strategies for chemical production in *C. necator* H16

To produce value-added chemicals in *C. necator* H16 and to be able to use a specific carbon source such as economical and sustainable biological feedstock, the cell requires to be engineered by manipulating its native metabolic pathways. Strategies such as *rational design engineering* (*e.g.*, metabolic engineering aided by bioinformatics studies and recombinant strain engineering to regulate expression of proteins); and *evolutionary engineering* with directed and adaptive evolution tools, are vital biomolecular strategies for strain engineering (**Figure 1.2**). In the bio-based industry, these tools are also coupled with bioprocess engineering, this, to achieve cost-effective processes for fermentation and chemical production.

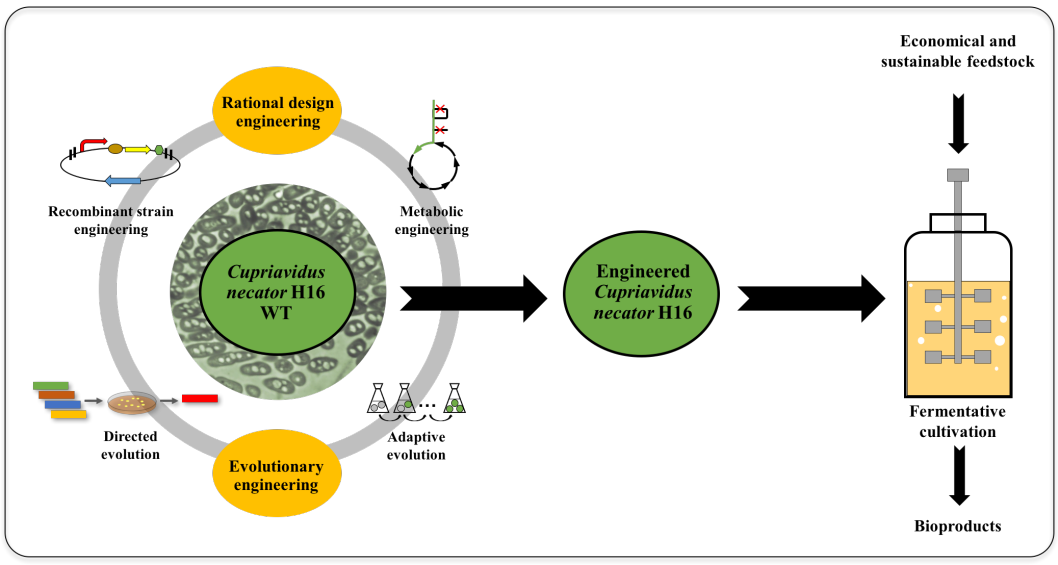


Figure 1.2 Engineering strategies for production of value-added chemicals in C. necator H16.

These strategies complement each other when the chassis cell requires to be engineered for chemical production. Engineers will usually choose a rational design approach (metabolic and recombinant strain engineering) for the development of a biocatalyst when the metabolic pathway of interest has been elucidated, because most of the details of the enzymes and genes responsible for the biosynthesis of the target value-added chemical and the undesired side products are well known, which contributes for an effective optimisation. Nevertheless, the cellular metabolism of microorganisms is very complex, and due to it is generally highly interdependent, it is difficult to account for all possible outcomes when a chassis is engineered with rational design engineering strategies. When this scenario is present, and there is no full understanding of the pathway or enzymes involved in the study of interest, the other strategies used to address this issue are the tools of evolutionary engineering such as directed evolution –aided by random mutagenesis- and adaptive evolution.

1.3.1 Rational design engineering

Rational design engineering (forward engineering) can be distinguished for controlled modifications of specific properties of proteins or genes, which deviates from random sequence perturbations, such as the class of optimisation techniques used in evolutionary engineering. Hence, this strategy creates molecules with new properties based on a known-predicted ability. Some of the main strategies used for rational design engineering are recombinant strain engineering and metabolic engineering, where metabolic engineering uses the first to modify the metabolic pathways of the cell factories.

1.3.1.1 Recombinant strain engineering

Recombinant strain engineering is defined as the improvement of cellular activity via the manipulation of regulatory, enzymatic and transport functions of the cell using recombinant DNA technology (Cakar *et al*., 2011) where in most cases, plasmids are used as the delivery platform of genes to be expressed in bacteria,

and the expression of these creates recombinant strains with the desired phenotype. Broad-host-range expression systems (*e.g.*, pBBR1-derived plasmids) have been reported to be transferable to *C. necator* H16 (Bi *et al*., 2013). It has been reported that usually plasmids that harbour origins of replication derived from broad-host-range plasmids can be replicated in *C. necator* H16, some examples of these plasmids are RSF1010, pSa, RP4 and pBBR1 plasmids (Gruber *et al.*, 2014).

As reviewed in Tee *et al*. (2014) and in Khalil *et al*. (2010), synthetic biology, defined as the engineering of biology, contributes to recombinant strain engineering by the re-design and construction of existing systems or the creation of novel biological systems with a desired phenotype, which is possible by the assembly of standardised and well characterised biological parts. One of the key biological bricks of a synthetic biology toolbox is the promoter. Promoters are the most critical parts of the synthetic toolbox, as they are in charge of regulating the expression of the downstream genes; if a maximised protein expression is desired, as it is for biomanufacturing purposes, usually strong promoters are selected for gene expression. The expression of the desired genes can be regulated with inducible promoters, which are prompt only when an exogenous chemical (inducer) or condition is present in the environment, but if a regulation of the expression is not required, then constitutive promoters are used. Some of the best proven inducible promoter systems for gene expression identified previously to work in *C. necator* H16, are the 1) L-arabinose-inducible P_{BAD} promoter and the 2) m-toluic acid-inducible promoter (Bi *et al*., 2013). On the other hand, for constitutive gene expression, Gruber *et al*. (2014) reported bacteriophage T5-derived promoters to be stronger than other constitutive promoters such as P_{tac} or P_{lac}, which had been previously reported as the strongest constitutive promoter for gene expression in *C. necator* H16 (Gruber *et al*., 2014).

As reviewed in Johnson *et al*., 2015 and Johnson *et al*., 2018, promoter engineering has been found as a critical strategy for tuning gene expression, which can be achieved through promoter with various strengths. A synthetic anhydrotetracycline-controllable gene expression constructed for *C. necator* H16 was reported by Li *et al*. (2015), where a modified *rrsC* promoter native of *C.*

necator H16 with two copies of integrated *tetO1* operators and a *tetR* regulated by a mutated version of *phaC1* promoter native of *C. necator* H16 were part of the new biological system. Another example of promoter engineering in *C. necator* H16, are the variants of the P_{j5} constitutive promoter; Gruber *et al*. (2014) reported promoters with various strengths which could potentially be used for regulation of gene expression in *C. necator* H16.

Other biological parts also play essential roles in the improvement of gene expression, such as the ribosome binding site (RBS). The *E. coli* consensus RBS has proven to be compatible with the *C. necator* H16 system, the RBS helps to improve the translational efficiency. The same scenario has been reported for the T7 mRNA stem-loop, which is located upstream to the RBS, this biological part has demonstrated to enhance mRNA stability in *C. necator* H16 (Bi *et al*., 2013).

1.3.1.2 Metabolic engineering

As reviewed in Yadav *et al*. (2018), metabolic engineering is based on the modification of a microorganism to improve its bioproduction by altering metabolic pathways to understand better and direct cellular pathways; it is motivated by commercial applications for the production of useful metabolites. Metabolic engineering of a strain can be achieved with different biomolecular tools such as genetic and recombinant strain engineering, as the downregulation or overexpression of specific proteins in a metabolic pathway is required; metabolic engineering would optimise either existing biochemical pathways or introduce new components of non-native pathways to obtain a high-yield production of the desired metabolite. In order to do so, an overall analysis of the metabolic pathways of the strain must be analysed, whereby the detection of metabolic bottlenecks is crucial for the efficient bioproduct synthesis under predefined environmental conditions.

As reviewed in Johnson *et al.*, 2015, metabolic engineering will aim to rewire or reconstruct metabolic networks that will yield a desired metabolic phenotype using a determined substrate convenient for the bioprocess. Most of the metabolic

engineering studies focus on the regulation of determined enzymes levels either by deletion, addition or amplification of specific pathways (Zha *et al*., 2009). Hence, metabolic engineering changes cell metabolism with the aid of recombinant strain engineering in order to improve the productivity during fermentation.

1.3.2 Evolutionary engineering

As reviewed in Dragosits *et al*. (2013), evolutionary engineering (reverse engineering) under a controlled laboratory setting is another scientific approach for generating robust and optimised engineered strains as production systems. Evolutionary engineering roots back to Antonie van Leeuwenhoek, Louis Pasteur and Charles Darwin days, when the first microorganisms and natural selection for biological evolution were discovered. This strategy represents an option to overcome the limitations of the rational design engineering (forward engineering), limitations such as the lack of extensive knowledge of how the metabolic networks work, lack of information on the regulatory factors, enzymes involved or kinetics of the metabolic pathway, or in some cases, even the difficulties of cloning industrial strains due to genetic complexity (Cakar *et al*., 2011).

Evolutionary engineering can improve microbial growth on specific relevant substrates or develop stress resistance. Usually, selection pressure and genetic mutations are used to create a pool of variant strains that have gone through natural selection adaptation changing temporarily or permanently the strain phenotype. Usually, the strategy consists of improving cellular viability on an alternative carbon source, or to develop tolerance on inhibitors, pH, organic solvents, or temperature (reviewed in Dragosits *et al*., 2013).

After the desired phenotype has been found, and the strain mutant or variant has been properly isolated, genome or RNA sequencing are the further steps to identify the genetics behind -reverse engineering-; these genetic changes can be beneficial mutations that combined with rational design -forward engineering- can result in further strain improvement.

As reviewed in Dragosits *et al*. (2013), in the past 30 years, an increasing number of experiments employing evolutionary engineering in microbial cells such as *Escherichia coli* and *Saccharomyces cerevisiae* have been reported, but despite being a powerful approach, this strategy is still limited in *C. necator* H16 (reviewed in Johnson *et al*., 2015).

1.3.2.1 Directed evolution

Directed evolution is the laboratory process where biological systems with desired traits are created through iterative rounds of mutation that create genetic diversification in order to find a specific phenotype. As reviewed in Cobb *et al*., (2013), a library of mutants is screened, and the best mutant is selected for characterisation. This strategy has become one of the most widespread and useful tools in bioengineering, as the generation of whole-cell mutants has helped to identify new genes and study its properties (Lawrence, 2002). This tool is capable of reaching even the most conserved biological processes by the creation of variation in a population which is then selected individually – be it an enzyme (protein level), operons (pathways) or entire genomes (whole-cell) (reviewed Tizei *et al*., 2016), where genomic mutations are introduced stochastically through external mutagenic compounds (reviewed in Tee *et al*., 2013).

It was in the 90s when the term of directed evolution began to take root, defined as an iterative process based on three main steps: 1) the generation of a library of mutants of a specific enzyme, pathway or biological system and 2) the screening of the library in a high-throughput fashion that identifies the mutants with the improved properties, and 3) the subsequent rounds of diversification and selection using the best mutant as the basis for the next round of mutagenesis, process that is repeated until the desired level of improvement has been achieved (reviewed in Cobb *et al*., 2013).

Directed evolution at different biological levels: protein, pathways, genome

As reviewed in Zhao *et al*. (2007) and in Cobb *et al*. (2013), directed evolution has its roots in classical adaptive evolution and strain engineering, although modern directed evolution emerged about 25 years ago with the demonstration of iterative rounds of polymerase chain reaction (PCR) for random mutagenesis where activity was screened for improved proteins, ever since, different strategies have been developed which have enabled not only the evolution of 1) proteins or 2) pathways, but also it has been able to evolve 3) entire microorganisms in its genome level, which helped to create novel functions previously not found in nature. As reviewed in Tobin *et al*. (2000), as an evolutionary engineering tool, it has the advantage that there is no need to know the protein structure, metabolic pathway or the entire genome of interest, which in some cases is very difficult to predict in advance.

Directed evolution of proteins, is a combinatorial approach that can be used to improve the functionality of enzymes by using the rational design site-directed mutagenesis and random mutagenesis, which allow scientists to produce focused libraries, that only concentrate in randomised regions that are expected to be richer in beneficial mutations. Other techniques such as error-prone PCR, are used for random mutagenesis, this is done in a specific gene sequence, and it creates diverse protein libraries of existing scaffolds (reviewed in Cobb *et al*., 2013; Kumar *et al*., 2013).

As reviewed in Dalby (2011) and Du *et al*. (2011), on the other hand, directed evolution of pathways is focused on enabling cost-effective microbial biosynthesis by optimizing the host-cell, as it dodges the limitations of rational design due to even simple pathways are not fully understood, as there are gaps in the regulation mechanisms and metabolic networks, techniques such as DNA shuffling have been used to improve the phenotype of host cells. Crameri *et al.* (1997) engineered an *E. coli* strain by introducing an evolved arsenate resistance operon from *Staphylococcus aureus*, composed of genes *arsR*, *arsB*, and *arsC* which encode for a repressor regulatory protein, an arsenite membrane pump, and an arsenate reductase respectively, the operon was evolved by using recursive

rounds of DNA shuffling, where after three rounds of mutagenesis, non-synonymous mutations were identified in *arsB* and *arsC*. This recombinant strain was able to survive in 0.5 M arsenate, which is 40 times higher than the wild-type strain, the use of directed evolution in this pathway allowed to use this strain as a candidate for applications in bioremediation arsenate from gold metallurgy.

Directed evolution can cover genome-scale as well. Directed evolution on genome-scale has generated new whole cells. The fundamentals of directed evolution at genome level is the incorporation of accelerated mutagenesis which links phenotype with genotype (**Figure 1.3**) (Liu *et al.,* 2005; reviewed in Cobb *et al*., 2013), it emerged as well as a solution for the complex cellular metabolism, where random mutagenesis of bacterial strains and the use of a selective pressure allows having a library of 10^9 to 10^{12} mutant strains, which would not be possible to achieve otherwise, this procedure goes through iterative cycles until the desired phenotype is reached. Nevertheless, a selection pressure (usually inhibitors) for the process is highly recommended, as the screening, identification, and analysis of mutants can be extremely laborious, hence, a proper planning, choice of strain, efficient mutagenesis and screening method are the key elements for directed evolution at genome scale (Connor *et al*., 2010). The drawback of this strategy is that in some cases, the selective pressure is not possible for all processes (Connor *et al*., 2010). For these reasons, in some biomanufacturing engineering processes, both forward engineering (rational design) and reverse engineering (evolutionary engineering) are used to engineer strains.

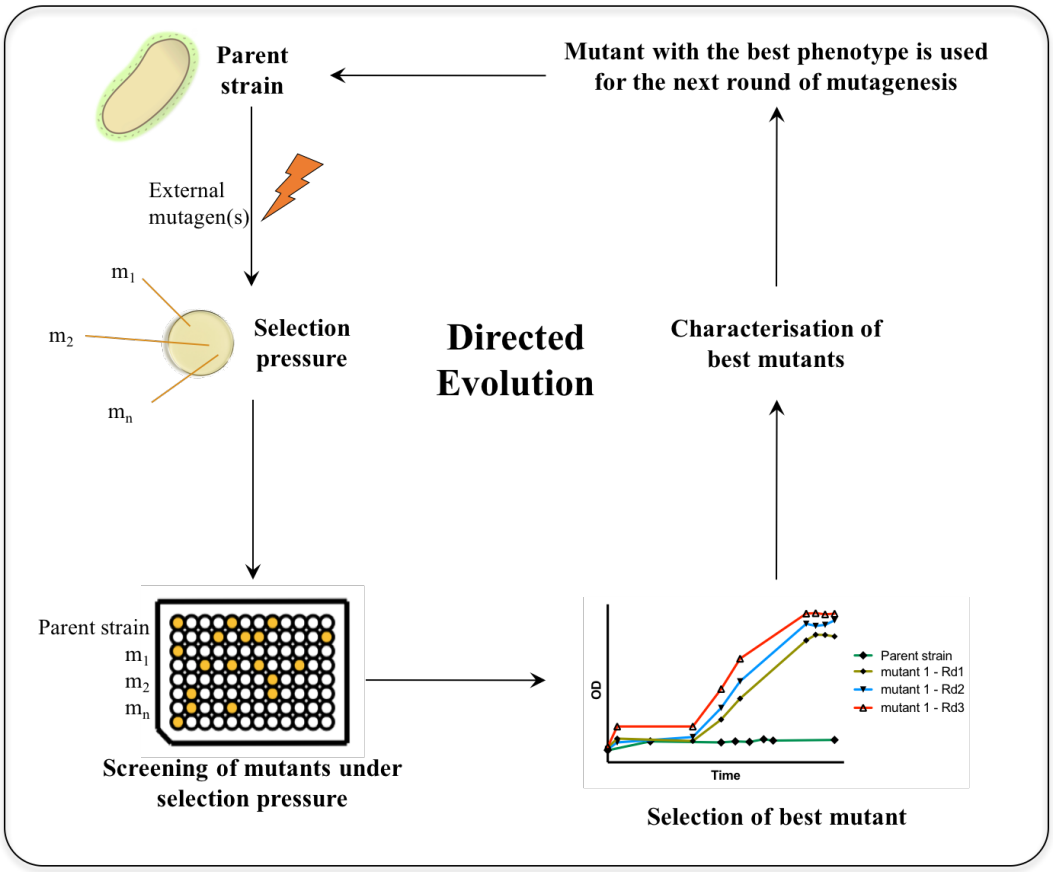


Figure 1.3 Directed evolution of the whole cell (genome level) process via random mutagenesis aided by chemical or physical mutagens. Cells are exposed to external mutagens and then selected for resistance to an inhibitor in agar plates. Resistant colonies are then screened for growth in liquid media usually in 96-well microtitre plates to cover a higher number of mutants under selective pressure. The best mutants are then characterised to confirm the desired phenotype, and then the best mutant is chosen as the parent strain for the next round of mutagenesis.

Connor *et al*. (2010) used directed evolution with random mutagenesis aided by a chemical mutagen to develop a mutant strain of *E. coli* to produce 3-methyl-1-butanol (3MB) in combination with rational design engineering by expressing heterologous genes from *Lactobacillus lactis*, *S. cerevisiae*, and *B. subtilis*, the genes used in this study are involved in the 3MB production, and this combinatorial strategy resulted in a mutant *E. coli* strain capable of producing 4.4 g/L of 3-methyl butanol. 4-aza-D,L-leucine was used as a selection pressure, which acts as an inhibitor to the cell as it is a structural analogue of L-leucine, this analogue interferes with the regulation of amino acid synthesis and its incorporation into polypeptides. The fundamental of the strategy is that cells can adapt to survive in the presence of an inhibitor by developing the ability of

producing the natural amino acid in higher enough quantities that will compete with the analogue for its incorporation into polypeptides, the mutant that has been best adapted is found in the screening selection method, afterwards a verification (characterisation) of the desired phenotype is performed before the best mutant is subjected to the next round of mutagenesis. Mutants with the desired improved phenotype will show an increased growth when grown in the presence of the inhibitor. As reviewed in Hall (1998), in directed evolution *via* random mutagenesis, beneficial mutations are typically observed when a population of evolved cells is spread onto a media upon which growth cannot occur unless a known mutation occurs. Another example is a study published by Sakurai *et al*. (1993), they isolated a mutant strain of *Serratia marcescens* resistant to acidomycin –a biotin analogue- capable of overproducing biotin, where cells were first subjected to random mutagenesis with a chemical mutagen and selected in agar plates supplemented with the biotin analogue.

On the other hand, this strategy has been used as well to elucidate unknown metabolic pathways, as reviewed in Streit *et al.* (2003), other studies performed in different strains such as *Corynebacterium glutamicum, Mesorhizobium loti, Methylobacillus flagellatum, Kurthia sp., Sinorhizobium meliloti*, where cells were subjected to detailed analysis for the elucidation of their biotin synthesis pathways or for biotin overproduction, aided by engineering tools of random mutagenesis and complementation studies, where biotin analogues were used as selection pressure.

Chemical and physical mutagens

As reviewed by Nielsen (1998), random mutagenesis is a common strategy in bioengineering as high mutation rates can emerge naturally, but also, as mentioned in previous paragraphs, mutations rates can be increased with the aid of a chemical or physical mutagen, which contributes to the formation of genetic novelty (Sniegowski *et al*., 1997; Bachmann *et al.,* 2012; reviewed in Tee *et al*., 2013).

After increasing the genetic diversity by using external mutagens, the next step is the direct selection on solid media of the mutants that have been able to survive

under selective pressure, such as high metal concentrations, low or high temperatures, high concentrations of inhibitors, etc., although it is important to consider that for this reverse engineering method, the major disadvantage is the fact that not too many repeated cycles of mutations are recommended as this may result in highly improved phenotype but crippled variant strains (Cakar *et al*., 2005).

The external mutagens frequently used for the increase in random mutagenesis can be divided into two categories: 1) chemical mutagens, and 2) physical mutagens. The mutations are known as transitions when interchanges that involve nucleotides of similar shape occur, such as two-ring purines (A ↔ G), or one-ring pyrimidines (C ↔ T), all four transitions are T_s: AT → GC and GC → AT; and transversions when there are interchanges of purine to pyrimidine bases (A ↔ T or C; G ↔ T or C), where all eight transversions are T_v: AT → TA, AT →CG, GC →CG, and GC → TA (reviewed in Tee *et al*., 2013). Chemicals mutagens induce high frequencies of base-pair substitutions and some lethality depending on the concentration used. Some of the most widely used chemical mutagens are the alkylating agents *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (NTG) and ethyl methanesulfonate (EMS), which produce transitions at G-C sites. On the other hand, one of the most commonly used physical mutagens is ultraviolet light (UV light, 254 nm), which has the advantage of producing a more full range of substitutions usually in pyrimidines, it includes transitions and transversions; a less common physical mutagen are X-rays. Some strategies of directed evolution consist of combining both mutagens –chemical and physical- which provide outcomes that can satisfy experimental needs. It is important to consider as well the optimal dose of the mutagens, since high concentrations may result in a 100 % killing; a balance between the competing needs for a high mutation frequency and a reasonable survival must be optimised, where usually a 50 % survival rate provides the highest proportion of mutants (Lawrence, 2002).

1.3.2.2 Adaptive evolution

As reviewed in Hall (1998), adaptive evolution is defined as the spontaneous mutations that occur in a microorganism as a consequence of prolonged periods of stress. During adaptive evolution, the strain of interest is cultivated under very well defined conditions of stress for prolonged periods, which can range from weeks to months or even years depending on the target phenotype and the condition set; the long cultivation period allows the selection of improved phenotypes. Microorganisms have simple nutrient requirements, which makes the process feasible, as they can be easily cultivated under controlled conditions in the laboratory, and due to cells usually grow very fast, they can be cultivated for many generations in a relatively short period (Cakar *et al*., 2011; reviewed in Dragosits *et al*., 2013).

With adaptive evolution, scientists can know that the phenotype changes are associated with the specific growth conditions provided, which can be genotype-phenotype elucidated with technologies such as whole genome sequencing (WGS) and proteomics (Cakar *et al*., 2011; reviewed in Dragosits *et al*., 2013). *C. necator* H16 is a perfect candidate for adaptive evolution since it is a ubiquitous inhabitant of freshwater and soil that has shown to be well adapted to the constantly changing environment (reviewed in Volodina *et al*., 2016).

The laboratory selection method of adaptive evolution usually consists in batch cultivation in shake flasks (although it can be performed in chemostat cultures as well) where propagation of the strain of interest is performed through serial cultivation (**Figure 1.4**). An initial population of wild-type cells is cultivated in the desired controlled media, and an aliquot of the batch is then withdrawn - generally daily and at late log or early stationary growth phase (Zambrano *et al*., 1993)- and transferred to a new flask for an additional round of growth with fresh media either with the same conditions or with variations of the initial conditions. Adaptive evolution has the advantage of being a cheap process and that various parallel cultures can be performed at the same time, if too many conditions need to be analysed or different strains, the cultures can be easily performed in 96-well

microtitre plates where hundreds of variations can be analysed in parallel (Gonzalez *et al*., 2013).

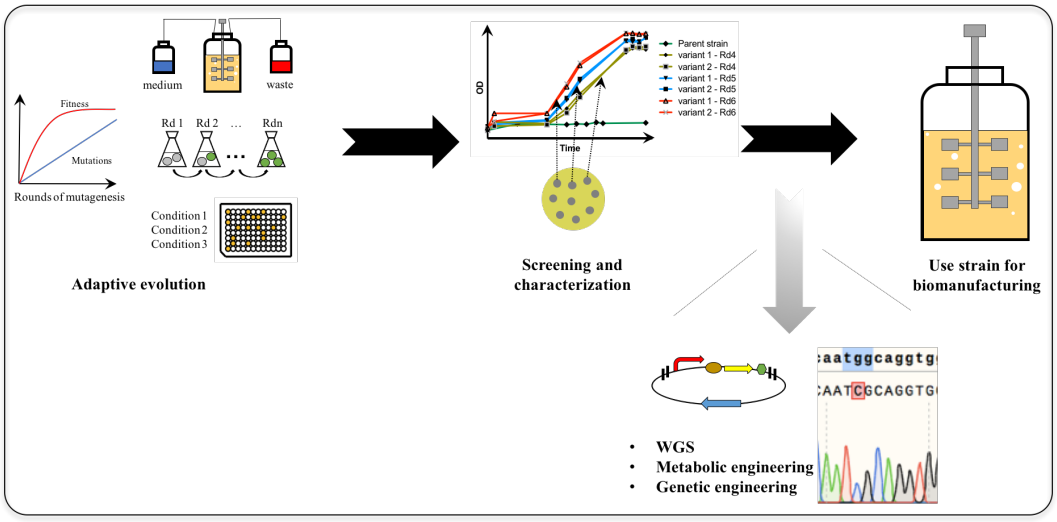


Figure 1.4 Adaptive evolution for biomanufacturing. Generally, the fitness increase tends to slow down after the first rounds of adaptive evolution due to inherent properties of biological metabolic networks, while the number of mutations increases with the rounds of mutagenesis; adaptive evolution can be performed in chemostat or batch cultures (conical flasks or microtitre plates when a variety of conditions are analysed). After the process of adaptive evolution, the screening and selection of the best variant strain is performed, where the variant can be used directly for the desired purpose, or if required, an identification of the genetic basis of the improved phenotype can be combined with metabolic engineering or other biomolecular tools to improve further the desired phenotype (Adapted from Connor et al., 2010).

As observed in **Figure 1.4** the improvement of fitness as a function of the rounds of mutagenesis tends to be non-linear, the linear increase of adaptation can be observed in the first rounds of mutagenesis, but it is considerably slower during the last rounds of adaptive evolution (Barrick *et al*., 2009), therefore, it is essential to consider this factor, as a prolonged adaptive evolution will not necessarily lead to better phenotypes (Hua *et al*., 2007). During adaptive evolution, cells can be challenged whit changing environmental conditions as selection pressure, which can be divided into two main categories: 1) environmental stress and 2) nutrient availability; these can be used separately or at the same time (a combinatorial approach that may lead to major findings).

As reviewed in Dragosits *et al*. (2013), various phenotypes will occur at the beginning of adaptive evolution, which will compete to dominate over the total

population, where only the most stable phenotypes will be accumulated rapidly, although the population will not be homogeneous at any point of the evolution experiment. Therefore the best variant strains must be identified by screening methods and isolated for further characterisation. In some cases, in the characterisation experiments of the variant strain, it has been observed that the selection of improved phenotype strains in a determined environment sometimes may lead to trade-offs when the mutant strain is transferred to another selective or stressful condition. Due to these reasons, sometimes, the last step of evolutionary engineering is to transfer the identified mutations to the wild-type strain rather than using the variant strains directly for bioproduct formation (Cakar *et al*., 2011).

A variety of studies on adaptive evolution for improving nutrient availability have been performed in different microorganism such as *E. coli, T. fusca,* and *L. lactis*, where cells have been evolved in glycerol, glucose, lactate media or milk (Ibarra *et al*., 2002; Conrad *et al*. 2010, Hua *et al*. 2007, Bachmann *et al*., 2012). *C. necator* H16 was evolved to grow in glucose by incubating cells in high concentrations of the substrate, where glucose-positive phenotype variant strains were identified, which showed elevated activities of glucose-6-phosphate dehydrogenase (Franz *et al*., 2012). A similar approach of adaptive evolution for glucose assimilation was applied to *C. necator* H16 strains but with a combinatorial approach by using as well a chemical mutagen (NTG) to develop mutant strains capable of growing in glucose (Kim *et al*., 1995).

Even though when we know in advance that DNA replication is a process with high fidelity wherein microbial cells it only has a mutation rate of 10^{-10} per base pair per replication, the emergence of mutations occurs during adaptive evolution due to large population sizes and different generations of these, which occur during adaptive evolution (Drake, 1991).

As reviewed in Dragosits *et al*. (2013), a variety of mutations have been detected in microbial cells after adaptive evolution such as single-nucleotide polymorphism (SNPs), small-scale insertions and deletions (Indels), or large

amplifications or deletions in genomic regions (**Figure 1.5**), mutations that may contribute to a change in the phenotype of cells.

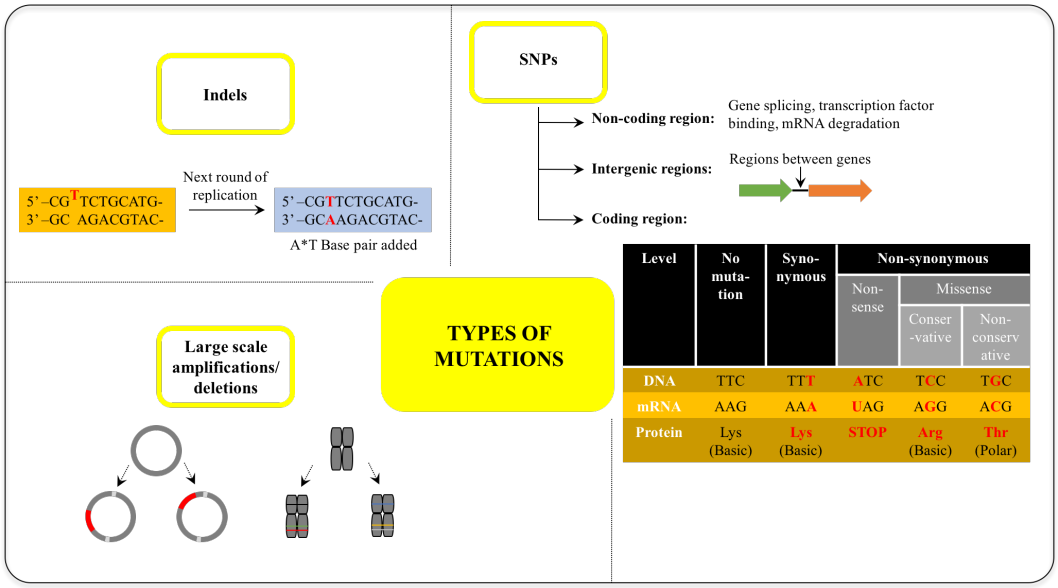


Figure 1.5 Mutations that are usually identified in evolutionary engineering. Insertions and deletions (Indels), Single nucleotide polymorphisms (SNPs) and larger deletions and insertions, which contribute to genetic and sometimes in gene regulatory changes which may impact in the phenotype of cells.

After identifying the desired phenotype of evolved microbial cells, these can be used directly for biomanufacturing purposes, nevertheless, scientist must be aware that there are some drawbacks related to this strategy, as mentioned earlier, evolved microbial strains are likely to show trade-offs developed during the stress exposure, therefore after the isolation of the desired mutant, different analysis such as the confirmation of true mutations and the performance of the variants under different environment or stress conditions must be considered if the biomanufacturing process requires it. For this reason, in many cases, the final step of evolutionary engineering is to transfer the identified beneficial mutations to a well-defined strain. The integration of evolutionary engineering and metabolic engineering offers a variety of tuning possibilities at multiple levels for bioengineering; therefore, it is a potential tool to explore in microorganisms such as *C. necator* H16.

1.4 Expanding carbon substrate range of *C. necator* H16

Even though when wild-type *C. necator* H16 can grow already in a wide range of carbon substrates, some carbon substrates that are economically viable for the production of value-added chemicals in the industry, are not assimilated or show limited growth in *C. necator* H16, such as glucose, and glycerol. Engineering *C. necator* H16 can expand the carbon substrate range in order to use sustainable and economical carbon feedstock, which will always be preferred for industrial scale production, fermentation, and for bioprocess development.

1.4.1 Glucose

C. necator H16 does not assimilate glucose, xylose, and arabinose (Fukui *et al*., 2014), the genomic sequence of *C. necator* H16 revealed that it lacks the genes required for glucose transport, although once it is available intracellularly, it can metabolise glucose (Pries *et al*., 1990; Pohlmann *et al*., 2006; Sichwart *et al.*, 2011). Due to glucose is one of the most abundant sugars commonly used as carbon substrate for microbial fermentation, and it can be obtained from diverse sources, mainly from most carbohydrate crops (Soucaille, 2002; reviewed in Ezeji *et al*., 2007; Linde *et al*., 2008), diverse strategies have been developed to create *C. necator* H16 strains capable of metabolising glucose.

With the aid of evolutionary engineering, scientists have developed *C. necator* H16 strains capable of utilising glucose as carbon source. One of the first reported strategies were *C. necator* H16 cells mutagenised with UV light for random mutagenesis, where Schlegel *et al*. (1965) achieved one of the first *C. necator* H16 strains capable of assimilating glucose. Kim *et al*. (1995), on the other hand, used a chemical mutagen (NTG) to evolve *C. necator* H16 glucose-utilising strain. Another successful approach that made use of adaptive evolution to obtain *C. necator* H16 glucose-utilising strains was performed by Franz *et al*. (2012), by incubating *C. necator* H16 strains with high glucose levels of up to 20 g/L for about ten days. Studies revealed that most of these mutant glucose-utilising strains

had alterations in gene function and expression, such as mutation in the *N*-acetylglucosamine phosphotransferase –a transporter that facilitates the diffusion of *N*-acetylglucosamine in the cell- which apparently was altered in its substrate specificity, by also accommodating glucose import; an increase in glucose-6-phosphate dehydrogenase expression was also detected, which is the essential enzyme for glucose metabolism via the Entner Duodoroff pathway (Raberg *et al*., 2011; Franz *et al*., 2012).

Some other strategies have been used to construct glucose utilising *C. necator* H16 strains with metabolic engineering, where *C. necator* H16 expressed constitutively the genes that encode for glucose kinase (*glk*) from *E. coli*, and the gene encoding for the facilitated diffusion protein (*glf*) from *Zymomonas mobile*. The expression of both genes resulted as well in a glucose-utilising *C. necator* H16 strain, where the *glf* gene conferred the assimilation of glucose and the *glk* gene helped in the glucose metabolism (Sichwart *et al*., 2011).

1.4.2 Sugars derived from lignocellulosic biomass

As reviewed in Hatti-Kaul *et al*. (2007), within the lignocellulosic biomass, mannose is mostly found, therefore in many microorganisms lignocellulosic biomass is considered a candidate as carbon source for biomanufacturing purposes, but as it occurs with glucose, *C. necator* H16 does not assimilate either metabolises mannose, therefore, different approaches have been successfully carried out to obtain mannose-utilising *C. necator* H16 strains. A recombinant *C. necator* H16 strain was designed by Sichwart *et al*. (2011) whereby with the heterologous expression of the *glf* gene (facilitated diffusion protein) and the *mak* and *pmi* genes (for mannofructokinase and phosphomannose isomerase respectively) the mannose-utilizing *C. necator* H16 was constructed. The *glf* gene conferred the mannose assimilation, whereas the *mak* and *pmi* genes conferred the mannose metabolism. The engineered strain was capable of assimilating and metabolising mannose due to the *glf* gene enabled mannose assimilation, further the *mak* expressed gene catalysed the reaction to convert mannose to mannose-6-phosphate, and the *pmi* expressed gene was in charge of the reversible

isomerization of mannose-6-phosphate to fructose-6-phosphate (due to *C. necator* H16 wild-type assimilates fructose) then, the fructose-6-phosphate was subsequently metabolised via the Entner-Doudoroff pathway.

1.4.3 Glycerol

The importance of reducing the cost of feedstocks to produce value-added chemicals is because feedstock represents up to 50 % of the entire production cost (Zhu *et al*., 2010). Sustainable manufacturing has become a trend in the recent years, and crude glycerol is an important waste stream which can help to reduce the production cost in bioprocess industry owing to its low price and abundance in the market; it is desirable carbon substrate as a building block chemical found as a byproduct of the transesterification of alcohols and oils (Cameron *et al*., 1998; Ashby *et al*., 2012). *C. necator* H16 shows slow growth when glycerol is used as the primary carbon source, but it has been previously demonstrated that with the aid of metabolic engineering, engineered strains of *C. necator* H16 have shown to enhance glycerol assimilation with the concomitant expression introduction of the aquaglyceroporin *glpF*, which encodes for a transport protein, and with the glycerol kinase *glpK* from *E. coli*; although it was demonstrated as well that the expression of *glpK* only was capable of enhancing glycerol assimilation (Fukui *et al*., 2014).

1.4.3.1 Sources of crude glycerol

As reviewed in Dobson *et al*. (2012) and in Yang *et al*. (2012), crude glycerol is a candidate to be used as a renewable feedstock due to its availability and meager prices; it can be found not only as a byproduct of the biodiesel industries but also from oleo-chemical, bioethanol industries and fat splitting processes. In the biodiesel production process, glycerol covers approximately 10 % (w/v) of the total product; in the bioethanol industry through the fermentation of sugars by yeast, glycerol comprises up to 2 % by volume of the liquid fraction in the whole stillage (reviewed in Rausch *et al*., 2006; Kurosawa *et al*., 2015).

Biodiesel is produced from vegetable oils or animal fats by the transesterification with alcohol –ethanol or methanol- and catalysed by NaOH or KOH where, as mentioned before, glycerol represents about 10 % (w/v) of the ester (**Figure 1.6**) (Paulo da Silva *et al*., 2009).

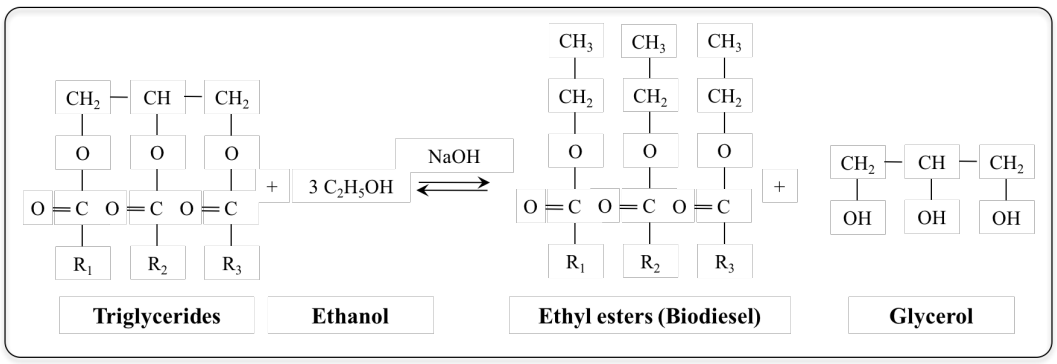


Figure 1.6 Synthesis of biodiesel. Transesterification reaction of a triglyceride with an alcohol (ethanol), using NaOH as a catalyst for biodiesel production (a mixture of fatty acids ethyl esters) and glycerol (Adapted from Paulo da Silva et al., 2009).

1.4.3.2 Chemical composition of crude glycerol

As reviewed in Dobson *et al*. (2012), crude glycerol has other components than glycerol that must be considered when engineering strains for chemical production if crude glycerol is used directly as feedstock, as these could inhibit cell growth. These components are usually alcohols (such as methanol), salts, heavy metals, free fatty acids, and soaps. The composition of crude glycerol depends on the parent feedstock (such as rapeseed, soybean, animal oils or even waste cooking oil); the transesterification process used for the biodiesel production (for instance if methanol or ethanol were used); the type of catalyst (potassium hydroxide or sodium methoxide); the recovery efficiency of the biodiesel; among other impurities in the feedstock. In crude glycerol, the content salt can vary and go as high as 5 % (w/v), and methanol can go as high as 32 % (w/v) (Posada *et al*., 2010; reviewed in Yang *et al*., 2012). Therefore, ideally, microorganisms which have a certain resistance to these impurities represent the perfect chassis for bioconversion with crude glycerol.

1.4.3.3 Annual production of crude glycerol and price

The low price of crude glycerol, which can contain from 50 % to up to 85 % glycerol, can be obtained in large quantities mainly from the biodiesel industry, where for every 100 lb of biodiesel that is produced by transesterification of animal fats or vegetable oils, about 10 lb of crude glycerol are produced (Durnin *et al*., 2009; Cavalheiro *et al*., 2009; Tanadchangsaeng *et al*., 2012). As reported by Oleoline (2018), in The Independent Oleo reporter, the prices of crude glycerol in 2018 in the USA ranged from only 0.05 to 0.18/lb.

As reviewed in Dobson *et al*. (2012), there was a 295 % increase of biodiesel production in 2005 with respect to 2001 due to biodiesel was a response of a green initiative by the governments, where a global production of 390 million litres per year of crude glycerol was produced mainly in Germany, France, Italy and USA, only in the USA the annual production of biodiesel in 2012 was 969 million gallons, which is about 3 times higher than the 343 million gallons produced in 2010 (U.S. Energy Information Administration, 2013), and it is expected that the biomass-based diesel production reaches up to 36 billion gallons by 2022; therefore the surplus production of biodiesel will generate about 3.6 billion gallons of crude glycerol as a waste by-product by then (reviewed in Moser *et al*., 2011).

The massive production of crude glycerol around the world is an opportunity for the biotechnological community for integrating biodiesel production with microbial fermentation, as at present, glucose is the most widely used chemical feedstock for biomanufacturing, and although the price of glucose is similar to crude glycerol (US$ 0.21-0.23/lb), a fluctuation in its price has been observed during the last 20 years, where it reached up to US$ 40/lb in 2010 (reviewed in Dobson *et al*., 2012); what is more, waste glycerol is not only a cheaper and sustainable substrate compared to glucose, but also it is not considered a suitable human source, as is the case for glucose and sucrose (reviewed in Li *et al*., 2013).

1.4.3.4 Purification of crude glycerol

As reviewed in Dobson *et al*. (2012) and in Min *et al*. (2011), although crude glycerol can be purified, and the refined glycerol can be used in the cosmetics, tobacco, and pharmaceutical industries, the industrial purification of glycerol is not economically viable for small industries, even in some cases large industries do not have much interest in purifying crude glycerol, as, over the last 20 years, refined glycerol dropped from US$ 1/lb to US$ 0.34/lb.

When crude glycerol is purified in industry, it is refined by filtration methods, fractional vacuum distillation, and chemical additives, and these processes produce different commercial pure glycerol grades, nevertheless, the difficulty of the process is that these treatments increase the production cost (reviewed in Johnson *et al*., 2007; Prada-Palomo *et al.,* 2012), therefore, high abundance, high degree of reduction, and low process make direct use of crude glycerol as an attractive feedstock for biomanufacturing purposes, due to the removal of the impurities from crude glycerol are cost-intensive, and the potential revenue for technical quality refined glycerol, very low (Lindlbauer *et al*., 2017).

1.4.3.5 Bacterial glycerol uptake

A variety of microorganisms can use glycerol as feedstock for microbial bioconversion, where value-added chemicals such as 1,3-propanediol, citric acid, succinic acid and the well-known PHAs bioplastics (Koutinas *et al*., 2007) can be produced; the advantage of some of these bioproducts is that they will not only be produced from renewable feedstock but also these final products can be biodegradable, which completes an environmentally friendly cycle strategy.

As reviewed in Dobson *et al*. (2012), microorganisms uptake glycerol into the cell by facilitated diffusion, and they encode metabolic pathways that can convert it into different metabolic intermediates, as glycerol is naturally found in different forms such as triglycerides. In *E. coli*, glycerol can cross the membrane by passive diffusion when high concentrations of glycerol are present (Truniger *et al.*, 1993),

nevertheless when the external concentration of glycerol is low, the glycerol facilitator protein (GlpF) transports glycerol by facilitated diffusion, subsequently glycerol is converted to glycerol-3-phosphate by glycerol kinase (GlpK), which is then further metabolised within the cell as it is no longer a substrate for GlpF (Paulo da Silva *et al.*, 2009). A similar utilisation of glycerol is believed to occur in *C. necator* H16, where there are two putative glycerol kinases (GlpK) and two putative glycerol-3-phosphate dehydrogenases (glpD) located in both chromosomes of *C. necator* H16 (**Table 1.2**)

Table 1.2 Genes potentially related to glycerol metabolism in *C. necator* H16.

Gene ID	Chromosome	Original annotation	Identity to *E. coli* protein (%)
H16_A3590	1	Glycerol facilitator or related permease	glpF (25 %)
H16_A2507	1	Glycerol kinase	glpK (52 %)
H16_A2508	1	Glycerol-3-phosphate dehydrogenase	glpD (29 %)
H16_B1199	2	Glycerol kinase	glpK (28 %)
H16_A0336	2	Glycerol-3-phosphate dehydrogenase	glpD (39 %)

Adapted from Fukui *et al.*, 2014.

Among the non-pathogenic microorganisms relevant for industry that have been reported to date which can use either pure and crude glycerol as feedstock for chemical production and obtain very similar tolerance to crude glycerol in comparison with pure glycerol are *Klebsiella pneumoniae* and *Clostridium butyricum* for the production of 1-3 propanediol, which has a great potential to be used in the plastic industry as a monomer of polyesters, polyethers, polyurethanes and poly(trimethylene) (used for a variety of product such as polymers, food, medicines, cosmetics, lubricants and fabrics) (Mu *et al*., 2006; Oh *et al*., 2008; González-Pajuelo *et al*., 2004; Moon *et al*., 2010; Jun *et al*., 2010); *Yarrowia lipolytica* for citric acid production used in the pharmaceutical and food industry due to its low toxicity (Rymowicz *et al*., 2008); *K. pneumoniae* for ethanol production (Oh *et al*., 2011); *Kluyvera cryocrescens, Escherichia coli*, and *C. butyricum* for lactic acid production (Hong *et al*., 2009; Choi *et al*., 2011); and *E. coli* (Andreeßen *et al*., 2009; Fukui *et al*., 2014) and *C. necator* H16 for PHA production (Fukui *et al*., 2014; González-Villanueva *et al*., 2018 (submitted)), which has demonstrated that crude glycerol can be used directly as a feedstock for chemical production without any pre-treatment.

1.5 Bioproduct synthesis in *C. necator* H16

In order to use *C. necator* H16 as a cell factory, key points must be considered such as 1) sustainable and economically viable feedstock that can produce the chemical of interest, 2) native or heterologous metabolic pathways or genes required for the production of the chemical of interest, 3) analysis of the product of interest, *i.e.*, if it is toxic to the cells or not, and 4) analysis of the carbon flux to check if it can be diverted and maintain at the same time sufficient cell viability. Some of the bioproducts that have been synthesised in *C. necator H16* considering these key points are bioplastics (PHAs), biofuel, methyl ketones, methyl citrates, feluric acid, chyral synthons, among others, where different strategies either using forward or reverse engineering have been used to engineer the strain.

1.5.1 Bioplastics

C. necator H16 is well known for its production of PHA, granules that are naturally produced as carbon storage materials. As reviewed in Lenz *et al*. (2005), bioplastics have similar properties to petroleum-based plastics found in the market, they are not only biodegradable (which can be biodegraded in most biological active environments), but they can also be produced from renewable feedstock. The biodegradation of these polymers depends on their chemical characteristics, *i.e*., crystallinity, monomer composition or stereoregularity. Considering these factors, some studies have shown that when the biopolymer is subjected to degradation under anaerobic sewage, degradation is accomplished in about 6 weeks, in soil environments in about 75 weeks and about 350 weeks in seawater (reviewed in Kessler *et al*., 2001; Loo *et al*., 2007; reviewed in Rehm, 2010).

PHAs can be produced in different chain lengths, there are short-chain length PHAs (scl-PHAs) such as poly-3-hydroxybutyrate (PHB) and poly-3-hydroxybutyrate-co-3-hydroxyvalerate [P(3HB-co-3HV)], which are composed of

2–5 carbon-chain-length monomer units; medium-chain-length PHAs (mcl-PHAs) such as poly-3-hydroxybutyrate-co-3-hydroxyhexanoate [P(3HB-co- 3HHx)] of 6-14 carbon-chain-length monomer units; and long-chain-length PHAs (lcl-PHAs) which are composed of more than 14 carbon-chain-length monomer units.

As reviewed in Brigham *et al*. (2012), some of the applications suitable for PHAs are directed for domestic use such as linings of milk cartons and diapers, and due to its biodegradability, PHAs have been also proposed for commercial use in garbage bags, so that they can be decomposed together with trash as compost. Other applications are fishing nets, as they can be easily degraded when lost in seawaters. As reviewed in Kessler *et al*. (2001), PHAs can be used as well for agricultural applications on delivery devices for fertilisers, nutrients, insecticides, and herbicides, as biodegradable mulch films and coatings for fertiliser pellets (reviewed in Brigham *et al*., 2012). As reviewed in Jajesniak *et al*. (2014), PHBs have also been used in the food industry for food packaging material due to its resistance to UV radiation, O_2 impermeability, and water resistance capabilities. In the medical field, PHA has been used for meshes and sutures due to its flexibility, durability, resorbability and tensile strength. The medical applications are still under testing, but expectations are high due to PHA biocompatibility. Both PHB and P(3HB-*co*-3HV) from *C. necator* H16 origin resulted highly compatible at a cell, tissue and macroorganism level, and can also be used in contact with blood (reviewed in Brigham *et al*., 2012).

Among the different chain length PHAs, PHB is one of the most pursued biomanufacturing products due to it can be transformed into others value-added chemicals such as acetoacetic acid or acetone; at the same time, acetone can be used for the synthesis of ketene, methyl methacrylate or diacetone alcohol. As reviewed in Yu (2014), propylene is another compound that can be obtained from PHB by thermal decarboxylation, which has reached high demands globally with a production of approximately 50 million tons per year (**Figure 1.7**).

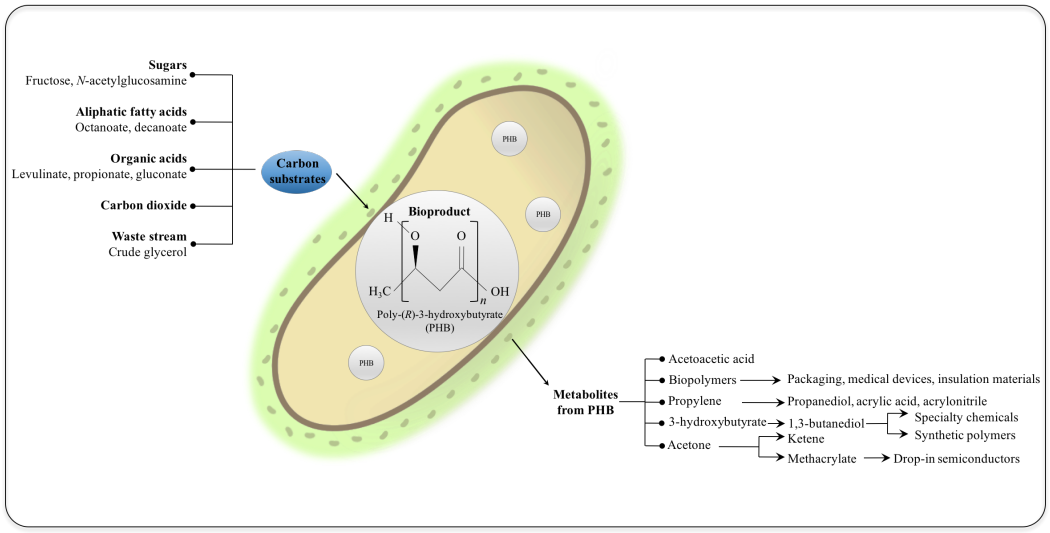


Figure 1.7 Metabolites derivable from polyhydroxybutyrates (PHB) (Adapted from Johnson et al., 2015).

Due to these potentials, diverse engineering strategies have been developed to optimise PHA biosynthesis in *C. necator* H16 and related species. The three genes responsible for PHB synthesis in *C. necator* H16 are encoded in the *phaCAB* operon, where *phaC* encodes for PHA synthase, *phaA* encodes for β-ketothiolase and *phaB* for acetoacetyl-coA reductase, being acetyl-CoA the primary metabolite in the metabolic pathway for PHB synthesis. As reviewed in Kessler *et al*. (2001), the *phaCAB* operon is constitutively expressed, but the PHB synthesis is regulated at enzymatic and transcription level. Therefore, strategies such as fed-batch in two-stage cultivation are employed to regulate the proportion of precursor substrates that enable both cell growth and PHB accumulation. The first stage oversees cell growth, which is performed under balanced growth conditions; and the second stage is in charge of product accumulation, performed under unbalanced growth conditions, usually under nitrogen limitation, which provokes the downregulation of TCA enzymes, mainly citrate synthase, as well as the reduction of acetyl-CoA flux used for cell growth (reviewed in Byrom, 1987; Jung *et al*., 2000).

As reviewed in Byrom (1987 and 1992), for the synthesis of scl-PHA copolymers, propionate has been used as a precursor co-substrate to produce P(3HB-*co*-3HV), due to its structural similarity with 3-hydroxyvalerate (3HV). Although some

challenges arise from the use of propionate when high concentrations are used, as this compound can be toxic to the cell (Steinbüchel *et al*., 2003), but at the same time regulations such as the fed-batch two-stage strategy can significantly decrease the negative impact of propionate toxicity (Lee *et al*., 1995; Grothe *et al*., 1999). There are some other alternative suitable precursors for 3HV monomer production, such as valerate (a direct 3HV precursor), levulinic acid, and some fatty acids as heptanoic and nonanoic acids (Steinbüchel *et al*., 2003). The molar fraction of the P(3HB-*co*-3HV) final product will depend on the choice and proportion of the 3HV precursors, along with the cultivation conditions, which will impact directly with the physical properties of the bioplastic (Choi *et al*., 2003; Albuquerque *et al*., 2007).

The copolymer P(3HB-*co*-3HHx) is an mcl-PHA that has been produced in engineered strains of *C. necator* H16 as well; Dennis *et al*. (1998) co-expressed *phaC1* and *phaB* genes from *C. necator* H16 in a PHA-negative mutant strain of *C. necator* H16, which showed accumulation of this copolymer using chain fatty acids as a carbon source. Another approach expressed heterologously the *ccr* gene (crotonyl-CoA reductase) from *Streptococcus cinnamonensis* and the *phaJ* and *phaC* genes from *Aeromonas caviae* (encoding for (R)-specific enoyl-CoA hydratase ((R)-ECH) and PHA synthase respectively) in a PHA-negative mutant strain of *C. necator* H16 for the production of P(3HB-*co*-3HHx) as well, where (R)-3HB-CoA is converted to crotonyl-CoA by the expression of the (R)-ECH enzyme and subsequently reduced by crotonyl-CoA reductase. PHA synthase then catalyses the copolymerization of 3HB-CoA and 3HHx-CoA monomers, ending with the P(3HB-*co*-3HHx) copolymer (Fukui *et al*., 2002). Different sustainable feedstocks such as plant oils (*e.g*., palm kernel oil, palm oil, soya bean oil) have been used for mcl-PHAs production with production of up to 87% (w/w) of the CDW (Kahar *et al.*, 2004; Loo *et al*., 2005; Riedel *et al.*, 2012). In **Table 1.3**, a comparison of the productivity of PHB and P(3HB-*co*-3HV) using different carbon souces is shown, where alternative carbon sources such as glucose, glycerol and waste glycerol were used for PHB production with mutant or engineered strains of *C. necator*.

Table 1.3 Production of PHB and P(3HB-co-3HV) from different carbon sources in *C. necator*.

Product	Carbon sources	Strain/Cultivation strategy	[PHA] (g/L)	Productivity (g/L/h)	References
PHB	CO_2	*C. necator* ATCC 17697[T] / Continuous one-stage cultivation with a stirred-tank fermenter.	61.9	1.55	Taga *et al*., 1997
	Glucose	*C. necator* NCIMB 11599 / Fed-batch cultivation.	121.0	2.40	Kim *et al*., 1994
		C. necator NCIMB 11599 / Fed-batch cultivation.	139.0	3.10	Shang *et al*., 2003
		C. necator DMS545 / Fed-batch two-stage cultivation.	125.0	2.03	Mozumder *et al.*, 2014
	Glycerol	*C. necator* DMS545 / Fed-batch cultivation.	51.2	1.52	Cavalheiro *et al*., 2009
	Waste glycerol	*C. necator* DMS545 / Fed-batch cultivation	65.6	1.36	Cavalheiro *et al*., 2009
P(3HB-*co*-3HV)	Glucose and valerate	*C. necator* NCIMB 11599 / Fed-batch two-stage cultivation.	90.4	1.81	Lee *et al*., 1995
	Glucose and propionic acid	C. *necator* NCIMB 40529 / Alternating feeding of carbon substrates in fed-batch cultivation.	65.5	1.60	Madden *et al*., 2000
		C. necator DMS545 / Optimised feeding strategy of substrates in a two-stage fed-batch cultivation. Stepwise reduction of propionate to glucose ratio to avoid propionate inhibition.	40.8	0.74	Du *et al*., 2001

Adapted from Johnson *et al.*, 2015.

1.5.2 Cyanophycin

As reviewed in Reinecke *et al*. (2009), other biopolymers can be produced in *C. necator* H16. Cyanophycin is a protein-like polyamide that is naturally produced in a variety of *Synechocystis* and *Anabaeana* spp.; it has been reported to be produced as well in recombinant strains of *C. necator* H16 at high product titters with the heterologous expression of *cphA* and *cphA1* that encode for the cyanophycin synthetases from *Synechocystis* and *Anabaeana* spp. in combination with metabolic engineering strategies, where 40% (w/w) of the CDW of cyanophycin has been achieved in 30-500 L fermenters (Voss *et al*., 2006). Due to cyanophycins are composed as well of polyaspartate backbone and arginine residues, the autotrophic cultivation with $^{13}CO_2$ and $^{15}NH_4Cl$ in *C. necator* H16 has been performed in order to produce $^{13}C/^{15}N$-labeled cyanophycin, from which $^{13}C/^{15}N$ arginine can be derived, which is a compound demanded for stable isotope (SI)-labeled biomolecule for quantitative proteomics (Lüte *et al*., 2012).

1.5.3 Biofuel

As reviewed in Johnson *et al*., 2015, biofuels represent another vital chemical for the bioeconomy industry, therefore, the understanding of the biochemical pathways of *C. necator* H16 is fundamental to the design of the routes that could lead to biodiesel production, and that includes the PHA biosynthetic pathway, due to acetyl-CoA is the central metabolite for PHA production, which is derived from pyruvate decarboxylation. PHA-negative *C. necator* H16 strains have a pyruvate-leaky phenotype when they are cultivated under PHA biosynthesis inducing conditions (Schlegel *et al*., 1970), hence, a strategy designed to produce branched-chain amino acids (such as valine and leucine), and alcohols, is based on the diversion of excess pyruvate flux from carbon metabolism to other biosynthetic pathways such as α-keto-acids.

Isobutanol, a branched-chain amino acid, is a promising biofuel that has been considered as an ethanol substitute in gasoline blends. Isobutanol and 3-methyl-1-butanol production have been produced with a recombinant strain of *C. necator* H16 by the overexpression of the native branched-chain amino acids leucine and valine synthesis genes, which leads to the conversion of pyruvate to α-keto acids; along with the overexpression of *kivD* gene (encoding for ketoisovalerate decarboxylase) from *Lactococcus lactis*, which is in charge of the α-keto acids decarboxylation; where the enzyme isobutyraldehyde dehydrogenase diverts the carbon flux from pyruvate to the biosynthesis of isobutanol and 3-methyl-1-butanol. Further optimisation of this approach was developed with the elimination of the genes *ilvE, bkdAB*, and *ace*, that encode for the three potential carbon sinks to reduce product consumption and toxicity (Lu *et al*., 2012). Another approach for production of isobutanol was reported by Li *et al*. (2012), where *C. necator* H16 was engineered to produce isobutanol and 3-methyl-1-butanol from CO_2 in an electro-bioreactor that uses only CO_2 as carbon source and electricity as the energy input.

1.5.4 Methyl ketones

Other value-added chemicals that can be obtained from *C. necator* H16 are methyl ketones, which can be used as blending agents for diesel fuels, or are used as well as chemical intermediates in the fragrance and flavour industries (Goh *et al*., 2012; Müller *et al*., 2013). The β-oxidation pathway (or fatty acid degradation pathway) is responsible for the breakdown of fatty acids into acetyl-CoA, based on this premise Müller *et al.* (2013) engineered the fatty acid catabolism of *C. necator* H16 to produce fatty-acid derived methyl ketones, by overexpressing TesA thioesterase. This approach was achieved with the replacement of the acyl-CoA dehydrogenase (*FadE*) from the native β-oxidation pathway of *C. necator* H16 by overexpressing the heterologous acyl-CoA oxidase from *Micrococcus luteus*, and FabB from *E. coli*, in combination with the deletion of the *fadA* gene in order to overproduce β-ketoacyl-CoAs; followed by the overexpression of thioesterase FadM from *E. coli*, which converts β-ketoacyl-CoAs into β-keto acids, from which decarboxylation ends up with methyl ketones.

1.5.5 Methyl citrate

2-methyl citrate (2MC) is a building block chemical used for the synthesis of various pharmaceuticals. Also, the compound is an essential intermediate for the synthesis of polymer plasticisers, and it is involved in the synthesis of citric acid. One of the pathways that compete for the propionyl-CoA flux is the 2MC cycle; propionyl-CoA is converted to succinate by the 2MMC enzymes, which are encoded in a gene cluster in the prp loci. The accumulation of 2MC was achieved in *C. necator* H16 by the inactivation of the *acnM* gene (encoding for 2-methyl-cis-aconitic acid dehydratase), which converts 2MC to 2-methyl-cis-aconitic acid, and with the insertion of an additional *prpC* gene, which is in charge of the production of 2MC from propionyl-CoA that generates a propionate-negative phenotype strain (Ewering *et al*., 2006).

1.5.6 Feluric acid

Vanillin, a compound that is used as a flavour additive in food and beverage industries, is synthesised by its precursor feluric acid, which can be produced in engineered strains of *C. necator* H16. A recombinant strain of *C. necator* H16 for industrial-scale production was designed to produce feluric acid, where the genes *ehyAB, calA*, and *calB* (which encode for eugenol hydroxylase, coniferyl alcohol dehydrogenase, and coniferyl aldehyde dehydrogenase, respectively) from *Pseudomonas* sp. strain HR199 -involved in the first reactions of eugenol catabolism-, were expressed in *C. necator* H16. Eugenol is an inexpensive compound that provides the biotransformation of feluric acid to vanillin; therefore, Overhage *et al*., (2002) used the recombinant *C. necator* H16 strain to convert eugenol to feluric acid, by the introduction of those three genes. The result was that the *C. necator* H16 recombinant strain (pBBR1-JO2ehyABcalAcalB) was capable of holding a five-fold-higher eugenol concentration than other bacteria, achieving effective biotransformation.

1.5.7 Potential bioproducts: pimelic acid for biotin production

1.5.7.1 Importance of pimelic acid and biotin

There are other exciting bioproducts such as pimelic acid and biotin that have been produced in other microorganisms; these bioproducts could potentially be produced as well in *C. necator* H16, with the advantage of using renewable feedstock or cheap substrates which have been used for bioproduction previously in this strain. Biotin is added in many foods and feed products, cosmetics for hair, skin, and nails, which has created a world market of hundreds of tons per year, where 1 kg of biotin costs around US$ 1000 (Finkenwirth *et al*., 2014). To date, most of the biotin is produced by chemical synthesis, previously, it was produced following the process developed by Goldberg *et al*. (1949) were biotin was synthesised in more than ten steps using succinic acid or D-mannose as precursors, and toxic compounds such as bromide were used for molecular

activations; lately, this process has been modified, but still generates considerable amounts of chemical waste such as tetrahydrofurane and toluene, ethyl acetate, small amounts of phenyl chloroformate, among other toxic compounds; the late patented biotin synthesis also needs high energy input, and it takes more than ten steps to be synthesised (reviewed in Streit *et al*., 2003).

For these reasons, and due to the environmental burden because of the accumulation of chemical waste, the development of cost-effective biotin-overproducing microorganisms is desirable. The production of biotin in microorganisms should reach production titres of at least 1 g biotin/L and use a cheap substrate in order to be cost-effective (reviewed in Streit *et al*., 2003), therefore *C. necator* H16 represents a good candidate as a biotin-overproducing microorganism, as it has been shown that it can use renewable feedstock as well as to be a versatile microorganism that can be engineered to obtain high production titres of bioproducts. The main strategies that have been used for biotin overproduction in different microorganisms to date are: 1) Selection of improved biotin-overproducing strains that have been obtained through chemical mutagenesis (evolutionary engineering), 2) recombinant strains that are capable of overproducing biotin *via* the expression of extra copies of biotin-biosynthesis genes (metabolic engineering) and, 3) the combination of both evolutionary and rational design engineering strategies.

One of the strategies used to improve biotin production in bioprocess engineering is the addition of pimelic acid in the fermentation process. Pimelic acid is a seven carbon α,ω-dicarboxylic acid, and it is an essential precursor in the biotin biosynthesis, although its synthesis in wild-type strains is low. The use of pimelic acid as an additive for biotin production represents a high-cost in the production chain; therefore alternatives for biosynthesis of pimelate are a fundamental requirement in order to reduce the production cost of biotin. It has been demonstrated that the production of dethiobiotin when pimelic acid is added in the fermentation, increases by ten-fold than without the addition of pimelate (Berkovitch *et al*., 2004).

The early steps of the biotin biosynthesis pathway of *C. necator* H16 have not been investigated yet, the only information available and published to date is the existence of the *bioFADB* operon described by Pohlmann *et al.* (2006) in the genome sequence of the strain, where it was found that *C. necator* H16 does have the standard *bioFADB* genes clustered in a single operon, these genes are involved in the final steps of biotin biosynthesis. On the other hand, there is a predicted *bioC* gene in a different location of the *bioFADB* operon in *C. necator* H16 (Rodionov *et al*., 2002), although no studies have been performed to date to confirm the function of the gene. In order to understand how the pimelate and biotin biosynthesis could be improved in *C. necator* H16, it is important to understand the biotin biosynthesis pathways that have been already elucidated in other microorganisms.

1.5.7.2 Catalytic role of Biotin

Biotin (also known as vitamin H and B7) is a fascinating sulphur-containing cofactor which functions when it is covalently attached to crucial metabolic enzymes; it is highly involved in the central metabolic pathways of gluconeogenesis and fatty acid synthesis of eukaryotic and prokaryotic cells. Biotin mediates the transport of CO_2 and can be divided into three different classes depending on their function: carboxylases (Class I, where CO_2 is transferred to an acceptor molecule), decarboxylases (Class II, where CO_2 is released as bicarbonate) or transcarboxylases (Class III, where a reversible transfer of a carboxyl group occurs from one compound to another) (reviewed in Knowles, 1989; Rodionov *et al*., 2002; Zhang *et al*., 2011; Grover *et al*., 2012). The number of biotin-dependent protein ranges from one to five in different microorganisms (Cronan *et al*., 2000) and in most cases, enzymes belong to biotin-dependent enzymes catalyse CO_2 fixation (Class I) (Grover *et al*., 2012). Without the attachment of biotin, biotin-dependent enzymes are not catalytically active, and therefore they cannot fulfil their metabolic roles.

One of the most important biotin-dependent enzymes is acetyl-CoA carboxylase, which is in charge of catalysing the first step in the fatty acid biosynthesis

pathway; acetyl-CoA carboxylase catalyses the ATP-dependent transfer of a carboxyl group from carbonate to acetyl-CoA to produce malonyl-CoA. The fatty acid biosynthesis pathway is essential for bacterial cell membrane maintenance (Polyak *et al*., 2012; Soares da Costa *et al*., 2012), hence, if biotin is inhibited –for instance, by biotin analogues-, cell growth could be compromised. Other biotin-dependent enzymes are pyruvate carboxylase, propionyl-CoA carboxylase, methylcrotonyl-CoA carboxylase, geranoyl-CoA carboxylase, oxaloacetate decarboxylase, methylmalonyl-CoA decarboxylase, transcarboxylase and urea amidolyase, which are found in prokaryotes or eukaryotes (Jitrapakdee *et al*., 2003) and participate in different central metabolic pathways as gluconeogenesis, amino acid metabolism, lipogenesis and energy transduction (Delli-Bovi *et al*., 2003).

The structure of biotin consists of a bicyclic ring that is fused to a valeryl side chain; the bicyclic rings are composed of an imidazol (ureido) ring and a sulphur-containing (tetrahydrothiophene) ring (DeTitta *et al*., 1976). When biotin is active, it is covalently attached to enzymes through an amide bond between the carboxyl group of biotin and a specific lysine residue in the ε-amino group of a conserved protein domain of about 80 residues (Lane *et al*., 1964). The valeryl side chain helps to extend the bicyclic rings away from the lysine residue, and the ureido ring N8 nitrogen is in charge of carrying the CO_2 moiety and forming N-carboxybiotin (reviewed in Knowles, 1989; Attwood *et al*., 2002).

Biotin-producing systems are limited to microbes, fungi, and plants; mammals are not capable of producing biotin, they obtain biotin from the intestinal microflora or from the diet. Humans and animals require hundreds of micrograms of biotin per day, these amounts of biotin have been provided mainly by chemical synthesis (reviewed in Streit *et al*., 2003; Lin *et al*., 2010, Cronan *et al*., 2011; Lin *et al*., 2011). Nevertheless, as mentioned before, the chemical synthesis of biotin is followed by environmental burden, and it is a lengthy process; hence new alternatives for biotin biosynthesis using microorganisms as a cell factory have become an attracted strategy for the biotechnological community.

1.5.7.3 Biotin biosynthesis in prokaryotes

Although biotin is essential, the knowledge of its biosynthesis remains fragmentary in many microorganisms. Comparative genomic analyses have indicated that the biotin biosynthesis pathway is largely conserved and can be divided into two stages: the first stage is the synthesis of the pimelate moiety -studies reported to date suggest that in most of the microorganisms the carbon atoms of biotin are derived from pimelic acid (Ifuku *et al*., 1994; Sanyal *et al*., 1994)-, and the second stage is the assembly of the bicyclic ring of biotin. The biotin biosynthesis pathway is best understood in some prokaryotic microorganisms such as *E. coli*, *Bacillus subtilis* and *Bacillus sphaericus,* these investigations have been performed using combined genetics, biochemical and complementation studies (Otsuka *et al*., 1978; Gloeckler *et al*., 1990; Kiyasu *et al*., 2001; Rodionov *et al*., 2002). As reviewed in Streit *et al*. (2003), other studies have analysed in detail the biotin biosynthesis pathway of other microorganisms such as *Kurthia* sp, *Mesorhizobium loti*, *Methylobacillus flagellatum*, *Corynebacterium glutamicum*, and *Sinorhizobium meliloti*.

The classic models of *E. coli* and *Bacillus* spp. show orthodox late steps of biotin biosynthesis –those involved in the assembly of the fused rings- which mainly differ in the first step of biosynthesis (**Figure 1.8**), where *Bacillus* spp. use *bioW* (that encodes for pimeloyl-CoA synthase) and bioX (acyl carrier protein) (Bower *et al*., 1996a; reviewed in Lemoine *et al*., 1996; Rodionov *et al*., 2002) to incorporate exogenous pimelate to synthesise pimeloyl-CoA, which afterwards undergoes to the biotin biosynthesis pathway; in addition, *B. subtilis* is thought to have an alternative route, where it has been proposed to use *bioI* (which encodes for cytochrome P450) to provide *de novo* synthesis of pimeloyl-ACP by the catalysis of oxidative C-C bond cleavage of long-chain acyl-ACPs for pimelic acid formation (Stock *et al*., 2000; Rodionov *et al*., 2002). In *B. subtilis*, the genes involved in the conversion of pimelic acid to biotin are found within a single operon as *bioWAFDBI* (Perkins *et al.*, 1996).

On the other hand, in *E. coli*, pimeloyl-CoA is not synthesised from pimelic acid but from acetate via acetyl-CoA (Ifuku *et al*., 1994); recent studies revealed that

the products of *bioC* (annotated as S-adenosyl-L-methionine (SAM) dependent methyltransferase) and *bioH* (which encodes for pimeloyl-ACP methyl ester esterase) are required for the synthesis of pimeloyl-CoA. BioC and BioH do not catalyse the synthesis of pimelic acid directly, but they provide the means to use fatty acid synthesis for the assembly of the pimelate moiety. BioC methylates the ω-carboxyl group of malonyl-CoA (or ACP) by transferring a methyl residue from SAM, this mimics the methyl ends of regular fatty acyl chains, hence, this atypical substrate is recognised by fatty acid enzymes (FabG, FabZ, FabI, FabB, FabF); the formed malonyl-CoA methyl ester enters the fatty acid pathway and undergoes two reiterations of elongation in the fatty acid synthesis and gives pimeloyl-ACP methyl ester, which is then hydrolysed to pimeloyl-ACP by BioH to terminate chain elongation and to liberate the ω-carboxyl group of biotin (Lin *et al*., 2010).

After pimelate synthesis, the biosynthesis pathway is similar in *E. coli* and *Bacillus* spp., with the products of *bioF*, *bioA*, *bioD*, and *bioB*. In *E. coli*, the biotin genes are not found in a single operon, it has a *bioBFCD* operon which is located divergently with *bioA* and a single *bioH* gene in a different location in the chromosome; if BioC is overexpressed and active, this will cause the elevated levels of malonyl-ACP methylated species, which will block fatty acid synthesis due to lack of malonyl-ACP, and that would reduce cell growth (Lin *et al*., 2012), therefore if *bioC* is overexpressed in a biological system, a proper regulation of its expression must be considered.

The *bioC* gene is widely found in bacteria, while *bioH* is not found in many *bioC*-containing bacteria as it is in *E. coli*; instead, other genes have been found to complement *bioH* function such as the *bioZ* gene from *Mesorhizobium loti* or the *bioG* gene found in some proteobacteria (Sullivan *et al*., 2001; Rodionov *et al.*, 2002). In bacteria, biotin is needed only in trace quantities, specifically in *E. coli*, where only few hundred molecules per cell are required for growth (concentrations of only a few nanomolar), due to this reason, biotin is expressed at shallow levels (Cronan, 2001; Lu *et al*., 2007).

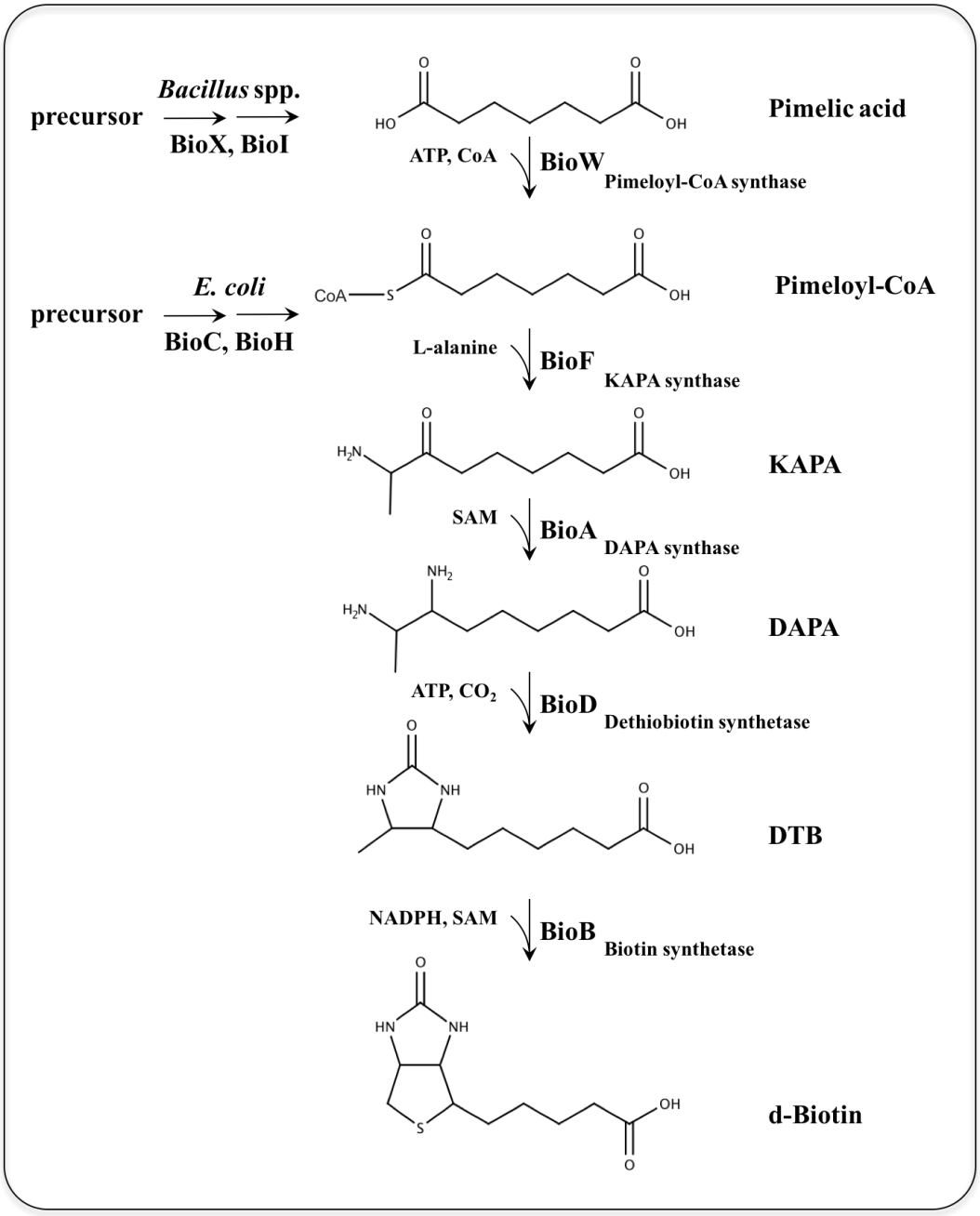


Figure 1.8 The orthodox last steps of the biotin biosynthesis pathway for the assembly of the bicyclic rings from studies of E. coli and Bacillus spp. Abbreviations: KAPA, 7-keto-8-aminoperlargonic acid; DAPA, 7,8-diaminopelargonic acid; DTB, dethiobiotin; SAM, S-adenosyl-L-methionine.

As reviewed in Cronan (2014), biotin must be covalently attached to metabolic enzymes to perform its role in cellular enzymology; free biotin is not physically useful, although it plays an indirect regulatory role. In *E. coli*, the regulation of biotin biosynthesis is carried on by the bifunctional protein BirA, which acts as a transcriptional repressor of the biotin operon as well as a biotin-protein ligase that catalyses the covalent attachment of biotin to biotin-dependent enzymes (Delli-

Bovi *et al*., 2010), where the *bioH* gene that is not located within the *bio* operon, is not regulated by BirA (Barker *et al*., 1980; Koga *et al*., 1996; reviewed in Cronan, 2014). Predicted BirA-binding sites are well conserved in a variety of eubacteria and archaeal genomes. The BirA protein of *B. subtilis* has a similar structure to the BirA of *E. coli*, it also acts as a repressor of the *bioWAFDBI* operon (Bower *et al*., 1996a). It has been shown that BirA is the protein most widely distributed biotin-related gene among bacteria (Rodionov *et al*., 2002).

As reviewed in Streit *et al*. (2003), many investigations have been done to analyse the genetics behind the biotin biosynthesis pathway in other microorganisms, apart from the more detailed investigations in *E. coli*, *B. subtilis*, and *B. sphaericus*, where experimental designs to identify the biotin biosynthesis genes included mainly complementation studies, or a combination of complementation studies and random mutagenesis to identify the genes related to the biotin biosynthesis pathway.

After the studies related to the biotin biosynthesis pathway in different microorganisms, it has been observed that the *bioFADB* operon is universally conserved among different prokaryotes, as can be observed in **Figure 1.9**, while the *bioFADB* operon is conserved in gram-positive, and gram-negative bacteria, the genes committed to pimelate synthesis vary among them, where some of the species of the gram-negative bacteria appear to have the genes *bioC* and *bioH* committed to pimelate synthesis (reviewed in Streit *et al*., 2003).

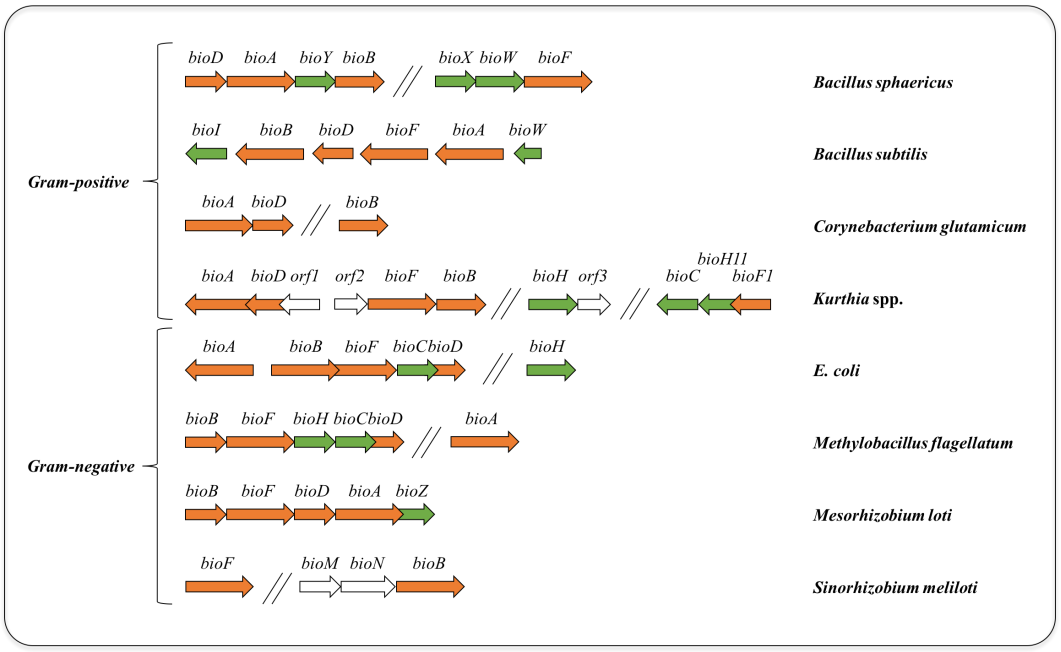


Figure 1.9 Physical organisation of biotin synthesis genes identified and tested for functionality in gram-positive and gram-negative microorganisms. In the illustration, different transcriptional units are separated by //. Genes linked to the assembly of the fused heterocyclic rings of biotin are indicated as orange arrows, the genes most likely involved in pimeloyl-CoA synthesis are indicated as green arrows, and the genes not directly linked to biotin biosynthesis are indicated as white arrows. Information of the genes involved in the biotin biosynthesis was extracted from GenBank (Adapted from Streit et al., 2003).

As mentioned before, one of the most common strategies to either understand the biotin biosynthesis pathway or to improve biotin production in a variety of microorganisms has been the use of directed evolution *via* random mutagenesis for all cell genome, as this strategy allows to cover the entire genome of the strains, as in many cases, the biotin biosynthesis genes of the microorganisms have not been predicted or identified.

1.5.7.4 Alternative pathways for pimelic acid production

One of the most intriguing biotin biosynthesis pathways is the one encoded by two *Desulfovibrio* spp. Rodionov *et al.* (2004) studied the metal-reducing genera in the δ subgroup of proteobacteria to compare genome sequences of seven representatives in order to investigate the regulatory and genetic factors in pathways that are involved in the biosynthesis of cofactors and building blocks. All δ-proteobacteria studied for the biotin cofactor showed to have the *bioFADB* genes for *de novo* biosynthesis of biotin, as well as the BirA bifunctional protein,

although the initial steps of biotin biosynthesis differed among these species (**Figure 1.10**). The *Geobacter* species have a *bioC-bioH* gene pair as *E. coli*, for the synthesis of pimeloyl-CoA, *Desulfuromonas* species have the same pair but also the *bioW* gene, which represents two different pathways for pimeloyl-CoA synthesis, and *Desulfotalea psychrophila* is predicted to use a *bioC-bioG* gene pair (Rodionov *et al*., 2002).

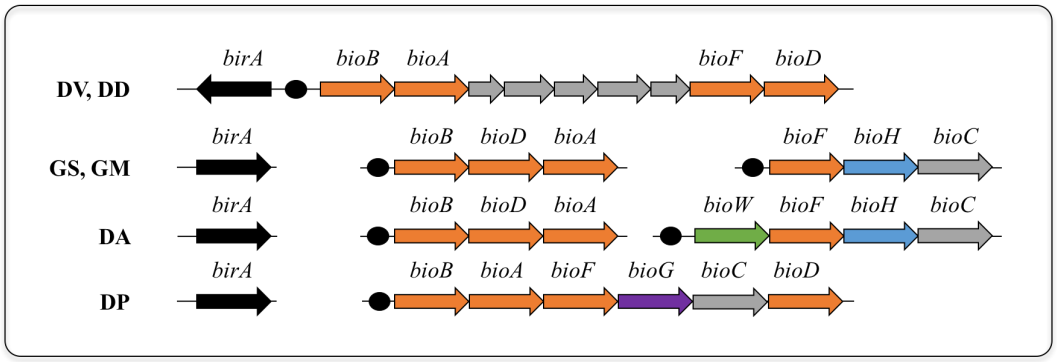


Figure 1.10 Genomic organisation of the biotin biosynthesis genes and regulatory elements of metal reducing proteobacteria. Abbreviations: DD, Desulfovibrio desulfuricans; DV, Desulfovibrio vulgaris; GS, Geobacter sulfurreducens PCA; GM, Geobacter metallireducens; DA, Desulfuromonas spp.; DP, Desulfotalea psychrophila (Adapted from Rodionov et al., 2004).

Interestingly, Rodionov *et al*. (2004) revealed that *Desulfovibrio vulgaris* (DV) and *Desulfovibrio desulfuricans* (DD) have an extended biotin operon within the standard *bioFADB* with five new genes that are related with the fatty-acid biosynthetic pathway (although only one of the genes within the extended biotin operon of DD (*DD5*) is annotated as a hypothetical protein) (**Figure 1.11**), these new genes –which are not present in other δ-proteobacteria-, are homologs of acyl carrier protein (ACP), 3-oxoacyl-(ACP) synthase, 3-oxoacyl-(ACP) reductase and hydroxymyristol-(ACP) dehydratase, due to this reason, they concluded that these genes might be functionally related to the biotin pathway. Hence, they hypothesised that these genes could encode a novel pimeloyl-CoA biosynthesis pathway, as the known genes for the biotin biosynthesis pathway for δ-proteobacteria: *bioC*, *bioH*, *bioG*, and *bioW* are not present in *Desulfovibrio* spp.

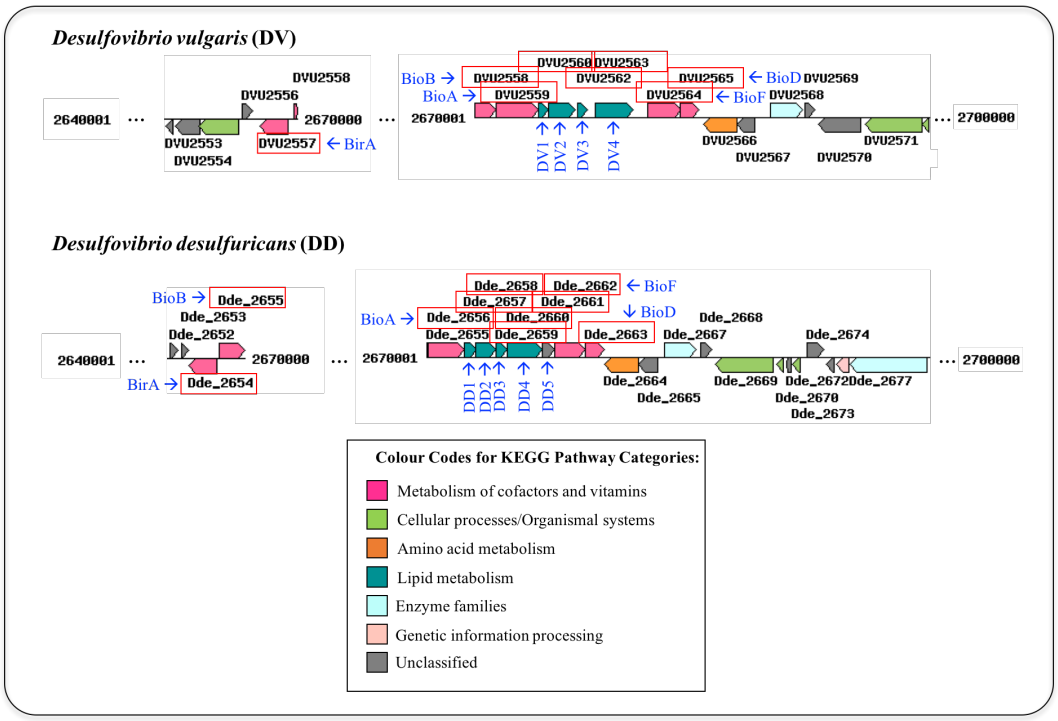


Figure 1.11 Biotin biosynthesis genes in Desulfovibrio vulgaris (DV) and Desulfovibrio desulfuricans (DD). The genome maps of the genes involved in the biotin biosynthesis pathway of both microorganisms were extracted from KEGG. The inset shows the colour codes used by KEGG to describe the function of the open reading frames; as can be observed, all the genes identified for biotin biosynthesis in pink colour for both microorganisms have been categorised under the section of "Metabolism of cofactors and vitamins", while the genes within the extended biotin operons of DV and DD are marked in aqua colour for both microorganisms under the section of "Lipid metabolism". DV1, DV2, DV3, and DV4 correspond to the new genes found within the bioFADB extended biotin operon of Desulfovibrio vulgaris (DV); DD1, DD2, DD3, DD4 and DD5 correspond to the new genes found within the bioFADB extended biotin operon of Desulfovibrio desulfuricans (DD).

1.5.7.5 Biotin biosynthesis in *C. necator* H16

As mentioned earlier, the second stage of biotin biosynthesis in charge of the assembly of biotin rings is encoded in the universal *bioFADB* biotin operon, which is found in the chromosome 1 of *C. necator* H16 (Pohlmann *et al.*, 2006), but the early steps of the biotin biosynthesis committed to pimelate synthesis have not been elucidated yet (**Figure 1.12**), therefore, the study of the biotin biosynthesis pathway of *C. necator* H16 represents a great opportunity and challenge to scientists to define the genes that are involved in the formation of the pimeloyl-CoA moiety, with the subsequent possibility of engineering *C. necator* H16 either to use it as a chassis for pimelate and biotin production.

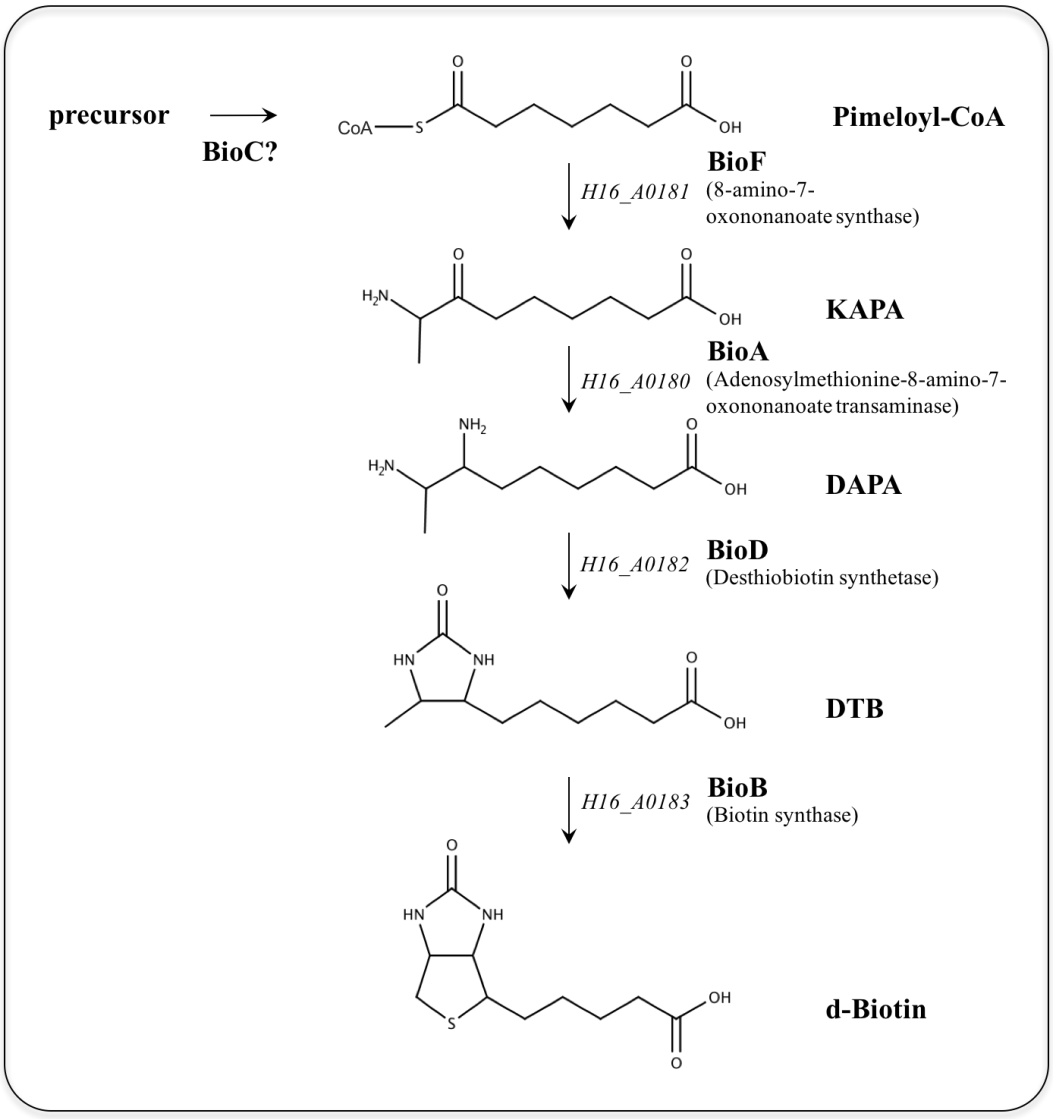


Figure 1.12 Key enzymes of biotin biosynthesis in C. necator H16. A putative bioC gene has been predicted in chromosome 1 of C. necator H16, which could be committed to pimelate synthesis.

Apart from the information provided after the genome sequencing of *C. necator* H16, Rodionov *et al*. (2002) predicted by comparative genomics and phylogenetic analysis a *bioC* gene found in chromosome 1of *C. necator* H16, where the sequence identity between the potential *bioC* in *C. necator* H16 (originally annotated as "Predicted malonyl-CoA O-methyltrasnferase) and the *bioC* of *E. coli* K-12 MG1655 (originally annotated as SAM-dependent malonyl-CoA O-methyltrasnferase) is of 30 %, with an overlap of 239 amino acids. This predicted *bioC* gene of *C. necator* H16 ($bioC_{H16}$), unlike the *bioC* of *E. coli* ($bioC_{EC}$) is not

found within the *bioFADB* biotin operon of *C. necator* H16, but in a distinct location in the chromosome 1 (about 150,000 bp downstream the *bioFADB* biotin operon), no other open reading frame under the category of "Metabolism of cofactors and vitamins" (pink colour arrows in **Figure 1.13** for *C. necator* H16) is indicated near the *bioFADB* biotin operon in *C. necator* H16.

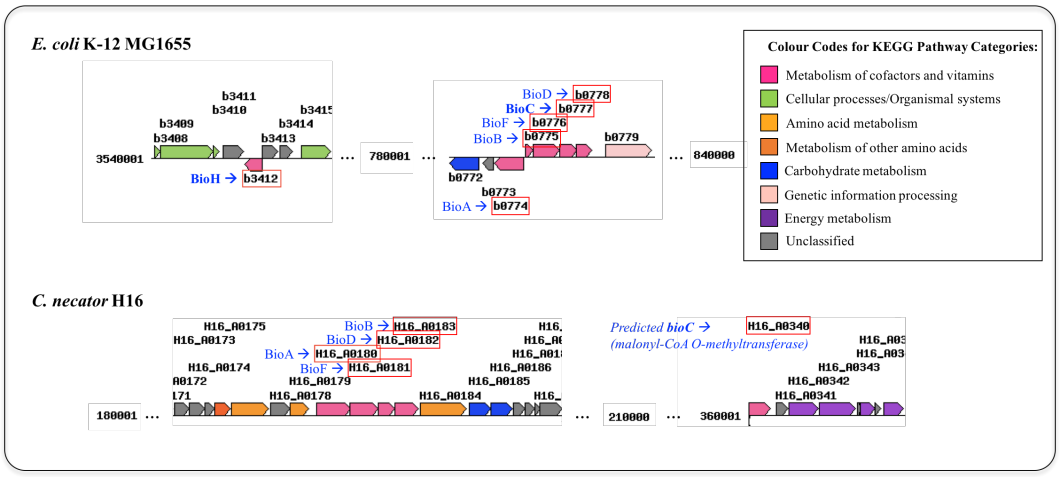


Figure 1.13 Biotin biosynthesis genes in E. coli K-12 MG1655 and C. necator H16. The genome maps of the genes involved in the biotin biosynthesis pathway of both microorganisms were extracted from KEGG. The inset shows the colour codes used by KEGG to describe the function of the open reading frames; as can be observed, all the genes identified for biotin biosynthesis in pink colour for both microorganisms have been categorised under the section of "Metabolism of cofactors and vitamins", which suggests that $BioC_{H16}$, could be involved in the biotin biosynthesis pathway.

Evolutionary engineering could be a possible strategy to use to understand more in detail the biotin biosynthesis pathway of *C. necator* H16. A whole genome random mutagenesis would expose the entire genome to mutagenesis and mutant strains could be selected after in an agar plate supplemented with a selection pressure –such as biotin analogues-, where it would be expected that only the mutants capable of growing in those conditions would be the mutants that have gone through genetic mutations in regions that could be related to the biotin biosynthesis pathway.

Once the best mutant strain had been selected, whole genome sequencing of the mutant strain could be carried out to compare the DNA sequences between the

mutant and the wild-type, which will help to identify the regions with genetic mutations, these regions could be further studied as they could be related to the biotin biosynthesis pathway, where perhaps genetic alterations on the predicted $bioC_{H16}$ gene could be found, or in a possible $bioH_{H16}$ gene (or a similar one such as *bioZ* or *bioG*) that could be identified in the strain, as it appears that in gram-negative microbes (as *C. necator* H16 and *E. coli*), the *bioC* and *bioH* (or a similar gene) genes are usually required for the first steps of biotin biosynthesis (reviewed in Streit *et al*., 2003).

1.5.7.6 Pimelate and biotin production using rational design and evolutionary engineering strategies

As reviewed in Streit *et al*. (2003), most of the recombinant biotin-overproducing bacterial strains reported to date have used forward engineering by employing cloning strategies that include the fusion of *bio* genes with strong promoters. Over-producing biotin recombinant strains have been constructed in different microorganisms such as *E. coli*, *B. subtilis*, and *Serratia marcescens* (Sakurai *et al*., 1995, Van Arsdell *et al*., 2005; Zhang *et al*., 2011), but still the production of biotin in these studies have not been able to yield enough biotin cost-effective yet.

Zhang *et al*., (2011) constructed a recombinant strain of *B. subtilis* for efficient pimelic acid synthesis with a maltose-inducible strong $P_{g/v}$ promoter inserted into the upstream of *bioI* and created a repeat of the cistron *bioI-orf2-orf3* –which are the genes involved in the pimelic acid biosynthesis pathway in *B. subtilis*- in *B. subtilis* chromosome DNA, where high level of production of pimelic acid was obtained with 4 times higher production fold compared to the single copy. A similar approach could be conducted in *C. necator* H16 once the genes committed to pimelate synthesis have been identified.

Sakurai *et al*. (1993) engineered *Serratia marcescens* Sr41 using evolutionary engineering by directed evolution *via* random mutagenesis; they isolated mutants of this strain resistant to the biotin analogue acidomycin (ACM, also known as actithiazic acid), this biotin analogue reported to inhibit the formation of D-biotin

from dethiobiotin by the biotin synthetase (BioB); up to 20 mg/L of biotin was produced with the mutant strains. *Serratia marcescens* Sr41 cells were mutagenised with *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (NTG), and mutated cells were grown in minimal medium agar plates supplemented with acidomycin.

On the other hand, Bower *et al*. (2001) constructed as well biotin analogue resistant mutants of *B. subtilis* that overproduced biotin, where cells were mutagenised with ethyl methanesulfonate (EMS) to give a 90 % killing. 5-(2-thienyl)-valeric acid (TVA) was used as a biotin analogue, which showed to inhibit cell growth in *B. subtilis* wild-type cells, while chemically mutagenised *B. subtilis* cells could grow in the presence of TVA. A different compound was also used to screen for biotin-overproducing mutants of *B. subtilis*: azelaic acid (AZ), which is a homologue of pimelic acid thought to be involved in the conversion of oleic acid to pimelic acid (Ohsugi *et al*., 1981). High concentrations of the AZ inhibited cell growth of the wild-type, which was reversed by the addition of biotin. According to experiments performed by Bower *et al*. (2001), AZ might act at the level of pimelyl-CoA synthetase (*bioW*) as a competitive inhibitor of pimelic acid, Bower *et al*. (2001) suggested that the mutants screened and selected in AZ were likely to have increased their capacity to produce or utilise pimelic acid probably with mutations found either in *bioI* or *bioW*. The mutant strains of *B. subtilis* selected produced from up to 0.3 mg/L of biotin.

Hoshino *et al*. (1999) also applied evolutionary engineering to produce biotin in high yields by fermentation using microorganism belonging to the genus *Kurthia*. *Kurthia* spp. mutant strains showed resistance to biotin analogues as ACM, TVA, α-methyldethiobiotin, amiclenomycin, bisnorbiotinol, among others, which have enabled the cell to accumulate large amounts of d-biotin in the culture broth, that can furtherly be recovered in excellent purity. Cells of *Kurthia* spp. were firstly randomly mutagenised with NTG, EMS, acridine orange, UV or X-rays; then cells were grown in biotin-free minimal medium supplemented with different biotin analogues (ACM, TVA or α-methyldethiobiotin), were mutants resistant to the different biotin analogues were obtained. The cultivation of *Kurthia* spp. takes about 2 to 4 days, were after that period of cultivation, biotin could be separated

from the culture broth and purified with yields of biotin 5 to 200 higher than the yields obtained from *Serratia marcescens* (Sakurai *et al*., 1993).

Another notable example was the one developed by Ifuku *et al*. (1993), where ACM and TVA were used as biotin analogues. *E. coli cells* were mutagenised with NTG chemical mutagen and then grown in two different biotin analogues TVA and ACM in the same agar plate, among the mutants selected a biotin productivity from 0.9 up to 4.5 mg/L was obtained. The mutations points of the biotin operon from the biotin-overproducing mutants of *E. coli* were analysed by DNA sequencing. One of the mutations was found in the operator overlapping the -10 region of the *bioB* promoter, and two other mutations were found before and after the initiation codon of the *bioB* gene; the first mutation was considered to have disrupted the operator structure, and thus, the promoter activity was enhanced, and the second two mutations were thought to have activated the translation efficiency.

Other scientists have used the same approach whereby the cells are randomly mutated either by chemical or physical mutagens and selected as biotin overproducing microorganisms under a biotin analogue pressure, but also, the isolated mutants are transformed to express heterologous *bio* operons mainly from *E. coli*. As reviewed in Streit *et al*. (2013), it is important to know that in addition to strain design, other factors contribute to higher biotin yields such as medium components and fermentation conditions.

1.6 Future perspectives

With increasing global attention on the environmental burden, approaches in biotechnology must consider the development of strategies that are sustainable for chemical production in bacteria; this potential can be fully exploited in *C. necator* H16 for biomanufacturing purposes. Understanding the biology of *C. necator* H16 is now easier to achieve since its complete genomic sequence annotation is already available, and the construction of a synthetic biology toolbox for *C. necator* H16 would allow the regulation and expression of genes that can

contribute to bioproduction. Expanding the range of utilisable substrates -mainly renewable substrates such as crude glycerol- would decrease production costs and make the bioprocess sustainable.

C. necator H16 remains a promising microbial cell factory for value-added chemical production. The integration of rational design and evolutionary engineering are key strategies to achieve a robust strain of *C. necator* H16 for chemical production.

1.7 Overall aim of the study

C. necator H16 has attracted biotechnological interest for the last decades due to its versatility and the relatively recent (2006) available genomic sequence, which aids to engineer the strain more easily. *C. necator* H16 is a non-pathogenic bacterium that awaits to be used as a cell factory. In this **Chapter 1**, some of the potential strategies and bioproducts that can be synthesised in the strain were mentioned and discussed; these strategies served as a basis and theoretical fundamental for the studies of this PhD project.

The overall aim of the study is *to engineer C. necator H16 with rational design and evolutionary engineering tools to explore the biotechnological potentials of the strain and fine-tune the properties of C. necator H16 for chemical production using waste stream as a feedstock.* The specific aims of each Chapter are listed below:

Chapter 2
Recombinant strain engineering: Development of a transformation method and a synthetic biology toolbox for metabolic engineering in Cupriavidus necator H16

- To optimise a transformation method by electroporation for *C. necator* H16.

- To construct a synthetic biology toolbox with inducible and constitutive promoters and other synthetic biology parts for metabolic engineering applications

Chapter 3

Metabolic engineering: Applicability of synthetic biology toolbox by the expression of heterologous extended biotin operons in *Cupriavidus necator* H16

- To demonstrate the applicability of the synthetic biology toolbox constructed in **Chapter 2** with the expression of operons driven by a single strong promoter (inducible promoter P_{BAD}) in *C. necator* H16, operons that could encode for a modified pathway committed to pimelate production.

Chapter 4

Directed evolution: Development of a directed evolution via random mutagenesis tool for *Cupriavidus necator* H16 to understand its biotin biosynthesis pathway

- To optimise a random mutagenesis method for *C. necator* H16 using EMS as a chemical mutagen.
- To explore the applicability of evolutionary engineering in *C. necator* H16 by using a directed evolution strategy *via* random chemical mutagenesis to generate a library of mutants which could help to understand the biotin pathway in *C. necator* H16.

Chapter 5

Rational design and Evolutionary engineering: Engineering *Cupriavidus necator* H16 to convert waste stream into useful chemicals by enhancing glycerol assimilation

- To apply rational design and evolutionary engineering tools explored and optimised in this PhD work to engineer *C. necator* H16 to enhance its

glycerol assimilation for chemical production using waste stream (crude glycerol) as feedstock.

1.8 Organization of this study

The different strategies used in this study to engineer *C. necator* H16 are described in **Figure 1.14**, which also shows how these may be applied synergistically, and how this PhD thesis is organised.

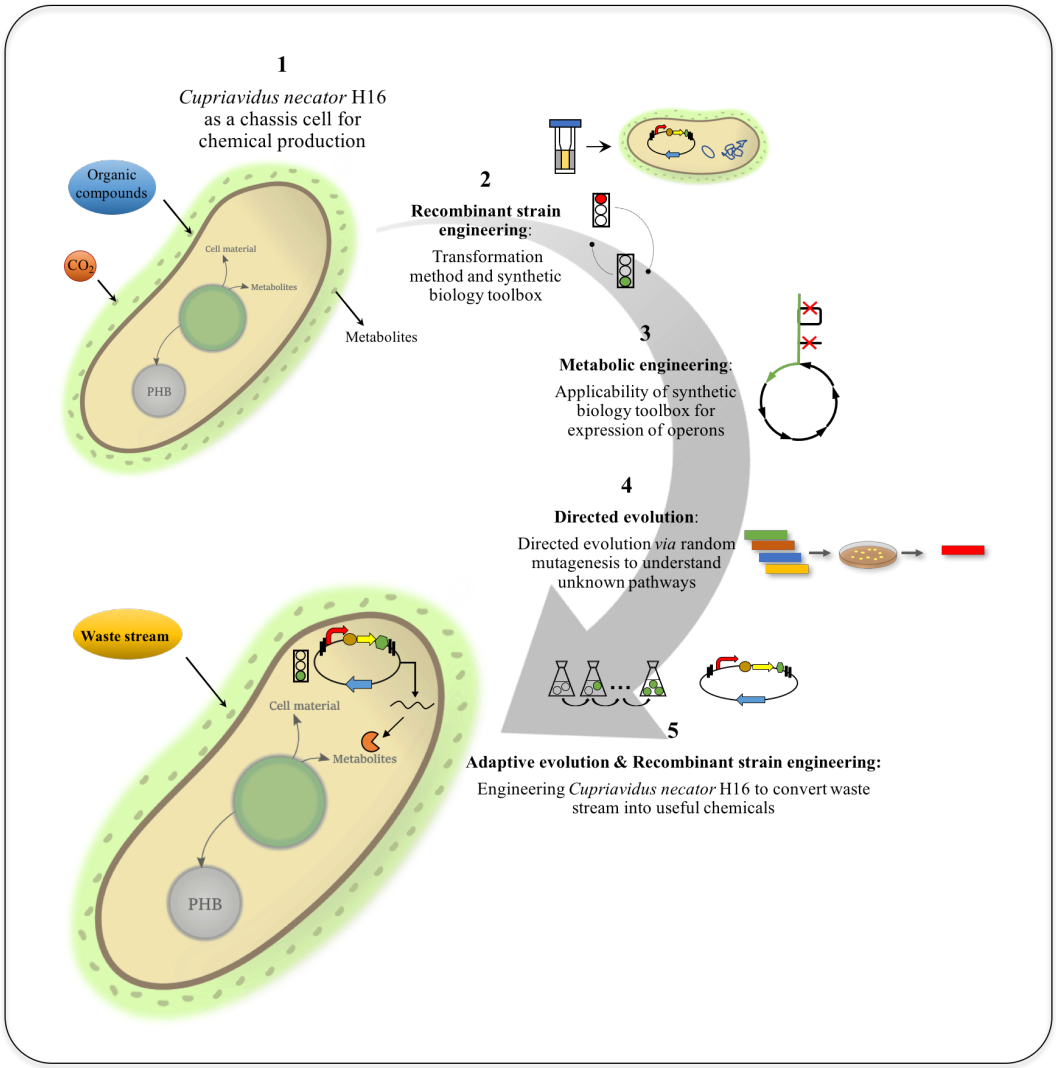


*Figure 1.14 Expanding engineering tools for Cupriavidus necator H16 to convert waste stream into useful chemicals via rational design and evolutionary engineering approaches. In Chapter 1 (**Scheme 1**), literature review about C. necator H16 is compiled to understand its natural metabolism and to analyse the different biomanufacturing potentials of the strain as well as some engineering strategies used for chemical production, afterwards, in Chapter 2 (**Scheme 2**) an optimised transformation method and a synthetic biology toolbox for C. necator H16 were developed in order to use the strain as a cell factory. Once the synthetic biology toolbox was tailored for C. necator H16, in Chapter 3 (**Scheme 3**) the synthetic biology toolbox was used for the expression of clusters of genes as well as possibly obtain bioproducts from the expression of these. Then, in Chapter 4 (**Scheme 4**), a different strategy to engineer C. necator H16 was studied: directed evolution via random mutagenesis, where an optimised protocol for random mutagenesis with a chemical mutagen was developed for C. necator H16 to understand its biotin biosynthesis pathway which could furtherly be used for chemical production. In Chapter 5 (**Scheme 5**), the knowledge and strategies studied in previous chapters and proven to work in C. necator H16, supported the approaches selected to construct recombinant and evolved strains from C. necator H16 for chemical production (bioplastics) using waste stream (crude glycerol) as a feedstock. Chapter 6 (not shown in the illustration) analyses the concluding remarks of the PhD study.*

1.9 Associated publications

- Paper (submitted, 2018)

Journal: Biological engineering
Title: "Unlocking glycerol utilization in *Cupriavidus necator* H16 for bioplastic synthesis"
Author(s): Gonzalez-Villanueva, Miriam; Tee, Kang Lan; Staniland, Paul; Staniland, Jessica; Savill, Ian; Wong, Tuck Seng

- Paper (published)

Johnson AO, Gonzalez-Villanueva M, Tee KL, and Wong TS. (2018) An engineered constitutive promoter set with broad activity range for *Cupriavidus necator* H16. *ACS Synth*. Biol. DOI: 10.1021/acssynbio.8b00136.

- Paper (published)

Tee KL, Grinham J, Othusitse AM, Gonzalez-Villanueva M, Johnson AO, and Wong TS. (2017) An Efficient Transformation Method for the Bioplastic-Producing "Knallgas" Bacterium *Ralstonia eutropha* H16. *Biotechnol J*, 12(11), DOI: 10.1002/biot.201700081.

- Book Chapter (published)

Johnson AO, Gonzalez Villanueva M, Wong TS. (April, 2015) Engineering *Ralstonia eutropha* for chemical production. In: Industrial Processes and Nanotechnology. 10. Studium Press LLC, USA. p. 255-292.

CHAPTER 2

Recombinant strain engineering:
Development of a transformation method and a synthetic biology toolbox for metabolic engineering of *Cupriavidus necator* H16

Abstract

The genome sequence of the Gram-negative lithoautotrophic β-proteobacterium *Cupriavidus necator* H16 is already available, as well as suitable plasmid systems that can be used to engineer the strain. In order to expand the biotechnological potential of *C. necator* H16, two different tools of rational design engineering were studied in this chapter. The first engineering tool consisted on developing a simpler and more robust method of transformation for *C. necator* H16, since transformation methods for *C. necator* H16 to date are based on bacterial conjugation, which requires 4 to 5 days to be accomplished. To overcome this challenge, in this **Chapter 2**, the effects of different parameters were studied to design a transformation method by electroporation optimised for *C. necator* H16. The final optimised method of transformation by electroporation consisted of a chemical treatment with 50 mM $CaCl_2$, the use of 0.2 M Sucrose, and an intermediate electric field strength (11.5 kV/cm), where the transformation efficiency was higher than 1 x 10^5 transformants/μg of DNA. After the transformation method was fully optimised, the second tool consisted on the construction of a synthetic biology toolbox tailored for *C. necator* H16, which included different biological parts such as different promoters (P_{BAD} for inducible expression of genes, and P_{j5} for constitutive expression of genes); all the plasmids constructed were based on a *C. necator* H16-compatible broad host range plasmid (pBBR1MCS-1). The transformation method and the construction of the synthetic biology toolbox are valuable tools for *C. necator* H16 that will further enable the strain to be used as a cell factory to produce value-added chemicals.

2 Chapter 2. Recombinant strain engineering: Development of a transformation method and a synthetic biology toolbox for metabolic engineering in *Cupriavidus necator* H16

2.1 Introduction

2.1.1 Transformation method by electroporation for *C. necator* H16

One of the overall aims of this PhD study is to expand engineering tools for *C. necator* H16 to explore the biotechnological potentials of the strain for chemical production, where recombinant strains of *C. necator* H16 play an important role. Low-time consuming, high efficient, and simple transformation methods to introduce foreign DNA into the strain are required for this purpose. To date, the transformation method by electroporation remained a challenge for *C. necator* H16, as the most common method for introducing plasmids to *C. necator* H16 was *via* bacterial conjugation.

Although bacterial conjugation is a well-studied and established transformation method for *C. necator H16*, it is time-consuming due to it requires two cultivation times and two transformations, as it first needs a transformation by employing $CaCl_2$ method with a donor strain (usually *E. coli* S17-1) with the plasmid of interest, before the plasmid is transferred to *C. necator* H16 (the recipient strain) by conjugation (Friedrich *et al*., 1981); this process makes the engineering of the strain slow if compared to the one-step method of transformation by electroporation in *C. necator* H16 developed in our laboratory by Tee *et al*., 2017.

Some publications have reported using both bacterial conjugation and electroporation to introduce plasmids in *C. necator* H16, and some others have transformed only by bacterial conjugation as plasmids had not been successfully transformed by electroporation into *C. necator* H16 (Park *et al*., 1995; Solaiman *et al*., 2010; Sato *et al*., 2013; Bi *et al*., 2013). Due to these reasons, the first step on this PhD investigation was to study different parameters to evaluate the effects

on transformation by electroporation of *C. necator* H16 in order to optimise the method and to achieve a low-time consuming method of transformation by electroporation for *C. necator* H16.

2.1.2 Synthetic biology toolbox for *C. necator* H16

Once having an optimised method for transformation in *C. necator* H16, it would be easier to engineer the strain, where synthetic toolboxes are required to control the expression of native or heterologous genes or pathways. In general, protein production in fermentation processes relies on a variety of factors such as gene dosage and stability of expression construct (Grabherr *et al*., 2002), promoter strength (Wilms *et al*., 2001) and induction times when inducible systems are used (Wu *et al*., 2001). In prokaryotes, the high-level protein expression is achieved by the modulation of gene dosage with the use of medium or high copy number plasmids, where usually an increase in plasmid copy number translates to an increase in the metabolic burden of the cell (Srinivassan *et al*., 2013); therefore, this must be considered when choosing a plasmid for protein expression. In *C. necator* H16, most of the heterologous expression and metabolic engineering studies conducted to date do not show a variety of suitable expression systems and neither promoter repertoires, only a few studies have standardised expression systems specifically for *C. necator* H16.

Gruber *et al*. (2014) designed stable vectors for efficient gene expression in *C. necator* H16 using promoters derived from bacteriophage T5; Bi *et al*. (2013) on the other hand, developed a toolbox with broad-host-range plasmids with a variety of origins of replication (including pBBR1), promoters, 5' mRNA stem-loop structures and ribosomal binding sites (an *E. coli* consensus RBS, a *C. necator* H16 native RBS, and a computationally automated designed RBS); while recently, in our laboratory group, Johnson *et al*. (2018) studied and constructed different promoters including the native P_{phaC1} promoter, a semi-synthetic P_{rrsC} promoter, and two coliphage T5 promoters P_{j5} and P_{g25} which were characterised with variable strength for *C. necator* H16. The promoters were genetically modified by either point mutations, length alteration, configuration alteration, or

by incorporation of regulatory genetic elements and promoter hybridisation, where a library of 42 constitutive promoters was created harbouring promoters that are stronger than the P_{j5} promoter with a ratio of promoter activities of 137 between the strongest promoter (derived from P_{j5}) and the weakest promoter (derived from P_{phaC1}).

The most commonly used plasmids for transformation in *C. necator* H16 are the medium-copy-number mobilisable plasmids pBBR1MCS and its derivatives (Kovach *et al*., 1994). In order to develop a highly stable plasmid expression system in *C. necator* H16, new plasmids derived from either pBBR1, RSF1010, RP4 and pSa have to be constructed, where pBBR1 plasmids have exhibited the less plasmid loss (Srinivasan *et al*., 2003; Voss *et al*., 2006). For this reason, pBBR1MCS-1 was used as plasmid backbone for the construction of the synthetic biology toolbox, as it is important to maintain a high rate of plasmid stability, as the decrease in productivity is often related to plasmid instability due to the lack of recombinant high-level protein expression (Voss *et al.*, 2006).

The versatility of the promoter systems must be considered as well, as this will expand the range of regulated expression; promoters derived from bacteriophage T5, such as P_{j5}, have been characterised as one of the strongest constitutive promoters in *C. necator* H16 (Gruber *et al*., 2014), other native constitutive promoters derived from PHB biosynthesis (P_{phaC}), pyruvate metabolism (P_{pdhE}), and acetoin metabolism (P_{acoD}, P_{acoX}, P_{acoE}) have been used for expression studies in *C. necator* H16, but these promoters showed comparatively weak activity (Delamarre *et al*., 2006). For inducible systems, Bi *et al*. (2013) demonstrated that the P_{BAD} and $P_{xyls/PM}$ inducible promoters provided the highest red fluorescence protein (*rfp*) expression upon induction in *C. necator* H16, hence, the promoters chosen for the synthetic biology toolbox constructed in this chapter were the P_{j5} and P_{BAD} promoters for constitutive and inducible protein expression respectively.

On the other hand, the inclusion of other biological parts to improve further protein expression have been reported and considered to be included in our synthetic toolbox; Bi *et al*. (2013) demonstrated that the insertion of a T7 stem-loop structure upstream the RBS and downstream the promoter, resulted in an

increase in RFP expression by approximately 2-fold compared to an identical plasmid without the T7 stem-loop biological part.

Regarding the RBS studied to date for *C. necator* H16, the results obtained by Bi *et al*. (2013) from three different RBS sites tested were that the *E. coli* consensus RBS sequence was the one that showed the highest RFP expression, followed by the computationally automated RBS; the RBS that showed the weakest RFP expression was a native RBS sequence from *C. necator* H16 (*nrdD* RBS).

The systems created by Bi *et al*. (2013) were used to produce hydrocarbon in *C. necator* H16, where results showed that the promoter with the highest hydrocarbon titres was the P_{BAD} promoter. On the other hand, the same plasmid with the P_{BAD} promoter was constructed with and without the T7 5' mRNA stem-loop structure, and both plasmids showed similar production titres; the *E. coli* consensus RBS achieved the highest hydrocarbon titres, as it had been observed in RFP expression.

Based on this studies performed previously in *C. necator* H16, the construction of a synthetic biology toolbox was the second step in the investigation of this PhD, since this would serve to expand the metabolic capabilities of *C. necator* H16. The optimised transformation method by electroporation developed in the laboratory was used to transfer the plasmids to *C. necator* H16.

2.2 Aim

The aims of **Chapter 2** are:

- To optimise a transformation method by electroporation for *C. necator* H16.
- To construct a synthetic biology toolbox with inducible and constitutive promoters and other synthetic biology parts for metabolic engineering applications.

2.3 Materials and methods

2.3.1 Materials

List of detailed information on the materials used (reagents, kits, equipment, software, media preparation and miscellaneous) can be found in **Appendix 1 - 6**.

2.3.2 Strains, plasmids and primers

All strains, plasmids and primers that were used in this **Chapter 2** are shown in **Table 2.1**.

Table 2.1 Strains, plasmids and primers used in this study.

Strains, plasmids or primers	Description	References or source
Strains		
E. coli DH5α	Standard cloning strain	Lab collection
C. necator H16	Wild-type Gen^r	DSM 428
Plasmid backbones		
pBHR1	Broad-host-range plasmid, Cam^r, Kan^r, pBBR1 Rep, pBBR1oriV, *mob*	MoBiTec GmbH, Germany
pBBR1MCS-1	Broad-host-range plasmid, Cam^r, pBBR1 Rep, pBBR1oriV, *mob*	Kovach *et al*., 1994
pBbA8k-RFP	Kan^r, *araC*, P_{BAD}, RBS, *rfp*	Lee *et al*., 2011
pEGFP	Amp^r, P_{lac}, *egfp*	Clontech
Plasmids constructed		
pBADk.rbs-RFP	Kan^r, pBBR1 Rep, pBBR1 oriV, *mob*, *araC*, P_{BAD}, RBS, *rfp*	This work
pBADc.rbs-RFP	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, *araC*, P_{BAD}, RBS, *rfp*	This work
pPj5c.T7rbs-RFP	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, P_{j5}, T7 5' mRNA stem-loop, RBS, *rfp*	This work
pPj5c.T7rbs-eGFP	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, P_{j5}, T7 5' mRNA stem-loop, RBS, *egfp*	This work
Primers (5' → 3')		
AvrII-pBBR1MCS1	GATCCCTAGGGAAGACGAAAGGGCCTCG TGATACG	This work
pBBR1MCS1-SacI	GATCGAGCTCAAATTGTAAGCGTTAATAT TTTGTTAAAATTCGCGTTAAATTTTTG	This work
AvrII-Cam-pBBR1	GATCCCTAGGATTGTTATCCGCTCACAAT TCCACACAACATAC	This work
pBBR1MCS1-PstI	GATCCTGCAGAAATTGTAAGCGTTAATAT TTTGTTAAAATTCGCGTTAAATTTTTG	This work
T7-RBS-RFP-fwd	GAATTCAAAAGATCT**GGGAGACCACAA**	This work

	CGGTTTCCCTCTAGAAATAATTTTGGA ATTCAAAAGATCT**TTTAAGAAGGAGATA TACAT**ATGGCGAGTAGCGAAGACGT	
Pj5-T7-fwd	GATCCTGCAG**AGCGGATATAAAAACCG TTATTGACACAGGTGGAAATTTAGAAT ATACTGTTAGTAAACCTAATGGATCGA CCTT**GAATTCAAAAGATCTGGGAGACC	This work
RFP-rev	GATCCTCGAGTTAAGCACCGGTGGAGTG ACGACCT	This work
NdeI-eGFP-fwd	GATCCATATGGTGAGCAAGGGCGAGGAG	This work
eGFP-XhoI-rev	GATCCTCGAGTTACTTGTACAGCTCGTCC ATGCCG	This work

All the RBS noted in all the plasmids in the list correspond to the sequence encoding for the *E. coli* consensus RBS. Nucleotide sequences highlighted in purple encode for T7 5' mRNA stem-loop, nucleotide sequences in blue encode for *E. coli* consensus RBS, and the nucleotide sequence highlighted in orange encodes for the P_{j5} promoter. Restriction enzyme sequences of primers are underlined. Abbreviations: Gen^r, gentamicin resistance; Cam^r, chloramphenicol resistance; Kan^r, kanamycin resistance; Amp^r, ampicillin resistance; RBS, ribosome binding site; *rfp*; red fluorescence protein gene; *egfp*, enhanced green fluorescence protein.

2.3.3 Cultivation of *E. coli* DH5α and *C. necator* H16

E. coli DH5α cells were cultivated at 37 ºC on 2x YT medium and 250 rpm, and when required with chloramphenicol [25 μg/mL] or kanamycin [50 μg/mL] according to the application, the strain was used for all molecular cloning, plasmid propagation, and maintenance. *C. necator* H16 cells were cultivated at 30 ºC using nutrient broth (NB) and 250 rpm always supplemented with gentamicin [10 μg/mL] and when needed with chloramphenicol [25 μg/mL] or kanamycin [250 μg/mL]. The optical density was measured at 600 nm with a BioPhotometer.

2.3.4 Bacterial transformation of *E. coli* DH5α and *C. necator* H16

*E. coli DH5*α cells were transformed using a standard chemical transformation method. *E. coli* DH5α was used for circularization and propagation of ligation mixtures of all plasmids constructed. *C. necator* H16 cells were transformed by the electroporation method optimised in our laboratory group.

2.3.4.1 Transformation using $CaCl_2$ heat-shock method for *E. coli* DH5α

A pre-culture of *E. coli* DH5α was cultivated in 5 mL of 2x YT for 16 h at 37 ºC; for competent cell preparation, a fresh 2x YT falcon tube was inoculated with the pre-culture at a 1:100 dilution and cultivated at 37 ºC, when cells reached an optical density at 600 nm (OD_{600}) of 0.5, 1 mL of cells were transferred to a sterile 1.5 mL microcentrifuge tube. Cells were centrifuged at maximum speed (17,000 x g) for 30 s, and the supernatant was removed by pipetting. The pellet was washed once with 500 µL of pre-chilled 50 mM $CaCl_2$. Then, the cell pellet was resuspended carefully in 500 µL of pre-chilled 50 mM $CaCl_2$ and incubated in ice for 30 min (for transforming intact plasmid, a 10 min incubation was sufficient). Plasmid DNA was added to a concentration of 1 µg or 5 µL of ligation mixture and mixed gently. A second 30 min incubation step was performed after the addition of the plasmid (for transforming intact plasmid, a 10 min incubation was sufficient). After the second incubation, the cells were heat-shocked at 42 ºC for 1 min and further incubated in ice for 2 min. After heat shock and ice incubation time, 800 µL of 2x YT was added, and cells were left to grow for a 1 h outgrowth at 37 ºC. After the outgrowth, cells were centrifuged at maximum speed for 30 s and most of the media removed, the remaining 200 – 300 µL media was used to resuspend the cells gently before plating them on TYE agar plates with the required antibiotic, and incubated overnight at 37 ºC.

2.3.4.2 Transformation by electroporation method for *C. necator* H16

Unless specified otherwise, the following transformation by electroporation method was used. A pre-culture of *C. necator* H16 was cultivated in 5 mL of NB with gentamicin for 40 – 44 h at 30 ºC; for electrocompetent cell preparation, a fresh NB supplemented with gentamicin falcon tube was inoculated with the pre-culture at a 1:50 dilution and cultivated at 30 ºC, when cells reached an OD_{600} of 0.5 – 0.7, cells were transferred to ice and chilled for 5 min. Two millilitres of the cells were then transferred to a sterile 2 mL microcentrifuge tube. Cells were

centrifuged at maximum speed (17,000 x g) for 30 s, and the supernatant was removed by pipetting followed by three cell washes. For experiments with chemical treatment (100 mM EDTA or 20 mM $MgCl_2$) the pellet was first washed once by resuspending cells in 1 mL of the chemical and incubated for 30 min in ice. Then, the cell pellet was washed twice by resuspension in 1 mL of pre-chilled transformation buffer (10 % (v/v) Glycerol or 0.3 M Sucrose), after the last washing step, the supernatant was removed by pipetting. When chemical treatments were not used, the first cell wash was replaced by resuspending the cell pellet with the transformation buffer instead of the chemical agent. After the final wash, the cell pellet was resuspended in 100 μL of pre-chilled transformation buffer. Plasmid DNA was added to a concentration of 0.5 μg to the resuspended cells and mixed gently. The resuspension was then transferred into a pre-chilled 2 mm electroporation cuvette and electroporated at 2.5 kV. After electroporation, 1 mL of NB was added immediately directly to the electroporation cuvette, then, cells were transferred to a new 2 mL microcentrifuge tube for a 2 h outgrowth at 30 ºC. After the outgrowth, cells were centrifuged at maximum speed for 30 s, and 950 μL of the supernatant was removed to resuspend all cells in the remaining 150 μL supernatant. Then, cells were plated in NB with gentamicin and chloramphenicol agar plates and incubated at 30 ºC for 40 – 48 h (Adapted from Tee *et al*., 2017).

2.3.5 DNA preparation

Standard procedures were used for isolation of plasmids, restriction enzyme digestions, polymerase chain reaction (PCR) with Pfu Turbo and Ultra DNA polymerase, *DpnI* digestions, DNA gel extraction, PCR purification, and T4 DNA ligation (Sambrook *et al*., 2011) and recommendations by the manufacturers. All primers were synthesised by Eurofins Genomics.

2.3.6 DNA Gel electrophoresis

An agarose gel electrophoresis was prepared to analyse DNA; 0.7 % or 1.0 % (w/v) agarose gels were prepared in 1x TBE buffer (prepared from a 5x TBE buffer composed of 54.0 g of Tris base, 27.5 g of Boric acid, and 20 mL of 0.5 M EDTA per 1 L, pH 8.3) by dissolving 0.35 g or 0.50 g of agarose respectively in 50 mL buffer. The percentage of gel used was dependent on the purpose of the gel, if only an analysis of the DNA was required, then 1.0 % (w/v) gels were used, but if gel extraction was required, then 0.7 % (w/v) agarose gels were used to excise the DNA band from the gel. To ensure that the agarose is fully dissolved in the buffer, the solution was microwaved until all the agarose was dissolved, then, when the gel had cooled down, 2 μL of ethidium bromide was added and the gel was cast using a gel caster, then the comb was inserted and the gel was left to room temperature until it solidified. 6 μL of DNA Ladder 1 kb and appropriate volume of DNA samples were loaded into the gel. The electrophoresis was run at a constant voltage of 100 V for 60 min. The gel image was captured with a gel documentation system.

2.3.7 Plasmid construction

All plasmid constructed in this **Chapter 2** were based on the plasmid backbones pBBR1MCS-1 and pBbA8k-RFP plasmids for the construction of the synthetic biology toolbox. The plasmid backbones are illustrated in **Figure 2.4** and **Figure 2.5** respectively. The chloramphenicol resistance (Cam^r) cassette (containing P_{cat} and Cam^r), pBBR1 Rep, pBBR1 oriV origin of replication, and *mob* gene were amplified from pBBR1MCS-1; the *rfp* gene, *E. coli* consensus RBS, L-arabinose-inducible system (containing *araC* and P_{BAD}), and the kanamycin resistance (Kan^r) cassette (containing P_{NEOKAN} and Kan^r) were amplified from pBbA8k-RFP; and the *egfp* gene was amplified from pEGFP. The primers and *in silico* analyses of plasmid construction were generated using the SnapGene software tool. Correct construction of all plasmids was confirmed by restriction enzyme analysis (ReA).

2.3.7.1 Construction of plasmid pBADk.rbs-RFP

pBADk.rbs-RFP was constructed by amplifying the *mob* gene, pBBR1 oriV, and pBBR1 Rep from pBBR1MCS-1 plasmid with AvrII-pBBR1MCS1 (forward primer) and pBBR1MCS1-SacI (reverse primer) primers. Subsequently, the amplified fragment was subjected to *DpnI* digestion, PCR purification and restriction enzyme digestion with *AvrII* and *SacI*, followed by a further PCR purification for ligation with T4 DNA ligase with the *AvrII* and *SacI* digested and gel extracted fragment containing Kan^r cassette, the L-arabinose-inducible system (*araC* and P_{BAD}), *E. coli* consensus RBS, and *rfp* gene from pBbA8k-RFP; resulting in the 6.3 kb plasmid pBADk.rbs-RFP. After transformation of *E. coli* DH5α, a single colony of the resulting plasmid pBADk.rbs-RFP was isolated and analysed by restriction enzyme analysis to confirm the right inserts before it was transformed into *C. necator* H16 by electroporation.

2.3.7.2 Construction of plasmid pBADc.rbs-RFP

pBADc.rbs-RFP was constructed by amplifying the Cam^r cassette, *mob* gene, pBBR1 oriV, and pBBR1 Rep from pBBR1MCS-1 plasmid with AvrII-Cam-pBBR1 (forward primer) and pBBR1MCS1-PstI (reverse primer) primers. Subsequently, the amplified fragment was subjected to *DpnI* digestion, PCR purification and restriction enzyme digestion with *AvrII* and *PstI*, followed by a further PCR purification for ligation with T4 DNA ligase with the *AvrII* and *PstI* digested and gel extracted fragment containing the L-arabinose-inducible system (*araC* and P_{BAD}), *E. coli* consensus RBS, and *rfp* gene from pBbA8k-RFP; resulting in the 6.7 kb plasmid pBADc.rbs-RFP. After transformation of *E. coli* DH5α, a single colony of the resulting plasmid pBADc.rbs-RFP was isolated and analysed by restriction enzyme analysis to confirm the right inserts before it was transformed into *C. necator* H16 by electroporation.

2.3.7.3 Construction of plasmid pPj5c.T7rbs-RFP

pPj5c.T7rbs-RFP was constructed by amplifying the *rfp* gene from pBbA8k-RFP plasmid with T7-RBS-RFP-fwd (forward primer) and RFP-rev (reverse primer) primers; the T7 stem-loop and *E. coli* consensus RBS sequences were contained within the T7-RBS-RFP-fwd (forward primer), resulting in an amplified fragment with a T7 stem-loop and *E. coli* consensus RBS upstream the *rfp* gene amplified from pBbA8k-RFP. Subsequently, the amplified fragment was subjected to *DpnI* digestion and PCR purification before a second PCR reaction was carried out with Pj5-T7-fwd (forward primer) and RFP-rev (reverse primer) primers to insert the P_{j5} promoter sequence upstream the T7 stem-loop. This fragment was then subjected to restriction enzyme digestion with *PstI* and *XhoI*, followed by a further PCR purification for ligation with T4 DNA ligase with *PstI* and *XhoI* digested and gel extracted fragment containing the Cam^r cassette, *mob* gene, pBBR1 oriV origin of replication and pBBR1 Rep from pBADc.rbs-RFP; resulting in the 5.3 kb plasmid pPj5c.T7rbs-RFP plasmid. After transformation of *E. coli* DH5α, a single colony of the resulting plasmid pPj5c.T7rbs-RFP was isolated and analysed by restriction enzyme analysis to confirm the right inserts before it was transformed into *C. necator* H16 by electroporation.

2.3.7.4 Construction of plasmid pPj5c.T7rbs-eGFP

pPj5c.T7rbs-eGFP was constructed by amplifying the *egfp* gene from pEGFP plasmid with NdeI-eGFP-fwd (forward primer) and eGFP-XhoI-rev (reverse primer) primers. Subsequently, the amplified fragment was subjected to *DpnI* digestion, PCR purification and restriction enzyme digestion with *NdeI* and *XhoI*, followed by a further PCR purification for ligation with T4 DNA ligase with the *NdeI* and *XhoI* digested and gel extracted fragment containing the Cam^r cassette, *mob* gene, pBBR1 oriV origin of replication, pBBR1 Rep, P_{j5} promoter, T7 stem-loop, and *E. coli* consensus RBS from pPj5c.T7rbs-RFP; resulting in the 5.3 kb plasmid pPj5c.T7rbs-eGFP. After transformation of *E. coli* DH5α, a single colony of the resulting plasmid pPj5c.T7rbs-eGFP was isolated and analysed by

restriction enzyme analysis to confirm the right inserts before it was transformed into *C. necator* H16 by electroporation.

2.4 Results and discussion

2.4.1 Optimisation of *C. necator H16* transformation by electroporation

One of the first experiments performed during the PhD was a collaboration with the laboratory group for the optimisation of the transformation by electroporation method for *C. necator* H16. Electroporation is a physical method of cell transformation that applies an electric field to the cell membrane of cells, which aids to open pores in the cell wall that permits the entry of external DNA or other molecules (RNA, proteins, etc.), where the electroporator generates an electromagnetic field in the cell solution. The advantages of electroporation are that it is not limited by plasmid size and that the DNA uptake is immediate as the electroporation does not require incubation.

In order to optimise this method, common factors responsible for the transformation efficiency by electroporation were studied, such as the choice of chemical agents and electroporation buffers, which are responsible for the increase in cell wall permeability. Other factors considered as well were heat-shock treatment and different gap sizes of electroporation cuvettes.

In this part of the study, *C. necator* H16 cells were cultivated in NB and transformants were selected on NB agar plates with gentamicin, and when required with chloramphenicol at 30 ºC. Plasmids used in this work were the broad-host-range mobilisable plasmids pBBR1MCS-1 (4.7 kb), pBHR1 (5.3 kb) and pBADc.rbs-RFP (6.7 kb). A negative control sample ((-)ve) using cells of *C. necator* H16 washed with 10 % (v/v) glycerol transformation buffer was always included in all parameters studied to check for any contamination that could occur during the process and interfere with the analysis of data. The transformation method followed for this (-)ve control sample was the same as described in

Section **2.3.4.2**, with the only difference in that no plasmid was added for these samples.

2.4.1.1 Electroporation cuvettes: 1 mm and 2 mm

In the early steps of the optimisation for transforming *C. necator* H16 by electroporation, 10 % (v/v) glycerol was used as standard transformation buffer to make *C. necator* H16 cells competent. This buffer was initially used as a constant in the optimisation process. Then the first parameter analysed was the gap size of the electroporation cuvettes.

Two different gap size cuvettes can be used for the electroporation step in transformation by electroporation for bacteria, the 1 mm cuvette and the 2 mm cuvette. Electroporation cuvettes (**Figure 2.1**) have aluminium electrodes and are available in three different sizes of gap widths of 1 mm (for up to 100 μL volume), 2 mm (for up to 400 μL volume), and 4 mm (for up to 800 μL volume). Usually, 1 mm cuvettes are used for electroporation of bacteria, 2 mm cuvettes are used for electroporation of yeast, and 4 mm cuvettes for electroporation of mammalian cells, although 2 mm cuvettes have been used for bacteria electroporation as well. The gap size is the distance between the electrodes, and it is one of the first optimisations to be considered for electroporation experiments, as the gap size is used for the determination of the field strength (kV) by the formula of voltage divided by gap size (mm) (Dower *et al*., 1988). Hence, the 1 mm cuvette is used for the highest field strength, the 2 mm cuvette is for intermediate requirements, and the 4 mm cuvette has the lowest field strength. Generally, the voltage used for 1 mm cuvettes is around 1.7 kV, while the voltage used for 2 mm cuvettes is around 2.5 kV.

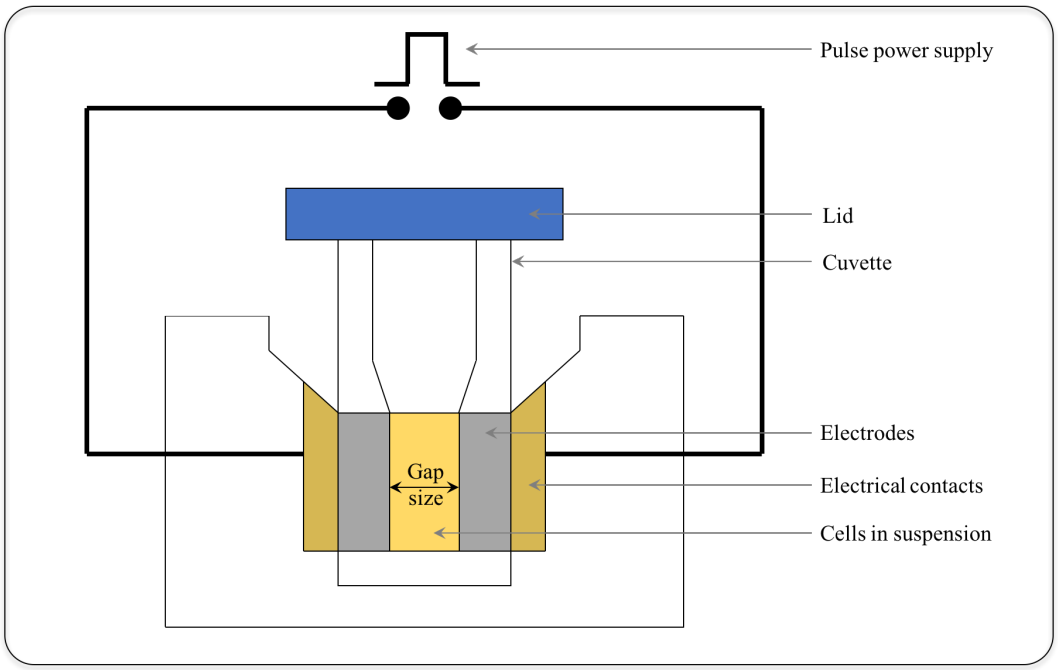


Figure 2.1 Electroporator cuvette components. Gap sizes can be 1 mm, 2 mm or 4 mm.

The method followed for transformation by electroporation for *C. necator* H16 to study the effects of using different gap sizes is described in **Section 2.3.4.2**, whereby 10 % (v/v) glycerol was used as transformation buffer for the three cells washes as no chemical treatment was performed. Plasmid DNA (pBBR1MCS-1) was added to a concentration of 0.5 µg and cells were electroporated at 1.7 kV when 1 mm cuvettes were used and at 2.5 kV when 2 mm cuvettes were used. After the 2 h outgrowth, all cells were plated in NB agar plates supplemented with gentamicin and chloramphenicol and incubated at 30 ºC for 40 – 48 h. The results of these experiments are shown in **Table 2.2** and **Figure 2.2**.

Table 2.2 Parameter studied for transformation by electroporation: Gap size of electroporation cuvettes.

		Transformation efficiency
Cell density at start of transformation		
OD_{600}	0.676	
DNA used		
pBBR1MCS-1	0.5 µg	
Other information		
Electroporation cuvette	1 mm, 2 mm	
Colony count	**10 % (v/v) Gly**	
	1 mm: 78 colonies	1.56×10^2 transformants/µg DNA
	2 mm: 697 colonies	1.39×10^3 transformants/µg DNA

Abbreviations: Gly, glycerol.

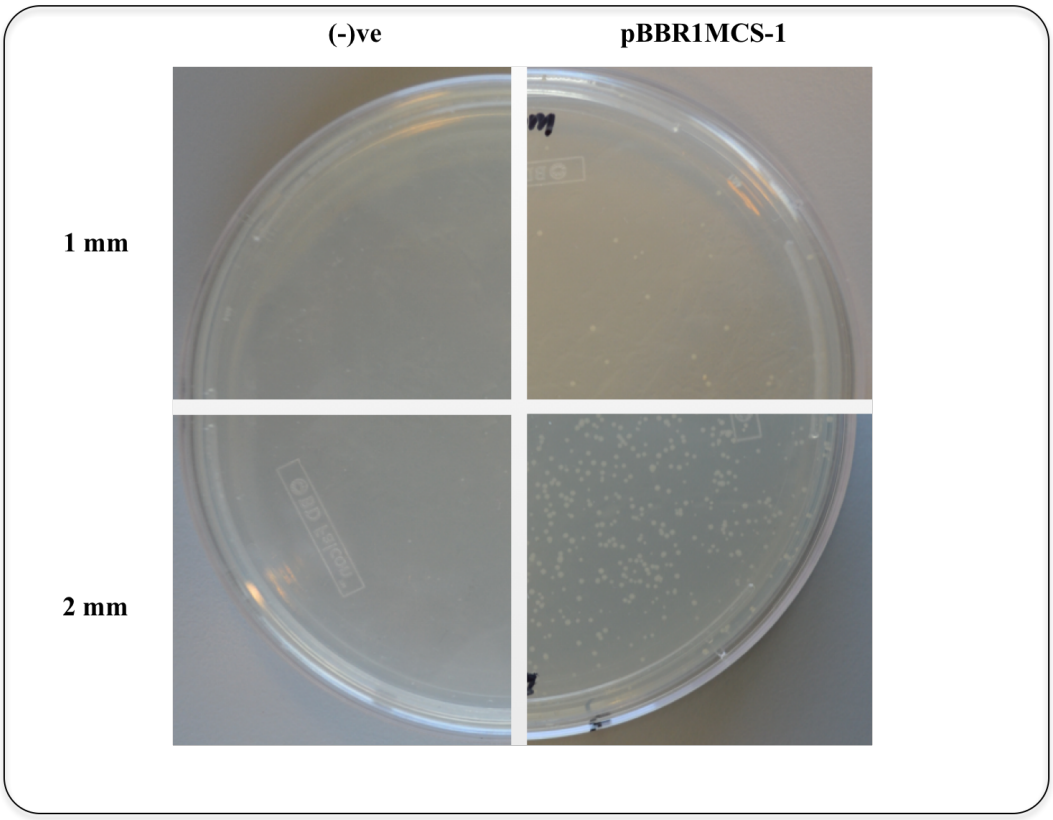


Figure 2.2 Transformation by electroporation in C. necator H16 with pBBR1MCS-1 using 1 mm and 2 mm electroporation cuvettes.

From this first parameter studied, it was clear that the 2 mm gap electroporation cuvette provided the best transformation efficiency, this might be due to the 1 mm gap electroporator cuvette that provided a higher field strength, killed a higher population of cells during electroporation, while the 2 mm gap electroporation cuvette provided and intermediate field strength allowing a higher number of cells to survive after electroporation. From this point onwards, only 2 mm electroporation cuvettes were used for the analysis of the rest of the experiments as they provided a higher transformation efficiency.

2.4.1.2 Physical treatment: Electroporation and heat shock

Another experiment was designed in order to improve the transformation efficiency based on the principle used for bacterial transformation of the $CaCl_2$ heat shock method. The principle behind the transformation by $CaCl_2$ heat shock method is that the cells are manipulated with positively charged calcium ions (a divalent cation, Ca^{2+}) followed by a temperature imbalance that helps to uptake the plasmid DNA. It has been reported that the naked plasmid DNA is bound to the lipopolysaccharide (LPS) receptor molecules of the bacterial cell wall, then the divalent cations of calcium chloride bind both to the negatively charged LPS in the bacterial cell wall and the negatively charged plasmid DNA; after the $CaCl_2$ treatment, is the heat-shock step that ferries the plasmid DNA into the bacterial cell due to pores are formed in the cell membrane once cells are chilled in ice after being heated at 42 ºC for a short time. The heat shock at 42 ºC depolarises the cell wall of the already chemically treated competent cells lowering its negativity potential, and the subsequent cold shock makes the cell membrane to raise its membrane potential and recover its original value. Hence, this experiment was designed to observe the effect of the heat shock treatment and observe if this could improve further the cell wall permeability and therefore increase the transformation efficiency.

The transformation by electroporation method followed for this part of the study is described in **Section 2.3.4.2**, whereby 10 % (v/v) glycerol was used as transformation buffer for the three cell washes and as no chemical treatment was performed. Plasmid DNA (pBADc.rbs-RFP and pBHR1) was added to a concentration of 0.5 μg, after adding the plasmid DNA, cells were incubated for 10 min in ice, followed by heat shock at 42 ºC for 1 min and a further incubation in ice for 2 min, then cells were electroporated at 2.5 kV in 2 mm cuvettes. After the 2 h outgrowth, all cells were plated in NB agar plates supplemented with gentamicin and chloramphenicol and incubated at 30 ºC for 40 – 48 h. The results of these experiments are shown in **Table 2.3**.

Table 2.3 Parameter studied for transformation by electroporation: Physical treatment with heat shock.

		Transformation efficiency
Cell density at start of transformation		
OD_{600}	0.604	
DNA used		
pBADc.rbs-RFP	0.5 µg	
pBHR1	0.5 µg	
Other information		
Electroporation cuvette	2 mm	
Colony count	**10 % (v/v) Gly**	
	pBADc.rbs-RFP: 658 colonies	1.32×10^3 transformants/µg DNA
	pBHR1: 5 colonies	1.00×10^1 transformants/µg DNA
	Heat shock & 10 % (v/v) Gly	
	pBADc.rbs-RFP: 3 colonies	6.00×10^0 transformants/µg DNA
	pBHR1: 1 colony	2.00×10^0 transformants/µg DNA

Abbreviations: Gly, glycerol.

This strategy yielded very few transformants after heat shock and electroporation treatment, this might be due to bacteria are not stable when they have a large number of pores on the cell wall, and this might lead to death easily; therefore, it is possible that the application of the two physical methods to the cells surpassed the limit of destabilisation that cells can stand and that could have lead to a decrease in cell's viability.

2.4.1.3 Chemical treatment: 100 mM EDTA

It has been reported that Ethylenediaminetetraacetic acid (EDTA) chemical treatment helps to increase the permeability of the cell membrane of gram-negative bacteria, where presumably, the external cell membrane loses its barrier function in relation to DNA after the EDTA treatment; it was also reported that EDTA treatment without a subsequent electric treatment was not successful in *E. coli* strains (Cymbalyuk *et al*., 1988). Considering what has been reported in literature, the design of this experiment was performed using EDTA as a chemical treatment for cell-weakening in combination with electroporation, using 10 % (v/v) glycerol and 0.3 M sucrose as transformation buffers.

The transformation by electroporation method followed for this part of the study is described in **Section 2.3.4.2**, whereby the chemical treatment was performed with 1 mL of 100 mM EDTA and an incubation of 30 min in ice in the first cell wash. Then, the cell pellet was washed twice by resuspension in 1 mL of pre-chilled transformation buffer of 10 % (v/v) glycerol or 0.3 M sucrose. Plasmid DNA (pBADc.rbs-RFP and pBHR1) was added to a concentration of 0.5 μg, and cells were electroporated at 2.5 kV in 2 mm electroporation cuvettes. After the 2 h outgrowth, all cells were plated in NB agar plates supplemented with gentamicin and chloramphenicol and incubated at 30 ºC for 40 – 48 h. The results of the experiments are shown in **Table 2.4**.

Table 2.4 Parameter studied for transformation by electroporation: Chemical treatment with 100 mM EDTA.

		Transformation efficiency
Cell density at start of transformation		
OD_{600}		
DNA used		
pBADc.rbs-RFP	0.5 μg	
pBHR1	0.5 μg	
Other information		
Electroporation cuvette	2 mm	
Colony count	**10 % (v/v) Gly**	
	pBADc.rbs-RFP: 467 colonies	9.34×10^2 transformants/μg DNA
	pBHR1: 9 colonies	1.80×10^1 transformants/μg DNA
	100 mM EDTA & 10 % (v/v) Gly	
	pBADc.rbs-RFP: 3 colonies	6.00×10^0 transformants/μg DNA
	pBHR1: 0 colonies	0 transformants/μg DNA
	100 mM EDTA & 0.3 M Suc	
	pBADc.rbs-RFP: 1 colony	2.00×10^0 transformants/μg DNA
	pBHR1: 2 colonies	4.00×10^0 transformants/μg DNA

Abbreviations: Gly, glycerol; EDTA, ethylenediaminetetraacetic acid; Suc, sucrose.

The treatment with EDTA not only did not improve transformation efficiency but it decreased the transformation efficiency significantly compared to the one obtained when only 10 % (v/v) glycerol was used as transformation buffer for pBADc.rbs-RFP. The same low transformation efficiency values were obtained when EDTA was used in combination with 0.3 M sucrose as transformation buffer. Even though it has been reported in literature that EDTA improved transformation efficiency in *E. coli* strains, it is well known that the

transformation efficiency can be different among different microorganims, as they have different cell membrane structures.

2.4.1.4 Chemical treatment: 20 mM $MgCl_2$

Other chemicals have been reported to improve transformation efficiency, such as combinations of calcium and magnesium (Taketo, 1974: Wensink *et al*., 1974) at prolonged incubations in ice (reviewed in Asif *et al*., 2017). In general, all divalent cation such as Ca^{2+} and Mg^{2+} enhance the transformation process. It has been reported that magnesium increases the transformation efficiency in bacteria by 15 to 20-fold compared with transformations performed with no magnesium treatment (Hanahan, 1983), where incubation of 30 min in ice enhanced further bacterial transformation (reviwed in Asif *et al*., 2017).

The transformation by electroporation method followed for this part of the study is described in **Section 2.3.4.2**, whereby the chemical treatment was performed with 1 mL of 20 mM $MgCl_2$ and an incubation of 30 min in ice in the first cell wash. Then, the cell pellet was washed twice by resuspension in 1 mL of pre-chilled transformation buffer of 0.3 M sucrose. Plasmid DNA (pBADc.rbs-RFP and pBHR1) was added to a concentration of 0.5 μg, and cells were electroporated at 2.5 kV in 2 mm electroporation cuvettes. After the 2 h outgrowth, all cells were plated in NB agar plates supplemented with gentamicin and chloramphenicol and incubated at 30 ºC for 40 – 48 h. The results of these experiments are shown in **Table 2.5**.

Table 2.5 Parameter studied for transformation by electroporation: Chemical treatment with 20 mM $MgCl_2$.

		Transformation efficiency
Cell density at start of transformation		
OD_{600}		
DNA used		
pBADc.rbs-RFP	0.5 µg	
pBHR1	0.5 µg	
Other information		
Electroporation cuvette	2 mm	
Colony count	**10 % (v/v) Gly**	
	pBADc.rbs-RFP: 751 colonies	1.50×10^3 transformants/µg DNA
	pBHR1: 12 colonies	2.40×10^1 transformants/µg DNA
	20 mM $MgCl_2$ & 0.3 M Suc	
	pBADc.rbs-RFP: 683 colonies	1.37×10^3 transformants/µg DNA
	pBHR1: 380 colonies	7.60×10^2 transformants/µg DNA

Abbreviations: Gly, glycerol; $MgCl_2$, magnesium chloride; Suc, sucrose.

Although these results did not show an enhancement in transformation efficiency with the pBADc.rbs-RFP plasmid compared to only using 10 % (v/v) glycerol as transformation buffer, it did show an improvement of transformation efficiency with the pBBHR1 plasmid compared to previous results obtained with all different buffers and chemicals studied before. From this point onwards, the investigation of the transformation buffer of 0.3 M sucrose was focused on testing whether the enhancement in transformation efficiency for pBHR1 transformation was due to $MgCl_2$ chemical treatment or due to the use of 0.3 M sucrose as transformation buffer, or if it was required to combine both to achieve this transformation efficiency.

2.4.1.5 Transformation buffer: 0.3 M sucrose

In order to have high efficient electroporation, a medium with low conductivity has to be used, for this end, cell washes are essential in electroporation as they remove the ions present in the growth media (Dower *et al*., 1998). Some of the most common transformation buffers used in microorganisms are double-distilled water, 10 % (v/v) glycerol, 0.3 M sucrose, 10 % (w/v) fructose, and 0.3 M glucose. The first one was used in the first set of experiments, and the second one had been used in combination with chemical treatments before this section.

In this section, the study of combining the transformation buffer 0.3 M sucrose with 10 % (v/v) glycerol was also investigated (a single buffer solution containing both glycerol and sucrose transformation buffers was prepared). The transformation by electroporation method followed for this investigation is described in **Section 2.3.4.2**, whereby the only difference is that the buffers of 0.3 M sucrose or 10 % (v/v) glycerol and 0.3 M sucrose were used as transformation buffers for the three cell washing steps. Plasmid DNA (pBADc.rbs-RFP and pBHR1) was added to a concentration of 0.5 μg and cells were electroporated at 2.5 kV in 2 mm electroporation cuvettes. After the 2 h outgrowth, all cells were plated in NB agar plates supplemented with gentamicin and chloramphenicol and incubated at 30 ºC for 40 – 48 h. The results of these experiments are shown in **Table 2.6**.

Table 2.6 Parameter studied for transformation by electroporation: Transformation buffer 0.3 M sucrose.

		Transformation efficiency
Cell density at start of transformation		
OD_{600}	0.631	
DNA used		
pBADc.rbs-RFP	0.5 μg	
pBHR1	0.5 μg	
Other information		
Electroporation cuvette	2 mm	
Colony count	**10 % (v/v) Gly**	
	pBADc.rbs-RFP: 483 colonies	9.66×10^2 transformants/μg DNA
	pBHR1: 10 colonies	2.00×10^1 transformants/μg DNA
	0.3 M Suc	
	pBADc.rbs-RFP: 1796	3.59×10^3 transformants/μg DNA
	pBHR1: 240 colonies	4.80×10^2 transformants/μg DNA
	10 % (v/v) Gly & 0.3 M Suc	
	pBADc.rbs-RFP: 2 colonies	4.00×10^0 transformants/μg DNA
	pBHR1: 1 colony	2.00×10^0 transformants/μg DNA

Abbreviations: Gly, glycerol; Suc, sucrose.

The buffer 0.3 M sucrose showed the best transformation efficiency obtained for both plasmids, and the combination of sucrose with glycerol transformation buffers resulted in poorer transformation efficiency for both plasmids transformed, so at this point of the investigation, the 0.3 M sucrose buffer showed the highest transformation efficiency.

2.4.1.6 Other chemical and transformation buffers studied

The results of the parameters shown and discussed in the previous sections developed in the first stages of my PhD served as a basis to plan further experiments that were performed by other members of our laboratory group to fine-tune and finalise the optimisation of the transformation method by electroporation for *C. necator* H16, which can be found more detailed in Tee *et al*., 2017. These studies were performed with the pBHR1 plasmid. **Figure 2.3** shows the replicates of the control samples (*C. necator* H16 electroporation using 10 % (v/v) glycerol as transformation buffer) from all the parameters studied, which show the reproducibility of the transformation by electroporation method when using 10 % (v/v) glycerol as transformation buffer.

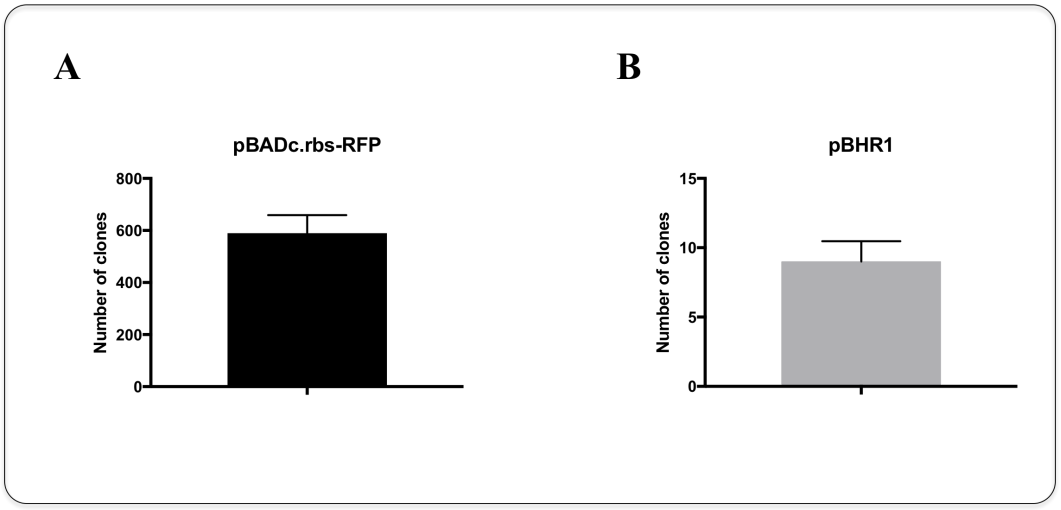


Figure 2.3 Replicates of control samples included in each study of transformation by electroporation for C. necator H16. ***A)*** *Control sample harbouring pBADc.rbs-RFP plasmid.* ***B)*** *Control sample harbouring pBHR1 plasmid. Error bars are ± SEM of four independent determinations.*

From all the parameters studied, the experiment that used 0.3 M sucrose as a transformation buffer resulted with the highest transformation efficiency, hence, further experiments were performed in the laboratory to find the optimal sucrose concentration, which was tested from 0.1 to 0.6 M sucrose, where 0.2 M sucrose resulted in the optimum concentration with the highest transformation efficiency.

Additionally, different chemical treatments were analysed, 50 mM $CaCl_2$, and TSS buffer, which showed to improve further transformation efficiency. Once the optimum transformation buffer -0.2 M sucrose- and the optimum chemical treatment -50 mM $CaCl_2$- were set, further investigations were performed to investigate the time dependency of 50 mM $CaCl_2$. Chemical treatments with 50 mM $CaCl_2$ ranged from 0 to 60 min, where it was found that the highest transformation efficiency was obtained after 15 min incubation, with a 5-fold increase in transformation efficiency compared to the sample without chemical treatment.

The optimum electroporation voltage was also investigated; voltages from 2.1 to 2.5 kV were tested, being 2.3 kV (11.5 kV/cm) the one that showed the highest transformation efficiency for 2 mm cuvettes. Also, different DNA concentrations were investigated, and very similar transformation efficiencies were obtained with DNA concentrations ranging from 0.25 to 1.00 μg DNA.

Hence, the final optimised method for an efficient transformation by electroporation for *C. necator* H16 consisted on cultivate cells in NB to an OD_{600} of 0.6, followed by the chemical treatment with 50 mM $CaCl_2$ for 15 min in ice and two further cell washes with 0.2 M sucrose before the cells were electroporated with 2.3 kV in 2 mm cuvettes, where a maximum transformation efficiency of 3.86×10^5 transformants/μg DNA was achieved.

The main advantage of optimising this protocol is that it represents a tool for constructing recombinant strains is a shorter time, as the transformation can be accomplished only after 40 h, while transformation by bacterial conjugation requires from 4 to 5 days to be achieved.

2.4.2 Synthetic biology toolbox construction

As mentioned in **Section 2.1.2**, the design of the synthetic biology toolbox was based on previous synthetic biology parts reported to function efficiently in *C. necator* H16, where four plasmids were constructed pursuing to develop

metabolic capabilities in future experiments; the plasmids were transferred to *C. necator* H16 using the transformation by electroporation method optimised in our laboratory.

All plasmid designs were based on the premise that *C. necator* H16 has shown preference and replicability with broad-host-range plasmids, being pBBR1 derived plasmids one of the plasmids most commonly used for *C. necator* H16. The pBBR1MCS-1 (4.7 kb) plasmid (**Figure 2.4**) was chosen as a plasmid backbone to create a synthetic toolbox for *C. necator* H16, as it has proven to be replicable in the strain due to its pBBR1 oriV origin of replication (Gruber *et al*., 2014), as the origin of replication has to be compatible with the replication machinery of the host cell to be able to regulate gene expression (Li *et al*., 2015).

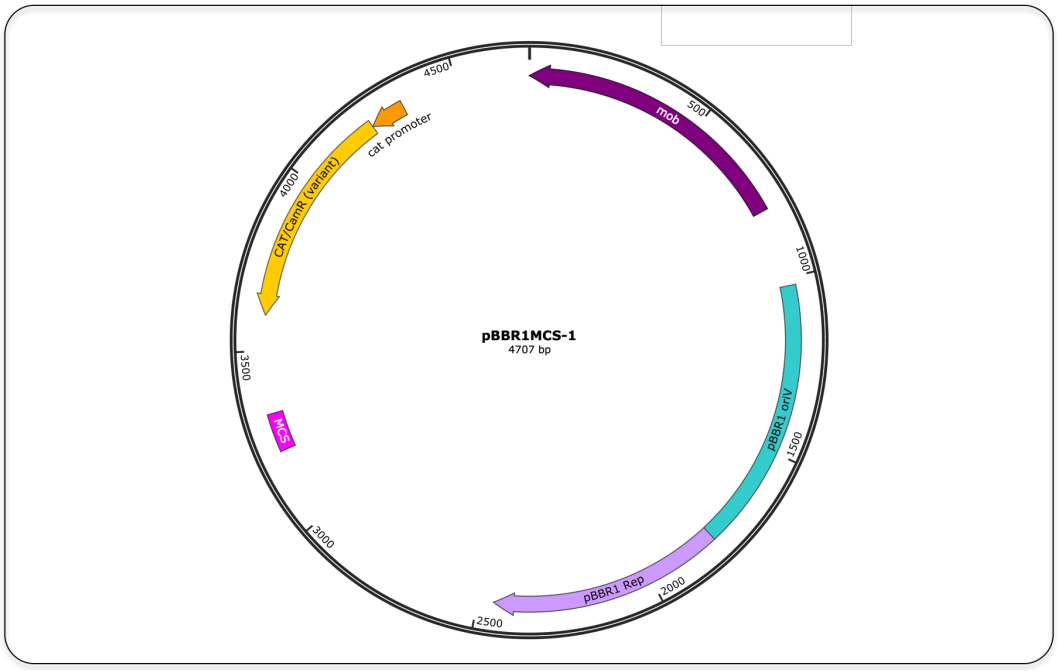


Figure 2.4 Illustration of pBBR1MCS-1 map (4.7 kb). The plasmid encodes for pBBR1 Rep, pBBR1 oriV origin of replication, mob gene (mob) for plasmid mobilisation, CAT promoter, and chloramphenicol resistance (CamR).

The promoters chosen for the versatile construction of plasmids for *C. necator* H16 were the inducible promoter P_{BAD}, which relies on the induction with L-arabinose; and the constitutive promoter P_{j5}. The inducible P_{BAD} promoter was cloned from the pBbA8k-RFP plasmid (3.9 kb) (**Figure 2.5**), this plasmid was not

used directly for gene expression in *C. necator* H16 due to it is a narrow range plasmid (Lee *et al*., 2011). The constitutive promoter was constructed by two subsequent PCR amplifications with the P_{j5} sequence contained within the primers. For this part of the study, RFP and enhanced green fluorescent protein (eGFP) proteins were used as fluorescence protein reporters to be expressed in the different constructed plasmids systems in *C. necator* H16 to demonstrate the successful replicability of all the plasmids constructed.

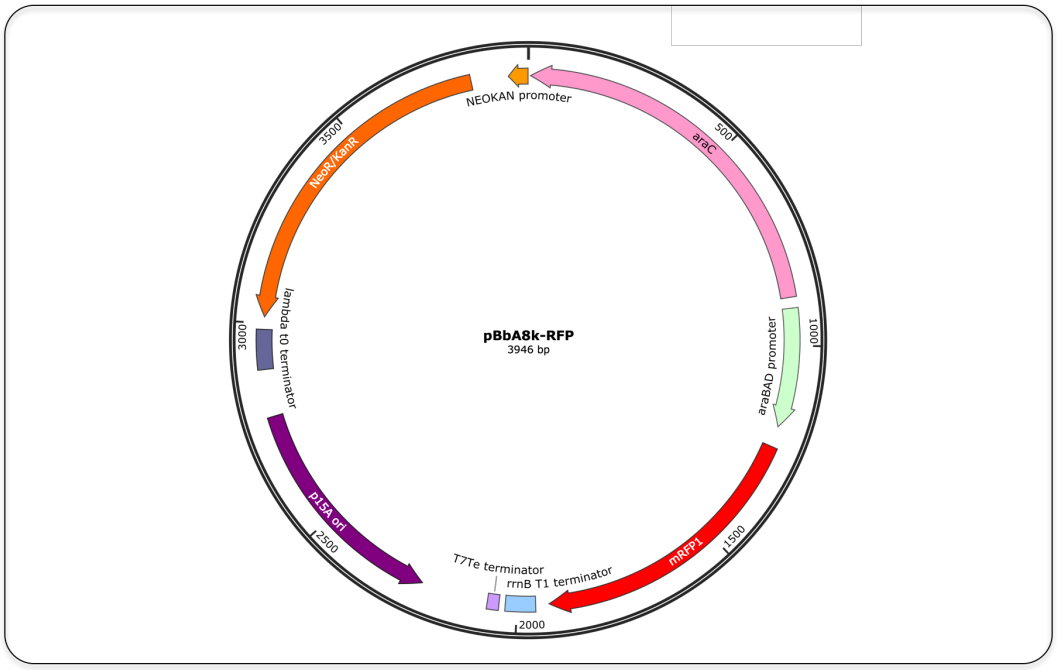


Figure 2.5 Illustration of pBbA8k-RFP map (3.9 kb). The plasmid encodes for the NEOKAN promoter, kanamycin resistance (KanR), araC and P_{BAD} promoter, and rfp protein.

The structure of the synthetic toolbox of the four plasmids constructed (pBADk.rbs-RFP, pBADc.rbs-RFP, pPj5c.T7rbs-RFP, and pPj5c.T7rbs-eGFP) with the different biological parts considered for the versatility of the toolbox – promoters, 5' mRNA stem-loop, RBS, terminators- is shown in **Figure 2.6**, where two different antibiotic resistance markers (chloramphenicol and kanamycin) and fluorescence reporters (RFP and eGFP) were included in the synthetic toolbox.

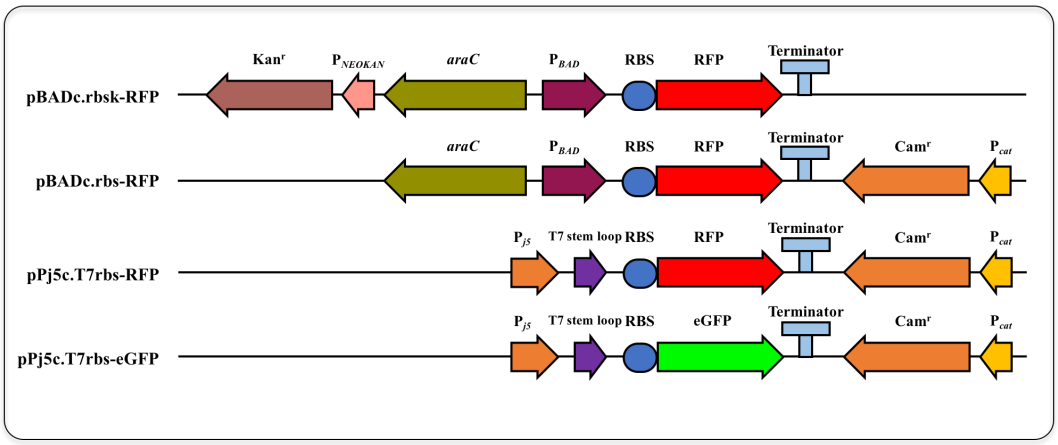


Figure 2.6 Synthetic biology toolbox for C. necator H16 for inducible and constitutive protein expression. All plasmids constructed harbour the pBBR1 oriV origin of replication, the pBBR1 Rep and mob genes.

2.4.2.1 Construction of inducible synthetic biology toolbox for protein expression

The construction of the arabinose-inducible pBADk.rbs-RFP plasmid was successfully achieved with the amplification of the pBBR1 Rep, pBBR1 oriV and *mob* gene from pBBR1MCS-1, which was inserted into the *AvrII* and *SacI* sites of pBbA8k-RFP harbouring the Kan^r cassette, L-arabinose-inducible system (*araC* and P_{BAD}), *E. coli* consensus RBS, and *rfp* gene. The RBS included in the synthetic biology toolbox was the *E. coli* consensus RBS taken from the digested pBbA8k-RFP plasmid, due to this synthetic RBS has been reported to provide the highest expression of RFP in *C. necator* H16 (Bi *et al*., 2013).

The PCR amplified fragment harbouring the pBBR1 Rep, pBBR1 oriV and *mob* gene from pBBR1MCS-1 was analysed to confirm the size of the insert, which was a 3.3 kb DNA band as expected (**Figure 2.7A**). Then, once the construction of the pBADk.rbs-RFP plasmid was successfully ligated and transformed into *E. coli* DH5α, a restriction enzyme analysis (ReA) was performed, where DNA bands of 3.3 kb and 3.0 kb confirmed the right sizes expected for the digested pBADk.rbs-RFP plasmid (**Figure 2.7B**). After *E. coli* DH5α transformation, protein expression of RFP with 0.1 % (w/v) L-arabinose at the early log phase was performed (**Figure 2.7C**). Then, the constructed plasmid was isolated from *E. coli* DH5α and used to transform *C. necator* H16 cells, followed by the induction

with 0.1 % (w/v) L-arabinose at the early log phase for RFP protein expression (**Figure 2.7D**), where the cell pellets of both cultures from L-arabinose-induced expression in *E. coli* DH5α and *C. necator* H16 were indeed red.

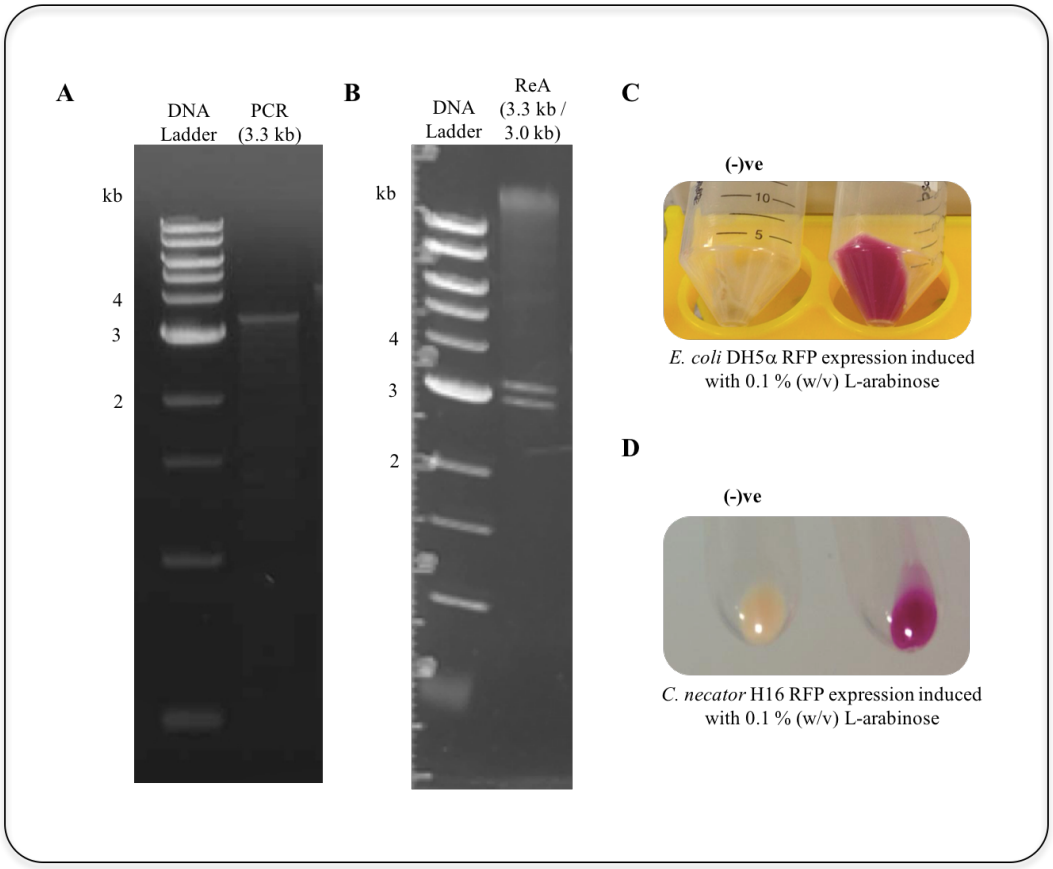


Figure 2.7 Analysis of the construction and RFP protein expression of pBADk.rbs-RFP in E. coli DH5α and C. necator H16. ***A)*** *Analysis of the PCR amplified insert from pBBR1MCS-1,* ***B)*** *Restriction enzyme analysis (ReA) with AvrII and SacI of the constructed pBADk.rbs-RFP,* ***C)*** *E. coli DH5α RFP protein expression after induction with 0.1 % (w/v) L-arabinose, left tube sample (-)ve control corresponds to a non-induced sample, right tube sample corresponds to an induced sample.* ***D)*** *C. necator H16 RFP protein expression after induction with 0.1 % (w/v) L-arabinose, left tube sample (-)ve control corresponds to a non-induced sample, right tube sample corresponds to an induced sample. Abbreviations: (-)ve, negative control; ReA, Restriction enzyme analysis.*

On the other hand, the construction of the other arabinose-inducible pBADc.rbs-RFP system with a chloramphenicol resistance cassette was successfully achieved as well with the amplification of the pBBR1 Rep, pBBR1 oriV, *mob* gene, and Cam^r cassette from pBBR1MCS-1, which was inserted into the *AvrII* and *PstI*

sites of pBbA8k-RFP harbouring the L-arabinose-inducible system (*araC* and P_{BAD}), *E. coli* consensus RBS, and *rfp* gene.

The PCR amplified fragment harbouring the pBBR1 Rep, pBBR1 oriV, *mob* gene, and Cam^r cassette from pBBR1MCS-1 was analysed to confirm the size of the insert, which was a 4.3 kb DNA band as expected (**Figure 2.8A**). Then, once the construction of the pBADc.rbs-RFP plasmid was successfully ligated and transformed into *E. coli* DH5α, a restriction enzyme analysis (ReA) was performed, where DNA bands of 4.3 kb and 2.4 kb confirmed the right sizes expected for the digested pBADc.rbs-RFP plasmid (**Figure 2.8B**). After *E. coli* DH5α transformation, protein expression of RFP with 0.1 % (w/v) L-arabinose at the early log phase was performed (**Figure 2.8C**). Then, the constructed plasmid was isolated from *E. coli* DH5α and used to transform *C. necator* H16 cells, followed by the induction with 0.1 % (w/v) L-arabinose at early log phase for RFP protein expression (**Figure 2.8D**), where the cell pellets of both cultures from L-arabinose-induced expression in *E. coli* DH5α and *C. necator* H16 were indeed red.

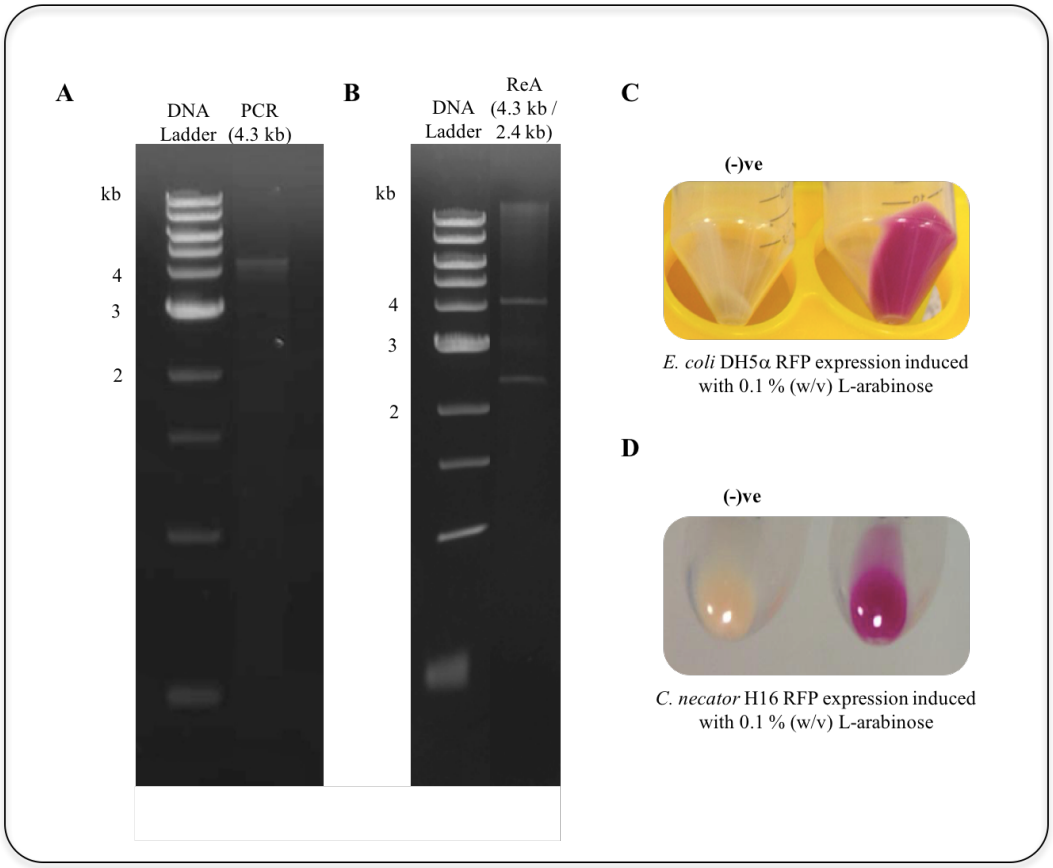


Figure 2.8 Analysis of the construction and RFP protein expression of pBADc.rbs-RFP in E. coli DH5α and C. necator H16. ***A)*** *Analysis of the PCR amplified insert from pBBR1MCS-1,* ***B)*** *Restriction enzyme analysis (ReA) with AvrII and PstI of the constructed pBADc.rbs-RFP,* ***C)*** *E. coli DH5α RFP protein expression after induction with 0.1 % (w/v) L-arabinose, left tube sample (-)ve control corresponds to a non-induced sample, right tube sample corresponds to an induced sample.* ***D)*** *C. necator H16 RFP protein expression after induction with 0.1 % (w/v) L-arabinose, left tube sample (-)ve control corresponds to a non-induced sample, right tube sample corresponds to an induced sample. Abbreviations: (-)ve, negative control; ReA, Restriction enzyme analysis.*

Both successful RFP expressions in *E. coli* DH5α and in *C. necator* H16 for the inducible synthetic biology toolbox constructed demonstrated that 1) the plasmids harboured the insert RFP, 2) that the transformation method optimised for *C. necator* H16 described in this **Chapter 2** allowed high transformation efficiencies with all plasmids transformed with transformation efficiencies higher than 1 x 10^5 transformants/µg of DNA, and that 3) the P_{BAD} promoter is compatible with *C. necator* H16 for protein expression.

2.4.2.2 Construction of constitutive synthetic biology toolbox for protein expression

For the construction of the synthetic biology toolbox for constitutive expression of proteins, the synthetic biology part T7 5' mRNA stem-loop structure was inserted upstream of the RBS for the *rfp* gene in both plasmids designed in order to increase further the RFP expression, as this biological part has been reported to enhance mRNA stability in *C. necator* H16 (Bi *et al*. 2013). The RBS included in this synthetic biology toolbox was the *E. coli* consensus RBS sequence as well, due to as mentioned before, this RBS has been reported to provide the highest expression of RFP in *C. necator* H16 (Bi *et al*., 2013).

The construction of the constitutive pPj5c.T7rbs-RFP plasmid was successfully achieved with the restriction enzyme digestion (ReD) of the pBADc.rbs-RFP plasmid previously constructed harbouring the pBBR1 Rep, pBBR1 oriV, *mob* gene, and Cam^r cassette; and from the insertion of the PCR amplified *rfp* gene from pBbA8k-RFP. The first PCR that amplified the *rfp* gene from pBbA8k-RFP was used to insert as well the *E. coli* consensus RBS and the T7 stem-loop sequences, which DNA sequences were contained within the T7-RBS-RFP-fwd primer, subsequently, after the purification of the first amplicon, a second round of PCR was carried out to insert the P_{j5} promoter sequence, which DNA sequence was contained within the Pj5-T7-fwd primer. The amplicon harbouring the P_{j5}, T7 stem-loop, *E. coli* consensus RBS, and the *rfp* gene was analysed in a gel to confirm its right size, which was a 0.9 kb DNA band as expected (**Figure 2.9A**). Then, once the construction of the pPj5c.T7rbs-RFP plasmid was successfully ligated and transformed into *E. coli* DH5α, a PCR amplification of the insert within the newly constructed plasmid was performed, as this short insert (0.9 kb) was not visible in the gel after restriction enzyme analysis (ReA); also a ReA was performed to confirm the 4.5 kb DNA band corresponding to the vector of the constructed pPj5c.T7rbs-RFP plasmid (**Figure 2.9B**). After *E. coli* DH5α transformation, constitutive protein expression of RFP was performed (**Figure 2.9C**). Then, the constructed plasmid was isolated from *E. coli* DH5α and used to transform *C. necator* H16 cells, followed by the constitutive RFP protein

expression (**Figure 2.9D**), where the cell pellets of both cultures in *E. coli* DH5α and *C. necator* H16 were indeed red.

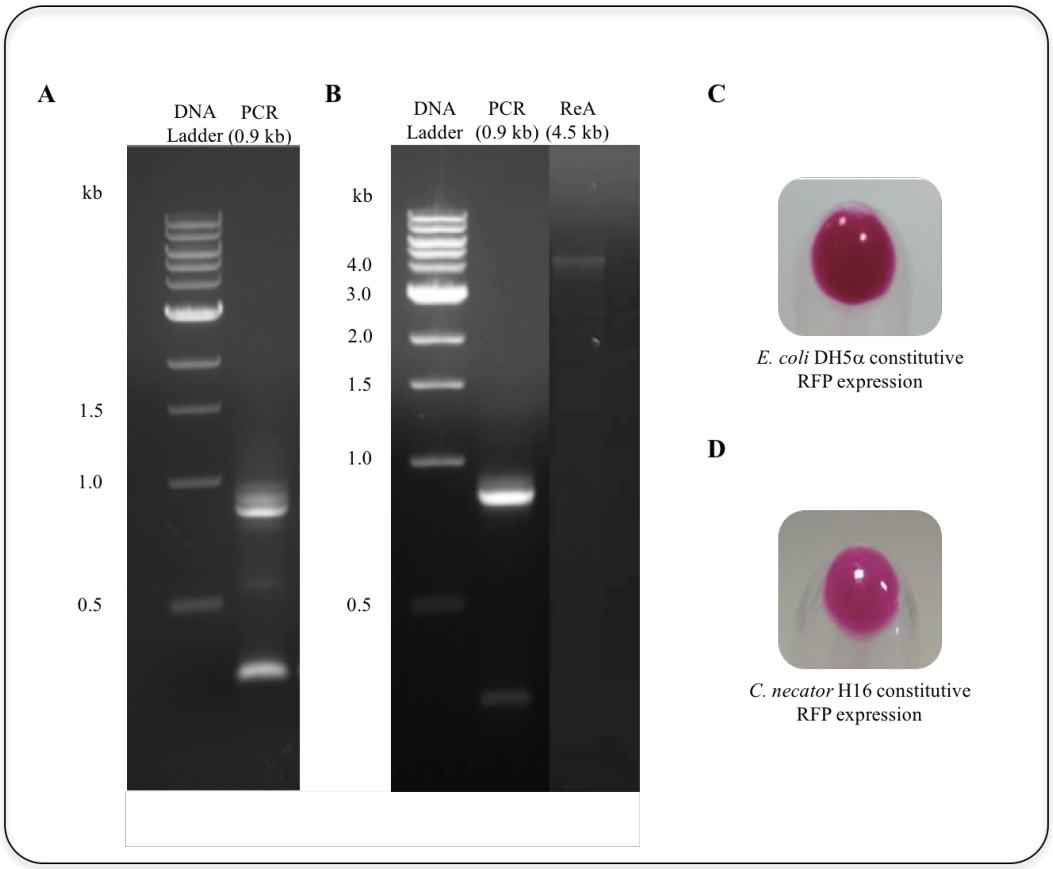


Figure 2.9 Analysis of the construction and RFP protein expression of pPj5c.T7rbs-RFP in E. coli DH5α and C. necator H16. ***A)*** *Analysis of the PCR amplified insert from pBbA8k-RFP,* ***B)*** *PCR of the amplified insert from pPj5c.T7rbs-RFP and restriction enzyme analysis (ReA) with PstI and XhoI of the constructed pPj5c.T7rbs-RFP,* ***C)*** *E. coli DH5α constitutive RFP protein expression.* ***D)*** *C. necator H16 constitutive RFP protein expression. Abbreviations: ReA, Restriction enzyme analysis.*

Then, the construction of the other constitutive pPj5c.T7rbs-eGFP plasmid with chloramphenicol resistance as well was successfully achieved using the previously constructed plasmid pPj5c.T7rbs-RFP as plasmid backbone, which was digested with *NdeI* and *XhoI* to replace the *rfp* gene for the *egfp* gene from pEGP plasmid. Thus, the pPj5c.T7rbs-eGFP is a constitutive system that harbours a P_{j5} promoter, T7 stem-loop, and *E. coli* consensus RBS.

The amplicon harbouring the P_{j5}, T7 stem-loop, *E. coli* consensus RBS, and the *egfp* gene was analysed in a gel to confirm its right size, which was a 0.9 kb DNA band as expected (**Figure 2.10A**). Then, once the construction of the pPj5c.T7rbs-eGFP plasmid was successfully ligated and transformed into *E. coli* DH5α, a PCR amplification of the insert within the newly constructed plasmid was performed, as this short insert (0.9 kb) was not visible in the gel after restriction enzyme analysis (ReA) as observed previously with the insert of pPj5c.T7rbs-RFP; also a ReA was performed to confirm the 4.6 kb DNA band corresponding to the vector of the constructed pPj5c.T7rbs-eGFP plasmid (**Figure 2.10B**). After *E. coli* DH5α transformation, constitutive protein expression of eGFP was performed (**Figure 2.10C**). Then, the constructed plasmid was isolated from *E. coli* DH5α and used to transform *C. necator* H16 cells, followed by the constitutive eGFP protein expression (**Figure 2.10D**), where the cell pellets of both cultures in *E. coli* DH5α and *C. necator* H16 were indeed green.

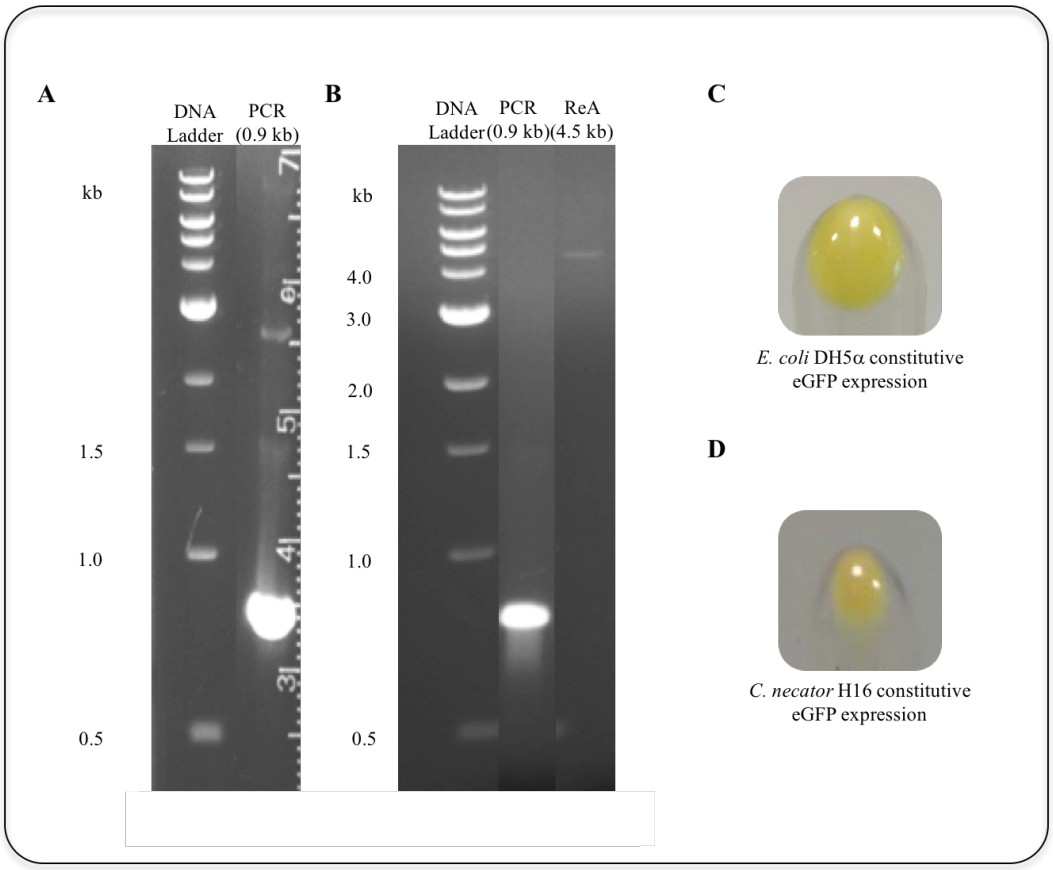


Figure 2.10 Analysis of the construction and eGFP protein expression of pPj5c.T7rbs-eGFP in E. coli DH5α and C. necator H16. ***A)*** *Analysis of the PCR amplified insert from pPj5c.T7rbs-RFP,* ***B)*** *PCR of the amplified insert from pPj5c.T7rbs-eGFP and restriction enzyme analysis (ReA) with NdeI and XhoI of the constructed pPj5c.T7rbs-eGFP,* ***C)*** *E. coli DH5α constitutive eGFP protein expression.* ***D)*** *C. necator H16 constitutive eGFP protein expression. Abbreviations: ReA, Restriction enzyme analysis.*

Both successful RFP and eGFP expressions in *E. coli* DH5α and in *C. necator* H16 for the constitutive synthetic biology toolbox constructed demonstrated that 1) the plasmids harboured the RFP and eGFP inserts, and 2) that the transformation method optimised for *C. necator* H16 described in this **Chapter 2** allowed high transformation efficiencies with all plasmids transformed with transformation efficiencies higher than 1 x 10^5 transformants/μg of DNA.

2.5 Conclusions

A contribution to optimise the transformation method of *C. necator* H16 by electroporation was developed in this part of the study. The final optimised method consisted of a chemical treatment (50 mM $CaCl_2$), the use of an efficient transformation buffer (0.2 M sucrose) and an intermediate electric field strength (11.5 kV/cm), which were key factors in the process, where a transformation efficiency higher than 1 x 10^5 transformants/μg of DNA was achieved. The development of this method was used to efficiently transform *C. necator* H16 with the versatile synthetic biology toolbox reported in this **Chapter 2**. The optimisation of the transformation method for *C. necator* H16 represents a significant milestone for genetic manipulation, which would help to develop metabolic engineering strategies.

The synthetic biology toolbox is a potential system for metabolic engineering applications in *C. necator* H16, as metabolic engineering and synthetic biology usually focus on a narrow set of microbial host, such as *E. coli* and *Saccharomyces cerevisiae*, due to the ease of genetic manipulation and the heterologous gene expression tools already available for these microorganisms, although these strains might have limitations for diverse industrial applications. *C. necator* H16 is a microbial host with both metabolic capabilities and growth conditions that are promising for industrial applications as it is a microorganism that can function as a chemolithoautotroph, therefore the construction of the synthetic toolbox explicitly created for *C. necator* H16 is required for chemical production.

In this study, plasmids with two different promoters were successfully transformed by electroporation in *C. necator* H16, with inducible and constitutive systems that have responded to the expression of RFP and eGFP as models proteins. The design of both systems will be useful for metabolic engineering approaches.

CHAPTER 3

Metabolic engineering:

Applicability of synthetic biology toolbox by the expression of heterologous extended biotin operons in *Cupriavidus necator* H16

Abstract

In **Chapter 3**, the synthetic biology toolbox with the inducible system previously constructed in this work was used to test the heterologous expression of two clusters of genes in *Cupriavidus necator* H16. The cluster of genes encoded for extended biotin operons from *Desulfovibrio vulgaris* (four genes) and *Desulfovibrio desulfuricans* (five genes) respectively. The function of the genes of the extended biotin operons remains unknown to date, although it is believed that these genes could be related to pimelic acid synthesis. The extended biotin operons were expressed in *E. coli* BW25113, *E. coli* BW25113 ΔbioC, and in *C. necator* H16. The protein expression after the induction of the P_{BAD} promoter showed the target proteins of most of the genes of the extended biotin operons of *Desulfovibrio* spp. in the SDS-PAGE in all host cells, although there was no full evidence that all the four and five genes of the extended biotin operons had been expressed simultaneously, and hence the function of these genes could not be determined.

3 Chapter 3. Metabolic engineering: Applicability of the synthetic biology toolbox by the expression of heterologous extended biotin operons in *Cupriavidus necator* H16

3.1 Introduction

The applicability of the synthetic biology toolbox described in **Chapter 2** was tested for metabolic engineering in *C. necator* H16 for the expression of a cluster of genes corresponding to extended biotin operons of *Desulfovibrio vulgaris* and *Desulfovibrio desulfuricans*. As mentioned in **Section 1.5.7** and **1.5.7.4**, an interesting bioproduct that could be produced in *C. necator* H16 is pimelic acid, which at the same time can be used to overproduce biotin in microorganisms, as it is an important precursor compound for biotin biosynthesis. The importance of applying metabolic engineering to microbial strains for overproduction of pimelic acid is because natural pimelate biosynthesis is low in wild-type strains, as biotin is required at very low concentrations in microorganisms.

The synthesis of biotin is widely conserved among microorganisms, and as mentioned in **Section 1.5.7.3**, it can be divided into two stages: 1) the synthesis of the pimelic acid moiety, which carbon atoms serve for biotin biosynthesis, and 2) the assembly of the bicyclic ring of biotin. The genes for the second stage of the biotin biosynthesis are encoded in the *bioFADB* operon in many microorganisms, while the first stage usually varies among different species, although microorganisms such as *E. coli*, and *Bacillus* spp., among other bacteria, have shown to use BioC and BioH, or BioX and BioI to synthesise the pimelate moiety (**Figure 1.8**).

As mentioned in **Section 1.5.7.4**, Rodionov *et al*. (2004) studied the regulatory and genetic factors in pathways that are involved in the biosynthesis of cofactors of the metal-reducing genera in the δ subgroup of proteobacteria, and they found that *Desulfovibrio vulgaris* (DV) and *Desulfovibrio desulfuricans* (DD) had an

extended biotin operon within the standard *bioFADB* with four and five new genes (respectively), and no *bioC*, *bioH* or similar genes were found in their biotin operon, while the rest of the δ-proteobacteria showed to have the standard *bioFADB* for the ring assembly of biotin and also *bioC* and *bioH*, *bioW* or *bioC* and *bioG* pairs known for its function for pimelic acid synthesis (**Figure 1.10** and **Figure 1.11**). This fact suggested that the new genes found within the *bioFADB* biotin operon of *Desulfovibrio* spp could be related to pimelic acid production. Among the genes found within the *bioFADB* operon of DV and DD, there are homologs of acyl carrier protein (ACP), 3-oxoacyl-ACP synthase, 3-oxoacyl-ACP reductase and hydroxymyristol-ACP dehydratase (**Table 3.1**). Due to the four (DV) and five (DD) new genes of the extended biotin operon might be related with the fatty-acid biosynthesis, this suggests that they could have similar functions as the *bioC* and *bioH* genes in *E. coli*, which are the genes that provide the means to use fatty acid synthesis for the assembly of pimelate.

Table 3.1 New genes found within the *bioFADB* operon of *Desulfovibrio vulgaris* and *Desulfovibrio desulfuricans.*

Genes	Locus tag (Old locus tag)	Original annotation[a]	Position
Desulfovibrio vulgaris (DV)			
birA	DVU2557	birA bifunctional protein	2668667..2669659
bioB	DVU2558	biotin synthase	2669895..2670827
bioA	DVU2559	adenosylmethionine-8-amino-7-oxononanoate transaminase	2670824..2672452
DV1	DVU2560	3-hydroxyacyl-[acyl carrier protein] dehydratase	2672449..2672835
DV2	DVU2561	3-oxoacyl-[acyl carrier protein] reductase	2672832..2673860
DV3	DVU2562	acyl carrier protein	2673960..2674367
DV4	DVU2563	3-oxoacyl-[acyl carrier protein] synthase II	2674645..2676093
bioF	DVU2564	8-amino-7-oxononanoate synthase	2676659..2677933
bioD	DVU2565	dethiobiotin synthetase	2677915..2678655
Desulfovibrio desulfuricans (DD)			
birA	DDE_RS12230 (Dde_2654)	bifunctional protein	2667956..2668987
bioB	DDE_RS12235 (Dde_2655)	biotin synthase	2669096..2670070
bioA	DDE_RS12240 (Dde_2656)	adenosylmethionine-8-amino-7-oxononanoate aminotransferase	2670067..2671446
DD1	DDE_RS17745 (Dde_2657)	3-hydroxyacyl-ACP dehydratase	2671439..2671894
DD2	DDE_RS12250 (Dde_2658)	short-chain dehydrogenase/reductase SDR (3-oxoacyl-[acyl-carrier protein] reductase)	2671885..2672640
DD3	DDE_RS12255	phosphopantetheine-binding protein	2672655..2673077

	(Dde_2659)		
DD4	DDE_RS12260 (Dde_2660)	3-ketoacyl-ACP synthase	2673067..2674377
DD5	DDE_RS12265 (Dde_2661)	hypothetical protein	2674374..2674883
bioF	DDE_RS12270 (Dde_2662)	8-amino-7-oxononanoate synthase	2674865..2676022
bioD	DDE_RS12275 (Dde_2663)	dethiobiotin synthase	2676019..2676735

DV1, *DV2*, *DV3*, and *DV4* correspond to the four new genes found within the *bioFADB* biotin operon of *Desulfovibrio vulgaris* (DV); *DD1*, *DD2*, *DD3*, *DD4* and *DD5* correspond to the five new genes found within the *bioFADB* biotin operon of *Desulfovibrio desulfuricans* (DD). Data of locus tag and original annotation was obtained from KEGG.

In order to explore the genes of unknown function of the extended biotin operons of DV and DD, the first stages of the experiments in this chapter were designed to express the extended biotin operons initially in *E. coli* BW25113 and nutrient-rich media to corroborate the expression of all the genes found within the biotin operons, then, the second stage was designed to express the extended biotin operons in an *E. coli* BW25113 ΔbioC strain -a mutant deficient in *bioC* synthesis, thus, the mutant strain is not capable of growing in media where biotin is not supplemented- in biotin-free media, this would verify if the extended biotin operons of DV and DD had a similar function to the *bioC* or *bioH* genes and were effectively related or not with pimelic acid synthesis. Moreover, if the extended biotin operons resulted to be related with pimelic acid biosynthesis, these could be transformed into *C. necator* H16 and be used for pimelate production.

The use of a single promoter for the expression of the extended biotin operon from DV and DD was based on the premise that in prokaryotic cells, proteins with related functions are usually clustered in an operon and are transcribed together with a single promoter, where the regulation of the transcription of all the genes that encode for proteins that catalyse a single biochemical pathway are controlled all at once, due to the cluster of genes are all required, or none of them needed.

3.2 Aim

The aim of this **Chapter 3** is:

- To demonstrate the applicability of the synthetic biology toolbox constructed in **Chapter 2** with the expression of operons driven by a single strong promoter (inducible promoter P_{BAD}) in *C. necator* H16, operons that could encode for a modified pathway committed to pimelate production.

3.3 Materials and methods

3.3.1 Materials

List of detailed information on the materials used (reagents, kits, equipment, software, media preparation and miscellaneous) can be found in **Appendix 1 – 7**.

3.3.2 Strains, plasmids, and primers

All strains, plasmids and primers that were used in this **Chapter 3** are shown in **Table 3.2**. The four genes of the extended biotin operon of DV are indicated as *bioDV1-4*, and the five genes of the extended biotin operon of DD are indicated as *bioDD1-5* in all the experiments and data shown in this **Chapter 3**.

Table 3.2 Strains, plasmids, and primers used in this study.

Strains, plasmids or primers	Description	References or source
Strains		
E. coli BW25113	Wild-type Kan^r	Datsenko *et al.*, 2000
E. coli BW25113 ΔbioC	Kan^r, mutant deficient in *bioC* synthesis	Lab collection
E. coli DH5α	Standard cloning strain	Lab collection
C. necator H16	Wild-type Gen^r	DSM 428
Plasmid backbones		

pBADc.rbs-RFP	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, *araC*, P_{BAD}, RBS, *rfp*	This work
pUC57-bioDV1-4	Plasmid carrying the *bioDV1-4* biotin synthetic operon codon optimised for *E. coli* BW25113 with four genes and restriction enzymes between the genes: *NdeI* and *SpeI* for *DV1*; *SpeI* and *SacI* for *DV2*; *SacI* and *ClaI* for *DV3*, and *ClaI* and *BamHI* and *XhoI* for *DV4*.	GenScript
pUC57-bioDD1-5	Plasmid carrying the *bioDD1-5* biotin synthetic operon codon optimised for *E. coli* BW25113 with five genes and restriction enzymes between the genes: *NdeI* and *SpeI* for *DD1*; *SpeI* and *SacI* for *DD2*; *SacI* and *ClaI* for *DD3*, and *ClaI* and *SalI* for *DD4*, *SalI* and *BamHI* and *XhoI* for *DD5*.	GenScript
Plasmids constructed		
pBADc.rbs-bioC	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, *araC*, P_{BAD}, RBS, *bioC*	This work
pBADc.rbs-bioDV1-4	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, *araC*, P_{BAD}, RBS, *bioDV1-4* biotin operon	This work
pBADc.rbs-bioDD1-5	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, *araC*, P_{BAD}, RBS, *bioDD1-5* biotin operon	This work
Primers (5' → 3')		
bioC-fwd	GATCCATATGGCAACGGTTAATAAACA AGCCATTG	This work
bioC-rev	GATCCTCGAGTAACGTTTACTCACGAG CAATCACTC	This work

All the RBS noted in all the plasmids in the list correspond to the sequence encoding for the *E. coli* consensus RBS. *DV1*, *DV2*, *DV3*, and *DV4* correspond to the new genes found within the *bioFADB* extended biotin operon of *Desulfovibrio vulgaris* (DV); *DD1*, *DD2*, *DD3*, *DD4* and *DD5* correspond to the new genes found within the *bioFADB* extended biotin operon of *Desulfovibrio desulfuricans* (DD). Restriction enzyme sequences of primers are underlined. Abbreviations: Kan^r, kanamycin resistance; Gen^r, gentamicin resistance; Cam^r, chloramphenicol resistance; RBS, ribosome binding site; *rfp*; red fluorescence protein gene.

3.3.3 Cultivation of *E. coli* DH5α, *E. coli* BW25113, *E. coli* BW25113 ΔbioC, and *C. necator* H16

E. coli DH5α cells were cultivated at 37 ºC on 2x YT medium and 250 rpm, and when required with chloramphenicol [25 μg/mL] according to the application, the strain was used for all molecular cloning, plasmid propagation, and maintenance. *E. coli* BW25113 and *E. coli* BW25113 ΔbioC cells were cultivated at 37 ºC on 2x YT medium, M9 minimal biotin-free medium, or M9 minimal medium supplemented with biotin at 250 rpm and with kanamycin [50 μg/mL] and when needed with chloramphenicol [25 μg/mL] according to the application. The *E. coli* BW25113 ΔbioC was used for complementation studies. *C. necator* H16 cells

were cultivated at 30 ºC using nutrient broth (NB) and 250 rpm always supplemented with gentamicin [10 μg/mL] and when needed with chloramphenicol [25 μg/mL]. The optical density was measured at 600 nm with a BioPhotometer.

3.3.4 Bacterial transformation of *E. coli* DH5α and *C. necator* H16

*E. coli DH5*α cells were transformed using a standard chemical transformation method. *E. coli* DH5α was used for circularization and propagation of ligation mixtures of all plasmids constructed. *C. necator* H16 cells were transformed by the electroporation method optimised in our laboratory group and described in **Chapter 2**.

3.3.4.1 Transformation using $CaCl_2$ heat-shock method for *E. coli* DH5α

A pre-culture of *E. coli* DH5α was cultivated in 5 mL of 2x YT for 16 h at 37 ºC; for competent cell preparation, a fresh 2x YT falcon tube was inoculated with the pre-culture at a 1:100 dilution and cultivated at 37 ºC, when cells reached an optical density at 600 nm (OD_{600}) of 0.5, 1 mL of cells were transferred to a sterile 1.5 mL microcentrifuge tube. Cells were centrifuged at maximum speed (17,000 x g) for 30 s, and the supernatant was removed by pipetting. The pellet was washed once with 500 μL of pre-chilled 50 mM $CaCl_2$. Then, the cell pellet was resuspended carefully in 500 μL of pre-chilled 50 mM $CaCl_2$ and incubated in ice for 30 min (for transforming intact plasmid, a 10 min incubation was sufficient). Plasmid DNA was added to a concentration of 1 μg or 5 μL of ligation mixture and mixed gently. A second 30 min incubation step was performed after the addition of the plasmid (for transforming intact plasmid, a 10 min incubation was sufficient). After the second incubation, the cells were heat-shocked at 42 ºC for 1 min and further incubated in ice for 2 min. After heat shock and ice incubation time, 800 μL of 2x YT was added, and cells were left to grow for a 1 h outgrowth

at 37 ºC. After the outgrowth, cells were centrifuged at maximum speed for 30 s and most of the media removed, the remaining 200 – 300 µL media was used to resuspend the cells gently before plating them on TYE agar plates with the required antibiotic, and incubated overnight at 37 ºC.

3.3.4.2 Transformation by electroporation method for *C. necator* H16

A pre-culture of *C. necator* H16 was cultivated in 5 mL of NB with gentamicin for 40 – 44 h at 30 ºC; for electrocompetent cell preparation, a fresh NB supplemented with gentamicin falcon tube was inoculated with the pre-culture at a 1:50 dilution and cultivated at 30 ºC, when cells reached an OD_{600} of 0.5 – 0.7, cells were transferred to ice and chilled for 5 min. Two millilitres of the cells were then transferred to a sterile 2 mL microcentrifuge tube. Cells were centrifuged at maximum speed (17,000 x g) for 30 s and the supernatant was removed. Then, the cell pellet was washed once with 1 mL of pre-chilled 50 mM $CaCl_2$ and incubated for 15 min in ice. Then the cell pellet was centrifuged at maximum speed for 30 s followed by two washes by resuspension in 1 mL of pre-chilled 0.2 M Sucrose. After the final wash, the cell pellet was resuspended in 100 µL of pre-chilled 0.2 M Sucrose. Plasmid DNA was added to a concentration of 0.5 µg to the resuspended cells and mixed gently. The resuspension was then transferred into a pre-chilled 2 mm electroporation cuvette and electroporated at 2.3 kV. After electroporation, 1 mL of NB was added immediately directly to the electroporation cuvette, then, cells were transferred to a new 2 mL microcentrifuge tube for a 2 h outgrowth at 30 ºC. After the outgrowth, typically about 10 % (v/v) cells were plated in NB with gentamicin and chloramphenicol agar plates and incubated at 30 ºC for 40 – 48 h.

3.3.5 DNA preparation

Standard procedures were used for genomic DNA isolation, isolation of plasmids, restriction enzyme digestions, polymerase chain reaction (PCR) with Q5 DNA

polymerase, *DpnI* digestions, DNA gel extraction, PCR purification, and T4 DNA ligation (Sambrook *et al*., 2011) and recommendations by the manufacturers. All primers were synthesised by Eurofins Genomics, *bioDV1-4* and *bioDD1-5* extended biotin operons were synthesised by GenScript (Piscataway, USA).

3.3.6 DNA Gel electrophoresis

An agarose gel electrophoresis was prepared to analyse DNA; 0.7 % or 1.0 % (w/v) agarose gels were prepared in 1x TBE buffer (prepared from a 5x TBE buffer composed of 54.0 g of Tris base, 27.5 g of Boric acid, and 20 mL of 0.5 M EDTA per 1 L, pH 8.3) by dissolving 0.35 g or 0.50 g of agarose respectively in 50 mL buffer. The percentage of gel used was dependent on the purpose of the gel, if only an analysis of the DNA was required, then 1.0 % (w/v) gels were used, but if gel extraction was required, then 0.7 % (w/v) agarose gels were used to excise the DNA band from the gel. To ensure that the agarose is fully dissolved in the buffer, the solution was microwaved until all the agarose was dissolved, then, when the gel had cooled down, 2 μL of ethidium bromide was added and the gel was cast using a gel caster, then the comb was inserted and the gel was left to room temperature until it solidified. 6 μL of DNA Ladder 1 kb and appropriate volume of DNA samples were loaded into the gel. The electrophoresis was run at a constant voltage of 100 V for 60 min. The gel image was captured with a gel documentation system.

3.3.7 Plasmid construction

All plasmid constructed in this **Chapter 3** were based on the plasmid backbone pBADc.rbs-RFP plasmid from the synthetic biology toolbox described in **Chapter 2** and illustrated in **Figure 3.1**. The pBADc.rbs-RFP plasmid was digested with *NdeI* and *XhoI* to harbour the pBBR1 Rep, pBBR1 oriV origin of replication, *mob* gene, chloramphenicol resistance (Cam^r) cassette (containing P_{cat}

and *Cam*^{r}), L-arabinose-inducible system (containing *araC* and P_{BAD}), and *E. coli* consensus RBS; the *rfp* gene was replaced with either *bioC*, *bioDV1-4* extended biotin operon or *bioDD1-5* extended biotin operon. The synthetic codon optimised *bioDV1-4* and *bioDD1-5* extended biotin operons DNA sequences for *E. coli* BW25113 can be found in **Appendix 7**. The primers and *in silico* analyses of plasmid construction were generated using the SnapGene software tool. Correct construction of all plasmids was confirmed by restriction enzyme analysis (ReA).

3.3.7.1 Construction of plasmid pBADc.rbs-bioC

pBADc.rbs-bioC was constructed by first extracting total genomic DNA from *E. coli* BW25113, where the *bioC* gene was amplified from genomic DNA by PCR with bioC-fwd (forward primer) and bioC-rev (reverse primer) primers. Subsequently, the amplified fragment was subjected to *DpnI* digestion, PCR purification and restriction enzyme digestion with *NdeI* and *XhoI*, followed by a further PCR purification for ligation with T4 DNA ligase with the *NdeI* and *XhoI* digested and gel extracted fragment containing the pBBR1 Rep, pBBR1 oriV origin of replication, *mob* gene, Cam^r cassette, L-arabinose-inducible system (containing *araC* and P_{BAD}), and *E. coli* consensus RBS from pBADc.rbs-RFP resulting in the 6.8 kb plasmid pBADc.rbs-bioC. After transformation of *E. coli* DH5α, a single colony of the resulting plasmid pBADc.rbs-bioC was isolated and analysed by restriction enzyme analysis to confirm the right inserts before it was transformed into *E. coli* BW25113, *E. coli* BW25113 ΔbioC and *C. necator* H16.

3.3.7.2 Construction of plasmid pBADc.rbs-bioDV1-4

The synthetic *bioDV1-4* extended biotin operon was synthesised by GenScript, where the entire operon (*bioDV1-4*) harboured the four new genes from the extended biotin operon of DV, the genes were codon optimised for *E. coli* BW25113 and specific restriction enzymes and spacers were inserted in between the genes. pBADc.rbs-bioDV1-4 was constructed by digesting with *NdeI* and

XhoI the *bioDV1-4* operon from the pUC57-bioDV1-4 plasmid (plasmid obtained from GenScript). The digested *bioDV1-4* operon was then used for ligation with T4 DNA ligase with the *NdeI* and *XhoI* digested and gel extracted fragment containing the pBBR1 Rep, pBBR1 oriV origin of replication, *mob* gene, Cam^r cassette, L-arabinose-inducible system (containing *araC* and P_{BAD}), and *E. coli* consensus RBS from pBADc.rbs-RFP; resulting in the 9.5 kb plasmid pBADc.rbs-bioDV1-4. After transformation of *E. coli* DH5α, a single colony of the resulting plasmid pBADc.rbs-bioDV1-4 was isolated and analysed by restriction enzyme analysis to confirm the right inserts before it was transformed into *E. coli* BW25113, *E. coli* BW25113 ΔbioC and *C. necator* H16.

3.3.7.3 Construction of plasmid pBADc.rbs-bioDD1-5

The synthetic *bioDD1-5* extended biotin operon was synthesised by GenScript, where the entire operon (*bioDD1-5*) harboured the five new genes from the extended biotin operon DD, the genes were codon optimised for *E. coli* BW25113 and specific restriction enzymes and spacers were inserted in between the genes. pBADc.rbs-bioDD1-5 was constructed by digesting with *NdeI* and *XhoI* the *bioDD1-5* operon from the pUC57-bioDD1-5 plasmid (plasmid obtained from GenScript). The digested *bioDD1-5* operon was then used for ligation with T4 DNA ligase with the *NdeI* and *XhoI* digested and gel extracted fragment containing the pBBR1 Rep, pBBR1 oriV origin of replication, *mob* gene, Cam^r cassette, L-arabinose-inducible system (containing *araC* and P_{BAD}), and *E. coli* consensus RBS from pBADc.rbs-RFP; resulting in the 9.7 kb plasmid pBADc.rbs-bioDD1-5. After transformation of *E. coli* DH5α, a single colony of the resulting plasmid pBADc.rbs-bioDD1-5 was isolated and analysed by restriction enzyme analysis to confirm the right inserts before it was transformed into *E. coli* BW25113, *E. coli* BW25113 ΔbioC and *C. necator* H16.

3.3.8 Depletion of biotin for corroboration of *E. coli* BW25113 ΔbioC phenotype

For depletion of biotin to corroborate the *E. coli* BW25113 ΔbioC phenotype, a pre-culture of the cells was cultivated in 5 mL of 2x YT for 16 h at 37 ºC, for depletion of biotin present in the spent media and biotin within the cells, a fresh 5 mL 2x YT falcon tube was inoculated with the pre-culture at a 1:100 dilution and cultivated at 37 ºC, when cells reached an OD_{600} of 0.8, 1 mL of cells was transferred to a sterile 1.5 mL tube. Cells were centrifuged at 2,800 rpm for 2 min, and the supernatant was removed by pipetting. The pellet was washed thrice with 500 µL of M9 biotin-free medium. Then, the cell pellet was resuspended carefully in 150 µL of M9 biotin-free medium and inoculated at a 1:100 dilution in a fresh falcon tube with 5 mL of 1) M9 biotin-free medium or 2) M9 supplemented with biotin and grown for 16 h at 37ºC (first generation of cultivation). Then, the first generation was used to inoculate the second generation at a 1:1000 dilution in a fresh falcon tube with 5 mL of 1) M9 biotin-free medium or 2) M9 supplemented with biotin; the second generation was grown for up to 21 h.

3.3.9 Protein expression

For protein expression, freshly transformed cells with the desired plasmid were prepared, and a single colony was picked and grown overnight (*E. coli* strains for 16 h, and *C. necator* H16 for 40 h) in 5 mL of rich medium (*E. coli* strains in 2x YT, and *C. necator* H16 in NB) supplemented with antibiotics, then a fresh falcon tube with 5 mL of new rich medium supplemented with chloramphenicol was inoculated with the pre-culture at a 1:200 dilution and cultivated at 37 ºC (*E. coli* strains) or 30 ºC (*C. necator* H16); samples were induced at the early log phase when cells reached an OD_{600} of ~0.6 with 0.1 % or 0.5 % (w/v) L-arabinose. Cells were grown for 5 h, or overnight before 1 mL of cells was harvested by centrifugation at maximum speed (17,000 x g) and 4 ºC for 15 min, the supernatant was removed by pipetting, and the cell pellet was stored at -20 ºC for further use.

3.3.10 Preparation of protein gel

For analysis of protein expression, a 15 % SDS-PAGE gel was prepared. In order to prepare the 15 % SDS-PAGE gel, two different steps are required, the first one is the preparation of the resolving gel (lower part), and the second step is the preparation of the stacking gel (upper part). The composition of resolving gel was 1.20 mL of DDI H_2O, 2.5 mL of 30 % Acrylamide/Bis, 1.25 mL of 1.5 M Tris-HCl (pH 8.8), 0.05 mL of 10 % (w/v) SDS, 2.5 µL of TEMED, and 25 µL of APS. The composition of the stacking gel was 2.05 mL of DDI H_2O, 1.65 mL of 30 % Acrylamide/Bis, 1.25 mL of 0.5 M Tris-HCl (pH 6.8), 0.05 mL of 10 % (w/v) SDS, 5 µL of TEMED, and 25 µL of APS. After the addition of the polymerisation initiators, the gel was allowed to solidify and then 5 µL of the PageRuler Broad Range Unstained Protein Ladder, and 10 µL of the samples were loaded in each of the wells. The electrophoresis was run at a constant voltage at 200 V for 70 min. Protein gel was stained with coomassie brilliant blue staining dye for 20 min, then rinsed three times with water before destaining the gel with a destaining solution (500 mL of methanol to a final concentration of 50 % (v/v), 100 mL Acetic acid to a final concentration of 10 % (v/v), and 400 mL DDI water to a final concentration of 40 % (v/v)). The gel image was captured with a gel documentation system.

3.3.10.1 Analysis of protein expression for total protein content

The gel for total protein content analysis was prepared as described in **Section 3.3.10**. The cell pellet (after protein expression) from a 1 mL culture was resuspended in 100 µL of 50 mM phosphate buffer (470 µL of 1 M KH_2PO_4, and 2.03 mL of 1 M K_2HPO_4, top-up to 50 mL at pH 7.5) by vortexing, then 100 µL of SDS reducing buffer (2x SDS reducing buffer supplemented with β-mercaptoethanol; 2x SDS buffer was composed of 5.25 mL of DDI H_2O, 6.25 mL of 0.5 M Tris-HCl pH 6.8, 25 mL of 50 % (v/v) Glycerol, 10 mL of 10 % (w/v)

SDS, and 1 mL of 0.5 % (w/v) Bromophenol blue, where 50 µL β-mercaptoethanol was added to 950 µL of 2x SDS reducing buffer prior to use) was added before the cells were incubated at 94 ºC using a block heater for 10 min, cells were then centrifuged at maximum speed (17,000 x g) for 5 min before 10 µL sample was loaded into the SDS-PAGE gel.

3.3.10.2 Analysis of protein expression for the soluble and insoluble fraction of proteins

The gel for the soluble and insoluble fraction of protein analysis was prepared as described in **Section 3.3.10**. The cell pellet (after protein expression) from a 1 mL culture was resuspended in 100 µL of lysomix (1,900 µL of 50 mM phosphate buffer, and 100 µL of 10 mg/mL lysozyme (lysozyme was dissolved in 50 mM phosphate buffer) by vortexing, then, cells were centrifuged at maximum speed (17,000 x g) for 10 min. Afterwards, the supernatant (soluble fraction) was removed from the pellet (insoluble fraction) by pipetting and transferred to a new microcentrifuge tube. To 1 part of the soluble fraction sample, 1 part of SDS reducing buffer (2x SDS reducing buffer supplemented with β-mercaptoethanol; 2x SDS buffer was composed of 5.25 mL of DDI H_2O, 6.25 mL of 0.5 M Tris-HCl pH 6.8, 25 mL of 50 % (v/v) Glycerol, 10 mL of 10 % (w/v) SDS, and 1 mL of 0.5 % (w/v) Bromophenol blue, where 50 µL β-mercaptoethanol were added to 950 µL of 2x SDS reducing buffer prior to use) was added and mixed by vortexing before the cells were incubated at 94 ºC using a block heater for 5 min, cells were then centrifuged at maximum speed (17,000 x g) for 2 min before 10 µL sample was loaded into the SDS-PAGE gel. For the insoluble fraction, the cell pellet was resuspended in 100 µL SDS reducing buffer (1x SDS reducing buffer supplemented with β-mercaptoethanol; 1x SDS reducing buffer was composed of 1 mL 2x SDS reducing buffer, and 1 mL of 50 mM Phosphate buffer) by vortexing and boiled at 94 ºC for 5 min, then, cells were centrifuged at maximum speed for 2 min, before 10 µL sample was loaded into the SDS-PAGE gel.

3.4 Results and discussion

3.4.1 Construction of inducible plasmid systems for biotin related genes expression

The plasmid pBADc.rbs-bioC was constructed as a positive control. All plasmids described in this **Chapter 3** used pBADc.rbs-RFP plasmid from the synthetic biology toolbox described in **Chapter 2** as plasmid backbone. The pBADc.rbs-RFP is illustrated in **Figure 3.1**, where the digested fragment of pBADc.rbs-RFP harboured the pBBR1 Rep, pBBR1 oriV origin of replication, *mob* gene, chloramphenicol resistance (Cam^r) cassette (containing P_{cat} and *Cam*^{r}), L-arabinose-inducible system (containing *araC* and P_{BAD}), and *E. coli* consensus RBS; the *rfp* gene was replaced with either *bioC*, *bioDV1-4* extended biotin operon or *bioDD1-5* extended biotin operon. The synthetic extended biotin operons of DV and DD were codon optimised by GenScript for *E. coli* BW25113 to improve the expression of recombinant proteins.

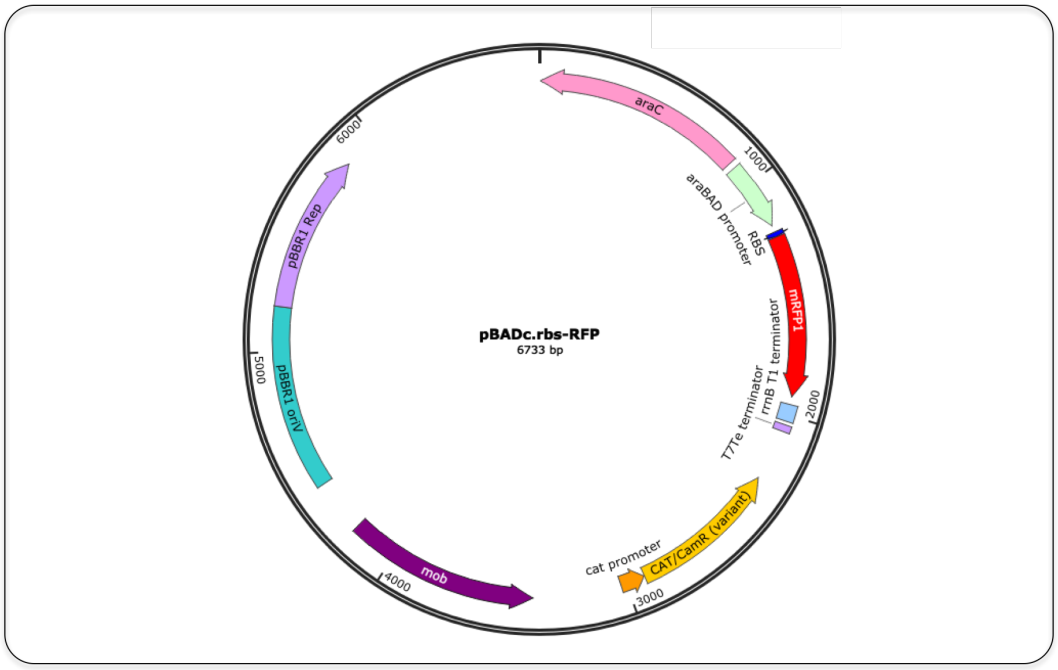


Figure 3.1 Illustration of pBADc.rbs-RFP map (6.7 kb). The plasmid encodes for pBBR1 Rep, pBBR1 oriV origin of replication, mobilisation sequence mob, CAT promoter, chloramphenicol resistance (Camr), araC and P_{BAD} promoter, and rfp gene.

3.4.1.1 Construction of plasmid pBADc.rbs-bioC

The construction of the arabinose-inducible pBADc.rbs-bioC plasmid was successfully achieved by replacing the *rfp* gene from pBADc.rbs-RFP with the *bioC* gene from *E. coli* BW25113. The pBADc.rbs-bioC plasmid was constructed to be used as a positive control for further experiments.

The PCR amplified *bioC* from *E. coli* BW25113 genomic DNA was analysed to confirm the size of the insert, which was a 0.8 kb DNA band as expected (**Figure 3.2A**). Then, once the construction of the pBADc.rbs-bioC plasmid was successfully ligated and transformed into *E. coli* DH5α, a PCR amplification of the insert within the newly constructed plasmid was performed, as this short insert (0.8 kb) was not visible in the gel after restriction enzyme analysis (ReA) as observed previously with other DNA bands of a similar size after ReA; also a ReA was performed, where DNA bands of 4.5 kb and 2.3 kb confirmed the right sizes expected for the digested pBADc.rbs-bioC plasmid (**Figure 3.2B**). After *E. coli* DH5α transformation, the constructed plasmid was isolated from *E. coli* DH5α and used to transform *E. coli* BW25113, *E. coli* BW25113 ΔbioC and *C. necator* H16 cells.

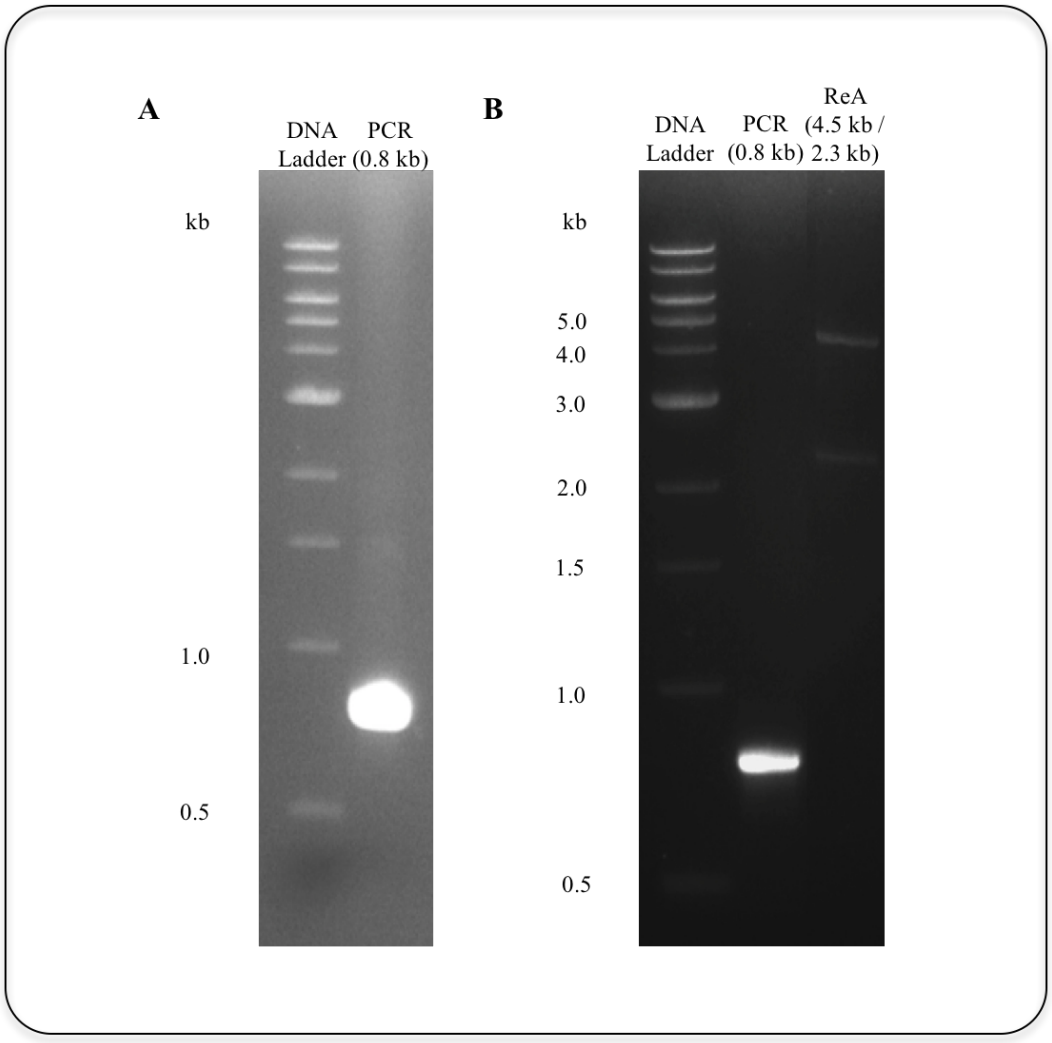


Figure 3.2 Analysis of the construction of pBADc.rbs-bioC. ***A)*** *Analysis of the PCR amplified insert from genomic DNA of E. coli BW25113,* ***B)*** *PCR of the amplified insert from E. coli BW25113 and restriction enzyme analysis (ReA) with PstI and XhoI of the constructed pBADc.rbs-bioC. Abbreviations: ReA, Restriction enzyme analysis.*

3.4.1.2 Construction of plasmid pBADc.rbs-bioDV1-4 and pBADc.rbs-bioDD1-5

The construction of the arabinose-inducible pBADc.rbs-bioDV1-4 and pBADc.rbs-bioDD1-5 plasmids was successfully achieved by replacing the *rfp* gene from pBADc.rbs-RFP with the synthetic operon *bioDV1-4* from pUC57-bioDV1-4 or the synthetic operon *bioDD1-5* from pUC57-bioDD1-5.

The synthetic operons were digested by restriction enzyme digestion (ReD) with *NdeI* and *XhoI* from pUC57-bioDV1-4 or pUC57-bioDD1-5 plasmids, which

yielded a DNA band of 3.4 kb corresponding to the *bioDV1-4* operon or a 3.6 kb DNA band for the *bioDD1-5* operon, and a DNA band of 2.5 kb corresponding to the pUC57 vector (**Figure 3.3A**). Then, once the construction of the pBADc.rbs-bioDV1-4 and pBADc.rbs-bioDD1-5 plasmids was successfully ligated and transformed into *E. coli* DH5α, a restriction enzyme analysis (ReA) was performed, where DNA bands of 3.4 kb confirmed the right size of the *bioDV1-4* operon or a 3.6 kb DNA band for the *bioDD1-5* operon, and a 6 kb DNA band expected for the vector (**Figure 3.3B**). After *E. coli* DH5α transformation, the constructed plasmids were isolated from *E. coli* DH5α and used to transform *E. coli* BW25113, *E. coli* BW25113 ΔbioC and *C. necator* H16 cells.

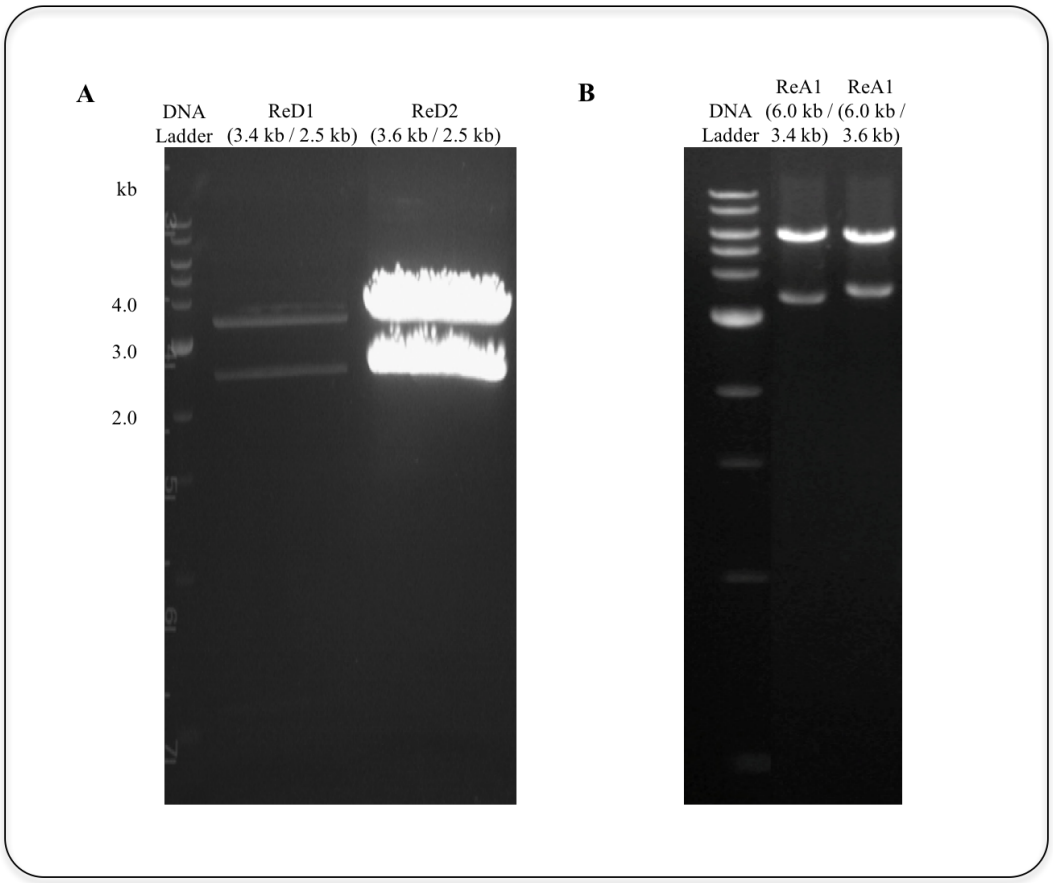


Figure 3.3 Analysis of the construction of pBADc.rbs-bioDV1-4 and pBADc.rbs-bioDD1-5. ***A)*** *Restriction enzyme digestion (ReD) with NdeI and XhoI of inserts bioDV1-4 from pUC57-bioDV1-4 (lane labelled as ReD1) and bioDD1-5 from pUC57-bioDD1-5 (lane labelled as ReD2).* ***B)*** *Restriction enzyme analysis (ReA) with NdeI and XhoI of the constructed pBADc.rbs-bioDV1-4 (lane labelled as ReA1) and pBADc.rbs-bioDD1-5 (lane labelled as ReA2). Abbreviations: ReD, Restriction enzyme digestion; ReA, Restriction enzyme analysis.*

3.4.2 Corroboration of *E. coli* BW25113 ΔbioC phenotype

E. coli BW25113 ΔbioC is a mutant deficient in *bioC* synthesis, the *bioC* gene is one of the essential genes in charge of the synthesis of pimelic acid; hence, this mutant strain is not capable of growing in biotin-free medium due to pimelate is a precursor for biotin synthesis. Critical metabolic enzymes of central metabolic pathways such as gluconeogenesis and fatty acid synthesis that mediate the transport of CO_2 cannot be active if biotin is not synthesised. In order to corroborate that the *bioC* gene had been knocked-out from *E. coli* BW25113 ΔbioC mutant strain, the mutant was subjected to grow in a biotin-free medium. This experiment was useful to confirm how many generations of cultivations were required to deplete biotin entailed from previous pre-cultures in rich media before cells could be able to grow in a biotin-free medium.

Thus, the first experiment performed in this set of experiments was to corroborate the phenotype of *E. coli* BW25113 ΔbioC, as this confirmation was a critical step in understanding the function of the *bioDV1-4* and *bioDD1-5* biotin operons of *Desulfovibrio* spp in further experiments. Once the phenotype of *E. coli* BW25113 ΔbioC was confirmed, it could be furtherly determined that if there were cell growth in *E. coli* BW25113 ΔbioC recombinant strains expressing *bioDV1-4* and *bioDD1-5* in biotin-free media, it would potentially be due the expression of the extended biotin operons was related to pimelate synthesis and not due to *E. coli* BW25113 ΔbioC could grow in biotin-free media.

In order to corroborate the phenotype of *E. coli* BW25113 ΔbioC, an overnight culture of the strain was first grown in 2x YT medium, then, an aliquot of the pre-culture was used to grow cells in fresh 2x YT medium to an OD_{600} of 0.8, once cells reached the specified OD_{600}, the spent media was removed before cells were washed in M9 biotin-free medium, this was the first step designed to remove any biotin left in the spent 2x YT medium. Then, the cell pellet was inoculated in two different media: 1) M9 biotin-free and 2) M9 supplemented with biotin, the second medium served as a control, where cell growth was always expected through all generations of cultivation. The two different media previously

mentioned correspond to the first generation of cultivation, where the pre-culture came from a rich medium (2x YT), therefore, growth in the M9 biotin-free medium was expected in the first generation of cultivation, as cells would carry some minimal but essential amounts of biotin within the cells to survive, and it is well known that in bacteria, biotin is needed only in trace quantities, specifically in *E. coli*, where only a few hundred molecules per cell are required for growth (nanomolar concentrations). The first generation was cultivated for 16 h, where effectively, there was cell growth of *E. coli* BW25113 ΔbioC in both conditions (biotin-free medium and medium supplemented with biotin); an aliquot of the first generation was then transferred to the second generation of cultivation.

In the second generation of cultivation, only the sample of *E. coli* BW25113 ΔbioC in M9 supplemented with biotin showed growth after 21 h of incubation (**Figure 3.4**). This experiment confirmed 1) the expected phenotype of *E. coli* BW25113 ΔbioC where there should be no growth when cells are cultivated in biotin-free medium, and 2) the number of generations of cultivation required to remove the trace quantities of biotin entailed from pre-cultures and biotin that could be generated within the cells.

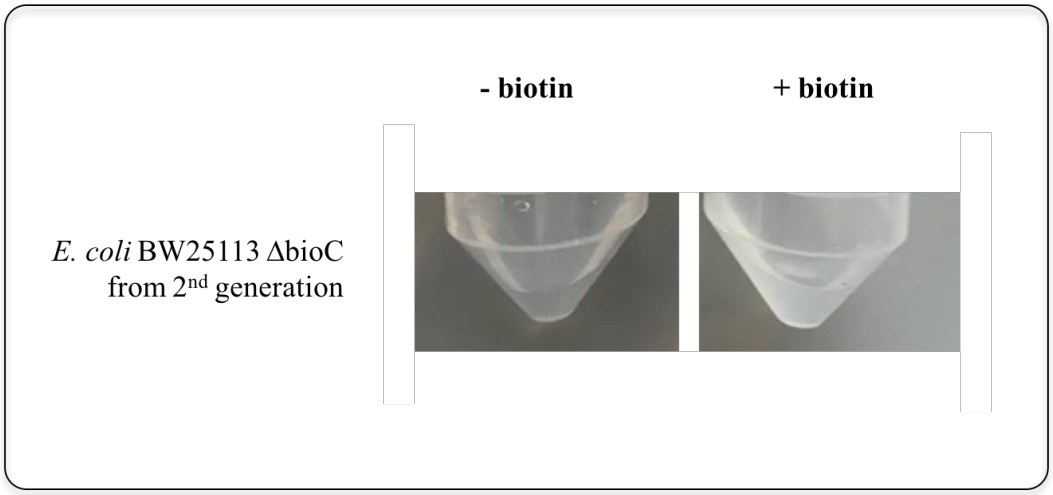


Figure 3.4 Growth of E. coli BW25113 ΔbioC cells from the second generation of cultivation in M9 biotin-free media (- biotin) and M9 supplemented with biotin (+ biotin).

3.4.3 Analysis of protein expression of *bioC*, *bioDV1-4* and *bioDD1-5*

Analysis of protein expression after induction of *bioC*, *bioDV1-4* and *bioDD1-5* operons in *E. coli* BW25113 was carried out in order to corroborate the expression of all the genes of the extended biotin operons before analysing the function of the extended biotin operons. Recombinant strains of *E. coli* BW25113 harbouring *bioC*, *bioDV1-4* and *bioDD1-5* were grown in 5 mL of 2x YT medium supplemented with chloramphenicol, where 0.1 % or 0.5 % (w/v) L-arabinose was used to induce the P_{BAD} promoter when cells reached early log phase to an OD_{600} of ~0.6. Different concentrations of L-arabinose in separate cultures were tested, this was done in order to obtain the highest protein expression yield at an inducer concentration that would not compromise cell growth.

After induction with L-arabinose, an aliquot of the samples was withdrawn from the cultures after 5 h and after 20 h of induction. Protein analysis samples were taken after 5 hours of expression as a longer expression time could sometimes potentially lead to protein degradation, hence compromising protein detection via SDS-PAGE. Protein analysis samples were also taken after 20 h of expression on the premise that longer expression time might enable better detection assuming the proteins are not degraded. Non-induced control samples were always included. The optical density at 600 nm was measured for all samples after the induction, as this could help to check whether the induction influenced or not in cell growth and thus, confirm if the induction was successfully achieved (**Table 3.3**).

Table 3.3 OD_{600} of samples of *E. coli* BW25113 expressing *bioC*, *bioDV1-4* and *bioDD1-5* after induction at the early log phase with 0.1 % and 0.5 % (w/v) L-arabinose in 2x YT medium.

Genes and operons	**OD_{600}**		
BioC	**0 h**	**5 h**	**20 h**
Non-induced	0.298	6.680	11.300
Induced 0.5 % (w/v) L-arab	0.254	**2.680**	**2.280**
Induced 1.0 % (w/v) L-arab	0.234	2.710	2.820
BioDV1-4			
Non-induced	0.224	7.070	13.360
Induced 0.5 % (w/v) L-arab	0.200	**4.750**	**4.750**
Induced 1.0 % (w/v) L-arab	0.212	8.620	10.160
BioDD1-5			
Non-induced	0.267	5.720	12.340
Induced 0.5 % (w/v) L-arab	0.289	**3.810**	**4.410**

Induced 1.0 % (w/v) L-arab	0.256	6.150	9.490

Abbreviations: L-arab, L-arabinose.

As observed in **Table 3.3,** the induction of the P_{BAD} promoter for the expression of the BioC, BioDV1-4 and BioDD1-5 resulted in significant cell growth decrease (≈ 77 % cell growth decrease when BioC was expressed at both different concentrations of the inducer, ≈ 64 % cell growth decrease when BioDV1-4 or BioDD1-5 were induced with 0.1 % (w/v) L-arabinose, and ≈ 24 % cell growth decrease when BioDV1-4 or BioDD1-5 were induced by 0.5 % (w/v) L-arabinose after 20 h of cultivation). This cell growth decrease might be attributed to metabolic burden ensuing from the diversion of energy resources and vital cellular building blocks towards the expression of the BioC and BioDV1-4 proteins. Expression of heterologous genes or pathways is generally known to lead to a hijacking of resources vital for cell viability, which have been known to unleash energetic inefficiency on the cell (Wu *et al*., 2006).

Overall, the observed cell growth decrease was more marked with the expression of BioC than with expression of BioDV1-4 and BioDV1-5, this could be due to the expression of BioC not only causes metabolic burden but also it is toxic to the cells. The toxicity of BioC to the cells might be due to as described by Lin *et al*. (2012), the *bioC* gene oversees the methylation of the *ω*-carboxyl group of malonyl-CoA (or malonyl-ACP), which mimics the methyl ends of regular fatty acyl chains, where this substrate is recognised by the fatty acid enzymes, entering thus to elongation in the fatty acid synthesis, which gives pimeloyl-ACP methyl ester. Henceforth, the overexpression of BioC could have caused elevated levels of malonyl-ACP methylated substrates, which could lead to the blockage of the fatty acid synthesis due to the lack of malonyl-ACP, thereby causing the decrease in cell growth.

In regard to the different concentrations of L-arabinose tested, it could be appreciated that for the extended biotin operons when a higher concentration of L-arabinose (0.5 % (w/v) L-arabinose) was used, there was less cell growth decrease compared to when a lower concentration of L-arabinose was used (0.1 % (w/v) L-

arabinose), meaning possibly that the genes were more highly expressed when 0.1 % (w/v) L-arabinose was used to induce than when 0.5 % (w/v) L-arabinose was used to induce protein expression. For further experiments, only 0.1 % (w/v) L-arabinose was used for induction to ensure the protein expression of the operons. The samples were withdrawn after 5 h and 20 h of cultivation after induction and analysed in an SDS-PAGE (**Figure 3.5**).

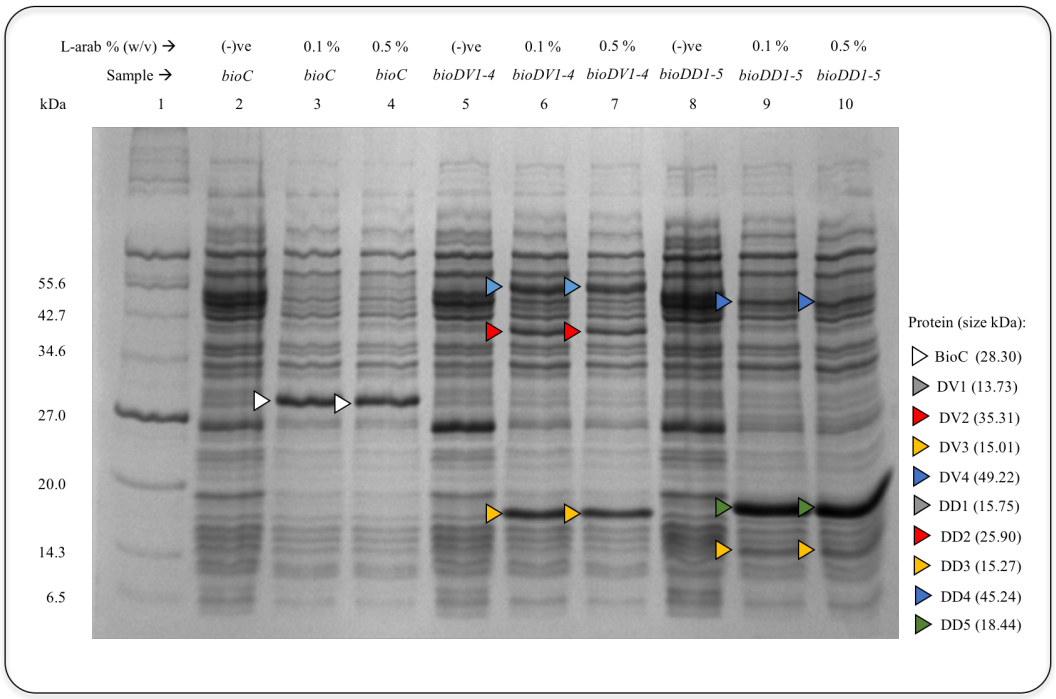


Figure 3.5 15 % SDS-PAGE showing the total protein content of bioC, bioDV1-4 and bioDD1-5 expression in E. coli BW25113 in 2x YT medium based on the induction of the P_{BAD} promoter with 0.1 % and 0.5 % (w/v) L-arabinose. Samples were withdrawn from the culture after 5 h of induction. The following plasmids were carried for protein expression: Lane 1: PageRuler Broad Range Unstained Protein Ladder, Lane 2: pBADc.rbs-bioC non-induced, Lane 3: pBADc.rbs-bioC induced with 0.1 % L-arabinose, Lane 4: pBADc.rbs-bioC induced with 0.5 % L-arabinose, Lane 5: pBADc.rbs-bioDV1-4 non-induced, Lane 6: pBADc.rbs-bioDV1-4 induced with 0.1 % L-arabinose, Lane 7: pBADc.rbs-bioDV1-4 induced with 0.5 % L-arabinose; Lane 8: pBADc.rbs-bioDD1-5 non-induced, Lane 9: pBADc.rbs-bioDD1-5 induced with 0.1 % L-arabinose, Lane 10: pBADc.rbs-bioDD1-5 induced with 0.5 % L-arabinose. The arrows show the protein bands of the different plasmids, where only the arrows of identified proteins are shown in the illustration. Abbreviations: (-)ve, negative control; L-arab, L-arabinose.

From the SDS-PAGE shown in **Figure 3.5**, which corresponds to the samples withdrawn after 5 h of induction, it could be appreciated that only three out of the four proteins from the *bioDV1-4* extended biotin operon could be identified in the

gel; where the target protein corresponding to the expression of the *DV3* gene seemed to have migrated slightly slower than expected, which appears to have a higher molecular weight of about 1 to 2 kDa higher, although this is not unusual as sometimes proteins migrate differently in the gels, due to in some occasions the abundance of a specific amino acid can affect protein migration, such as a high content of acid or basic amino acids (Guan *et al*., 2015).

On the other hand, the expression of the *bioDD1-5* biotin operon showed similar results compared to the *bioDV1-4*, where only three out of the five proteins of the extended biotin operon could be identified in the gel. Both extended biotin operons did not show their first protein of the operon in the gel corresponding to the *DV1* and *DD1* genes; therefore, the SDS-PAGE gel of the samples withdrawn after 20 h of induction (**Figure 3.6**) was performed as well to check if the proteins corresponding to *DV1* and *DD1* could be identified in the gel, as well as the *DD2* from the *bioDD1-5* extended biotin operon.

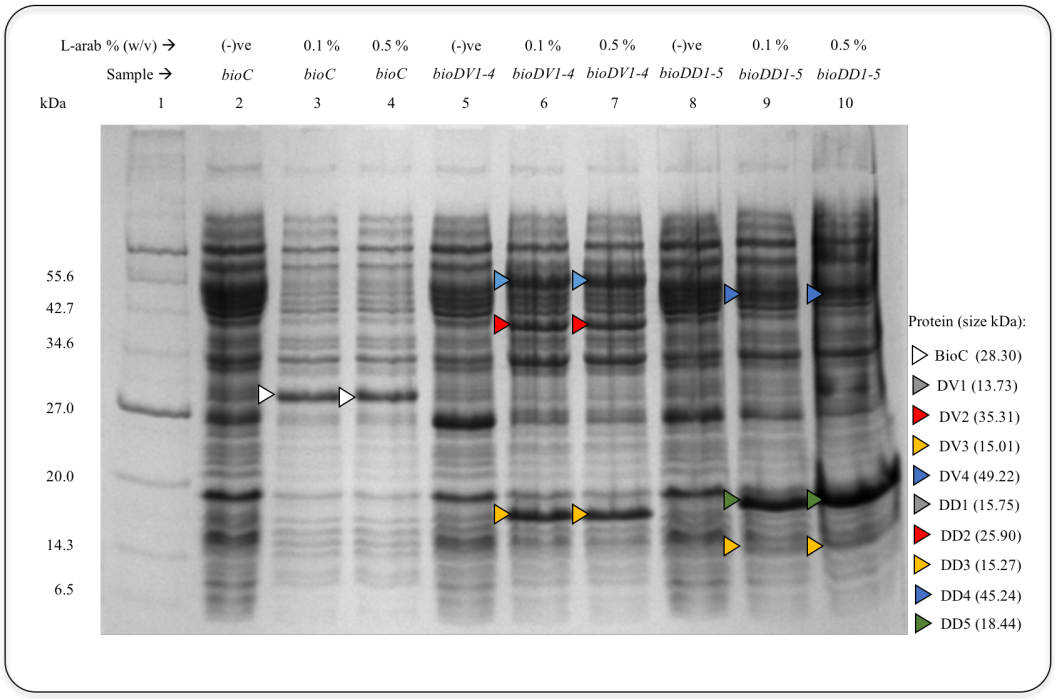


Figure 3.6 15 % SDS-PAGE showing the total protein content of bioC, bioDV1-4 and bioDD1-5 expression in E. coli BW25113 in 2x YT medium based on the induction of the P_{BAD} promoter with 0.1 % and 0.5 % (w/v) L-arabinose. Samples were withdrawn from the culture after 20 h of induction. The following plasmids were carried for protein expression: Lane 1: PageRuler Broad Range Unstained Protein Ladder, Lane 2: pBADc.rbs-bioC non-induced, Lane 3: pBADc.rbs-bioC induced with 0.1 % L-arabinose, Lane 4: pBADc.rbs-bioC induced with 0.5 % L-arabinose, Lane 5: pBADc.rbs-bioDV1-4 non-induced, Lane 6: pBADc.rbs-bioDV1-4 induced with 0.1 % L-arabinose, Lane 7: pBADc.rbs-bioDV1-4 induced with 0.5 % L-arabinose; Lane 8: pBADc.rbs-bioDD1-5 non-induced, Lane 9: pBADc.rbs-bioDD1-5 induced with 0.1 % L-arabinose, Lane 10: pBADc.rbs-bioDD1-5 induced with 0.5 % L-arabinose. The arrows show the protein bands of the different plasmids, where only the arrows of identified proteins are shown in the illustration. Abbreviations: (-)ve, negative control; L-arab, L-arabinose.

From the analysis of proteins of samples withdrawn after 20 h of induction (**Figure 3.6)**, the same results of the previous gel of samples withdrawn after 5 h (**Figure 3.5)** were observed, whereby none of the first genes from the extended biotin operons could be appreciated in the gel (*DD1*, 13.73 kDa and *DV1,* 15.75 kDa) and neither the protein band for the second gene of the *bioDD1-5* biotin operon (*DD2*, 25.90 kDa). The fact that these proteins were not overexpressed migh indicate that these proteins could be toxic to the cells, and therefore, they are not highly expressed.

Due to not all the target proteins were detected in the SDS-PAGE analysis of total protein content after 5 h or 20 h of induction, a repeat of the experiment, alongside with a growth curve to observe clearer the impact in cell growth after induction of the P_{BAD} promoter was performed again for the expression of *bioC*, *bioDV1-4*, and *bioDD1-5*, but this time in *E. coli* BW25113 ΔbioC, and in *C. necator* H16, as ultimately, the synthetic biology toolbox with the inducible system was tailored for optimised protein expression in *C. necator* H16. In *C. necator* H16, only one of the extended biotin operons (BioDD1-5) was transformed into the cell to analyse protein expression.

In this set of experiments, the pBADc.rbs-RFP plasmid, used as a plasmid backbone for the construction of the plasmids described in this chapter, was also included in the samples as an external control. All samples were induced at the early log phase to an OD_{600} of 0.5 with 0.1 % (w/v) L-arabinose, and samples for protein analysis were withdrawn after 5 h of induction. The growth curve of *E. coli* BW25113 ΔbioC and its recombinant strains is shown in **Figure 3.7A** and **Figure 3.7B**, and the growth curve of *C. necator* H16 and its recombinant strains is shown in **Figure 3.7C** and **Figure 3.7D**.

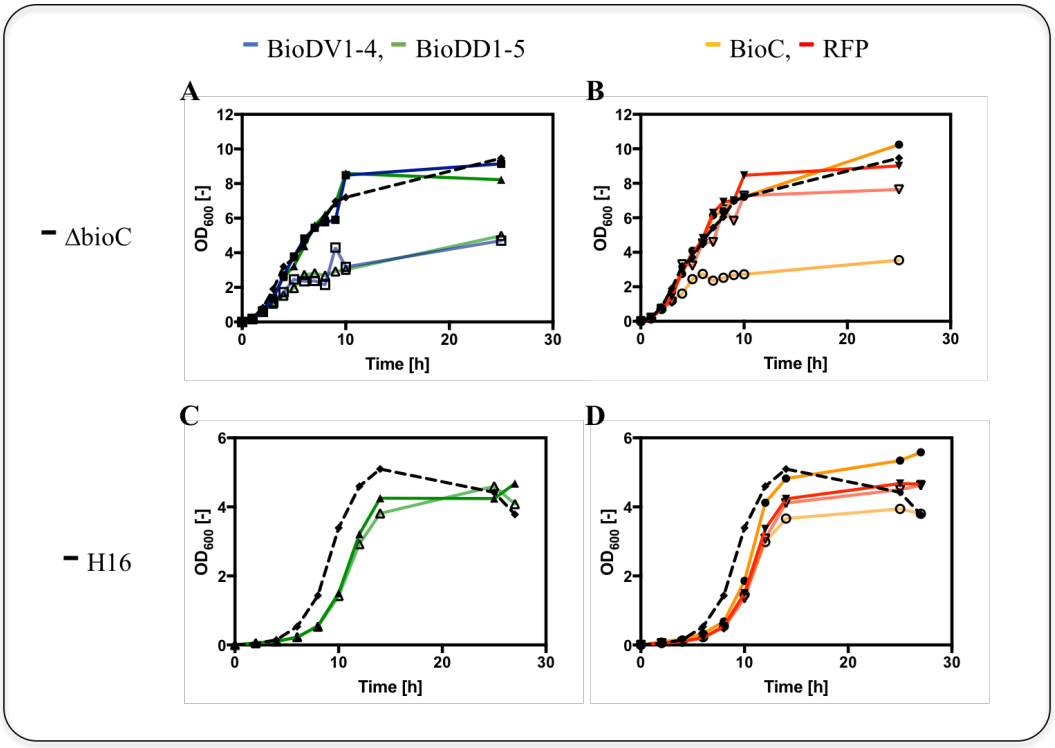


Figure 3.7 Cell growth study of E. coli BW25113 ΔbioC (shown as ΔbioC) and C. necator H16 (shown as H16) in nutrient-rich media. ***A)*** *Cell growth curve of non-induced samples of E. coli BW25113 ΔbioC (no plasmid) (diamonds, black dotted line), ΔbioC_bioDV1-4 (square, blue line), and ΔbioC_bioDD1-5 (triangle, green line); and its induced samples with 0.1 % (w/v) L-arabinose ΔbioC_bioDV1-4 (open square, blue line), and ΔbioC_bioDD1-5 (open triangle, green line) in 2x YT medium.* ***B)*** *Cell growth curve of non-induced samples of E. coli BW25113 ΔbioC (no plasmid) (diamonds, black dotted line), ΔbioC_bioC (circle, orange line), and ΔbioC_rfp (inverted triangle, red line); and its induced samples with 0.1 % (w/v) L-arabinose ΔbioC_bioC (open circle, orange line), and ΔbioC_rfp (open inverted triangle, red line) in 2x YT medium.* ***C)*** *Cell growth curve of non-induced samples of C. necator H16 (no plasmid) (diamonds, black dotted line), and H16_bioDD1-5 (triangle, green line); and its induced samples with 0.1 % (w/v) L-arabinose H16_bioDD1-5 (open triangle, green line) in NB medium.* ***D)*** *Cell growth curve of non-induced samples of C. necator H16 (no plasmid) (diamonds, black dotted line), H16_bioC (circle, orange line), and H16_rfp (inverted triangle, red line); and its induced samples with 0.1 % (w/v) L-arabinose H16_bioC (open circle, orange line), and H16_rfp (open inverted triangle, red line) in NB medium. The graphs show the results of single determinations per sample.*

From the growth curve shown in **Figure 3.7A** and **Figure 3.7B** corresponding to the cell growth study of *E. coli* BW25113 ΔbioC and its recombinant strains, it was observed that the induced samples depicted in **Figure 3.7A** of *bioDV1-4* (open square, blue line) and *bioDD1-5* (open triangle, green line) were expressed after the induction with 0.1 % (w/v) L-arabinose, as the decrease in cell growth was evident with a 49 % cell growth decrease when BioDV1-4 was expressed, and 39 % cell growth decrease when BioDD1-5 was expressed after 25 h of

cultivation. On the other hand, in **Figure 3.7B** the decrease in cell growth when BioC was expressed was again significantly higher compared to its non-induced sample, showing a 65 % cell growth decrease when BioC was expressed after 25 h of cultivation. The impact in cell growth decrease for the expression of these proteins was similar to what was previously observed with the impact in cell growth decrease of the *E. coli* BW25113 recombinant strains. The expression of RFP did not impact in cell growth significantly (only a 15 % cell growth decrease was observed when RFP was expressed by 0.1 % (w/v) L-arabinose after 25 h of cultivation) compared to the impact in cell growth decrease when BioC, BioDV1-4 and BioDD1-5 were expressed, although a slight decrease in cell growth was noticeable; this negligible impact in cell growth decrease when RFP was expressed might be due to the expression of RFP does not affect the natural metabolism of the host cell, since this protein might not interfere with any of its native pathways, as it is not the case for BioC, which does interfere with the natural metabolism of the cell.

On the other hand, the growth curves shown in **Figure 3.7C** and **Figure 3.7D** demonstrate the impact in cell growth of *C. necator* H16 and its recombinant strains. In **Figure 3.7C**, the impact in cell growth was not significant after the expression of BioDD1-5 (open triangle, green line), contrary to what was observed with protein expression of the extended biotin operons of *E. coli* BW25113 ΔbioC and *E. coli* BW25113 recombinant strains. In the case of the expression of BioC in **Figure 3.7D**, the same results than the ones observed for *E. coli* BW25113, and *E. coli* BW25113 ΔbioC were obtained, although the impact in cell growth decrease was lower (26 % cell growth decrease was observed when BioC expression was induced by 0.1 % (w/v) L-arabinose after 25 h of cultivation in H16_bioC). The cell growth decrease after BioC expression was expected as we know that the overexpression of BioC might block the fatty acid synthesis which can directly impact in cell growth. Also, as observed in *E. coli* BW25113 ΔbioC cell growth curve, the expression of RFP did not decrease cell growth of H16_rfp, which might be attributed to the fact that the expression of this protein does not interfere with the natural metabolism of the cell. The fact that the impact in cell growth decrease of H16_bioDD1-5 was lower compared to *E. coli*

recombinant strains harbouring the same extended biotin operon, might be due to the same concentration of L-arabinose was used to induce the samples of both microorganisms (*E. coli* and *C. necator* H16), where 0.1 % (w/v) L-arabinose might be an optimum L-arabinose concentration for protein expression in *E. coli*, but not for *C. necator* H16; recent publications mention that the generally the P_{BAD} promoter reaches expression maxima at 0.2 % (w/v) L-arabinose in *C. necator* H16 (Johnson *et al*., 2018).

An aliquot of all samples shown in **Figure 3.7** was withdrawn after 5 h of cultivation after induction and was analysed in an SDS-PAGE. Although for this analysis, the soluble and insoluble fractions of the proteins were analysed (**Figure 3.8**, and **Figure 3.9**).

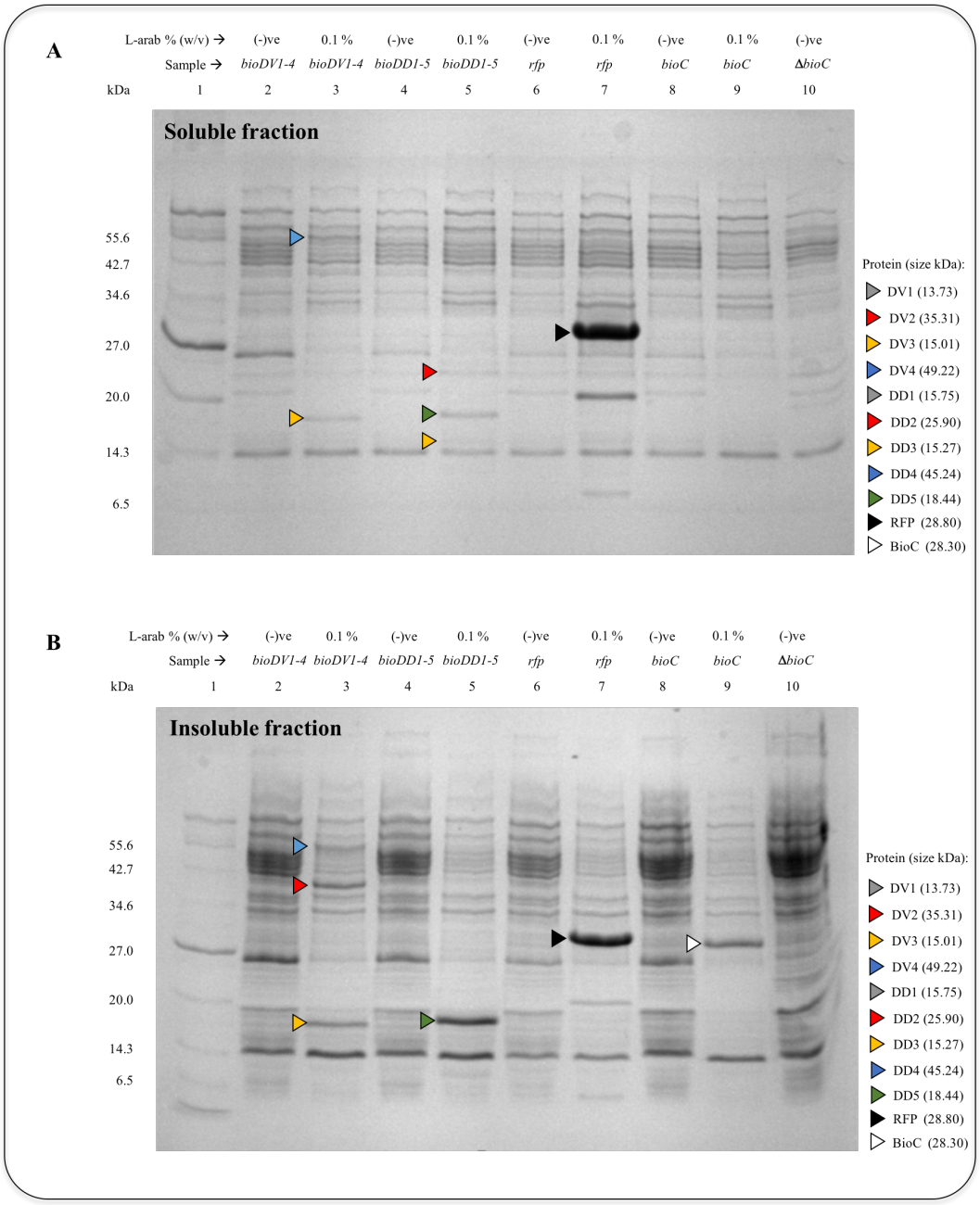


*Figure 3.8 15 % SDS-PAGE showing the soluble and insoluble fractions of proteins of bioDV1-4, bioDD1-5, rfp and bioC in E. coli BW25113 ΔbioC in 2x YT medium. Protein expression was based on the induction of the P_{BAD} promoter with 0.1 % (w/v) L-arabinose. Samples were withdrawn from the culture after 5 h of induction and treated to analyse the soluble (**A**) and insoluble (**B**) fractions of proteins. The following plasmids were carried for protein expression for both protein gels: Lanes 1: PageRuler Broad Range Unstained Protein Ladder, Lanes 2: pBADc.rbs-bioDV1-4 non-induced, Lanes 3: pBADc.rbs-bioDV1-4 induced with 0.1 % (w/v) L-arabinose, Lanes 4: pBADc.rbs-bioDD1-5 non-induced, Lanes 5: pBADc.rbs-bioDD1-5 induced with 0.1 % (w/v) L-arabinose, Lanes 6: pBADc.rbs-RFP non-induced, Lanes 7: pBADc.rbs-RFP induced with 0.1 % (w/v) L-arabinose; Lanes 8: pBADc.rbs-bioC non-induced, Lanes 9: pBADc.rbs-bioC induced with 0.1 % (w/v) L-arabinose, Lanes 10: E. coli BW25113 ΔbioC. The arrows show the protein bands of the different plasmids, where only the arrows of identified proteins are shown in the illustration. Abbreviations: (-)ve, negative control; L-arab, L-arabinose.*

The previous gel (**Figure 3.8**) corresponds to the soluble and insoluble fractions of proteins expressed in *E. coli* BW25113 ΔbioC, where a target protein that had not been observed in the SDS-PAGE analysis of total protein content of *E. coli* BW25113 samples (**Figure 3.5** and **Figure 3.6**) could be observed this time in the soluble fraction, which corresponded to the *DD2* gene (red arrow in Lane 5). Although once more, no target protein bands could be observed for the first genes of the extended biotin operons of either DD or DV and neither the DD4 protein band that had been observed in previous gels, while the RFP protein was highly expressed in both soluble and insoluble fractions.

On the other hand, the BioC protein band could be appreciated in the insoluble fraction gel, while BioC was not observed in the soluble fraction gel; this fact is understandable since usually, a high level of expression of recombinant proteins in bacteria leads to the formation of inclusion bodies, which are the insoluble fractions of proteins; inclusion bodies are highly aggregated proteins that are generally formed in the cytoplasm, and can be observed in the SDS-PAGE after lysing the cells. After the analysis of the soluble and insoluble fraction of proteins in *E. coli* BW25113 ΔbioC, the same SDS-PAGE was carried out for *C. necator* H16 samples.

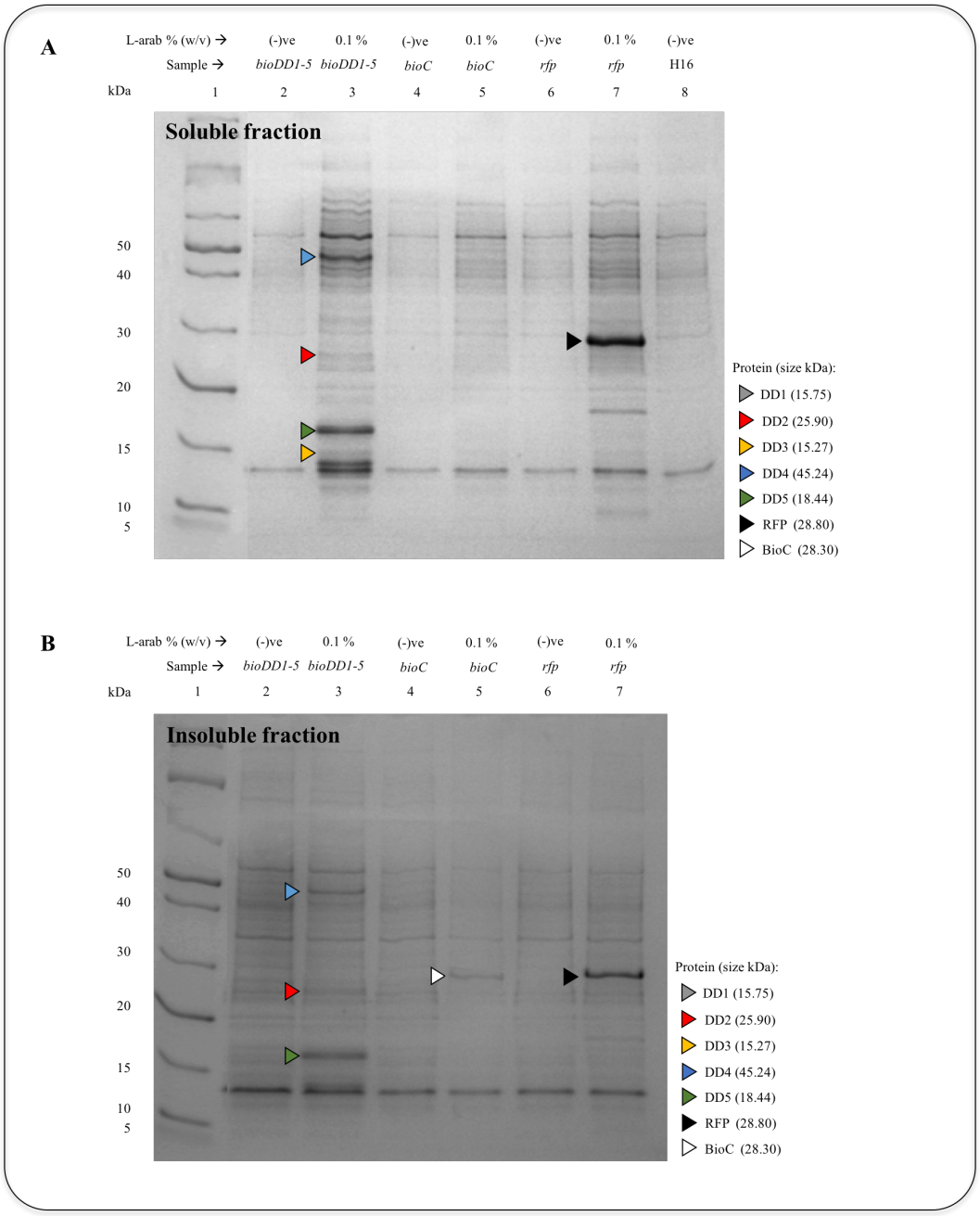


*Figure 3.9 15 % SDS-PAGE showing the soluble and insoluble fractions of proteins of bioDV1-4, bioDD1-5, bioC, and rfp expression in C. necator H16 in NB medium. Protein expression was based on the induction of the P_{BAD} promoter with 0.1 % (w/v) L-arabinose. Samples were withdrawn from the culture after 5 h of induction and treated to analyse the soluble (**A**) and insoluble (**B**) fractions of proteins. The following plasmids were carried for protein expression for both protein gels: Lanes 1: PageRuler Broad Range Unstained Protein Ladder, Lanes 2: pBADc.rbs-bioDD1-5 non-induced, Lanes 3: pBADc.rbs-bioDD1-5 induced with 0.1 % (w/v) L-arabinose, Lanes 4: pBADc.rbs-bioC non-induced, Lanes 5: pBADc.rbs-bioC induced with 0.1 % (w/v) L-arabinose; Lanes 6: pBADc.rbs-RFP non-induced, Lanes 7: pBADc.rbs-RFP induced with 0.1 % (w/v) L-arabinose, Lane 8: C. necator H16. The arrows of different colours show the protein target bands of the different plasmids, where only the arrows of identified proteins are shown in the illustration. Abbreviations: (-)ve, negative control; L-arab, L-arabinose.*

The soluble and insoluble fractions of the proteins of the extended biotin operon expressed in *C. necator* H16 (**Figure 3.9**) did not show the DD1 protein either as observed previously in *E. coli* recombinant strains. On the other side, the expression of the rest of the proteins of the BioDD1-5 operon could be observed in the SDS-PAGE analysis in both soluble and insoluble fractions. The RFP protein was highly expressed in both soluble and insoluble fractions (mainly in the soluble fraction as observed for *E. coli* BW25113 ΔbioC samples); and the BioC protein band could be appreciated in the insoluble fraction gel, as it occurred with the *E. coli* BW25113 ΔbioC.

The summary of the genes observed in the total protein content and soluble and insoluble fractions of the SDS-PAGE protein gels are shown in **Table 3.4**.

Table 3.4 Target protein bands observed in SDS-PAGE after 5 h and 20 h of induction of *bioDV1-4*, *bioDD1-5*, *bioC* and *rfp* in *E. coli* BW25113, *E. coli* BW25113 ΔbioC and *C. necator* H16.

Operon and genes	**5 h of induction** *E. coli* BW25113	**20 h of induction** *E. coli* BW25113	**5 h of induction** *E. coli* BW25113ΔbioC		**5 h of induction** *C. necator* H16	
bioDV1-4	**Total pt**	**Total pt**	**Soluble pt**	**Insoluble pt**	**Soluble pt**	**Insoluble pt**
DV1 (13.73 kDa)	-	-	-	-	NA	NA
DV2 (35.31 kDa)	+	+	-	+	NA	NA
DV3 (15.01 kDa)	+	+	+	+	NA	NA
DV4 (49.22 kDa)	+	+	+	+	NA	NA
bioDD1-5						
DD1 (15.75 kDa)	-	-	-	-	-	-
DD2 (25.90 kDa)	-	-	+	-	+	+
DD3 (15.27 kDa)	+	+	+	-	+	-
DD4 (45.24 kDa)	+	+	-	-	+	+
DD5 (18.44 kDa)	+	+	+	+	+	+
BioC						
BioC (28.30 kDa)	+	+	-	+	-	+
RFP						
RFP (28.80 kDa)	NA	NA	+	+	+	+

The signs + and – denote the existence or absence of the target protein bands in the SDS-PAGE. Abbreviations: Pt, protein; NA, not applicable.

As depicted in **Table 3.4**, not all the proteins of *bioDV1-4* and *bioDD1-5* extended biotin operons could be observed in the protein gels in any of the strains, although all proteins were observed throughout the different samples, unless for

DV1 and DD1. The lack of ability to visualise protein bands of the DV1 and DD1 genes is not conclusive in terms of determining the expression of the proteins. Alternative approaches could be performed to corroborate the expression of these proteins, which could include proteomic analysis by mass spectrometry (MS) (Washburn *et al*., 2001).

Proteomic analysis can be performed using a so called 'shotgun' bottom-up approach for protein identification, by analysis at the peptide level, where proteins are proteolytically digested, typically using trypsin. The resulting peptides are analysed by LC-MS/MS analysis, which couples peptide separation by reversed-phase high performance liquid chromatography (LC) with accurate mass measurement by tandem MS (MS/MS). Peptide masses are measured together with their fragments obtained by collision-induced dissociation. The resulting MS/MS spectra lead to peptide sequence information and thus protein identification by *in silico* matching against protein sequence database(s) (Perkins et al., 1999). The approach is applicable to proteins in-solution or post separation by SDS-PAGE. The advantage of using SDS-PAGE is that it provides a one-step sample clean up with high protein resolution. GeLC-MS/MS, uses entire gel lanes from SDS-PAGE (Lundby *et al*., 2011). In this procedure, entire gel lanes are excised as a series of gel slices, followed by in-gel digestion, typically using trypsin as the proteolytic enzyme of choice.

If after the above-mentioned analysis to corroborate the expression of DV1 and DD1 proteins the outcome shows that the proteins have not been expressed, it could be possible that the fact that DV1 and DD1 were not overexpressed in any of the different microorganisms tested indicate that these proteins are potentially toxic to the cells, and therefore, they are not highly expressed.

3.5 Conclusions

The plasmids constructed and described in this **Chapter 3** used the synthetic biology toolbox with the inducible system described in **Chapter 2**, which were

successfully transformed into *C. necator* H16 by electroporation with the optimised protocol method described in **Chapter 2**, although the expression of the cluster of genes of the extended biotin operons could not be confirmed, the inducible system demonstrated to be able to express other functional single genes apart from the reporter gene *rfp*. The *bioC* gene was successfully expressed in *C. necator* H16, and in *E. coli* BW25113 strains.

The protein expression after induction of the P_{BAD} promoter showed the target proteins of most of the genes of the extended biotin operons of *Desulfovibrio* spp. in the SDS-PAGE protein gels in the *E. coli* BW25113, *E. coli* BW25113 ΔbioC and *C. necator* H16 host cells, although there was no full evidence that all 4 and 5 genes of the extended biotin operons had been expressed simultaneously. If alternative analysis (*e.g.*, MS) of protein expression would not show either the successful expression of DV1 and DD1, the possible reason why the DV1 and DD1 proteins are not being expressed could be due to these genes might cause an imbalance in the cells due to toxicity, and hence these proteins can not be highly expressed.

In the SDS-PAGE protein gels, both *E. coli* BW25113 ΔbioC and *C. necator* H16 showed high expression of the positive controls RFP and BioC proteins, this indicated that the P_{BAD} promoter was successfully induced for protein expression. Due to the expression of the entire extended biotin operons was not proven in the experiments described in this **Chapter 3**, it could not be investigated if the extended biotin operons are committed to pimelate synthesis. Further studies should be performed to analyse the function of the extended biotin operons of *Desulfovibrio* spp., which might include studying different dose-responses of the L-arabinose inducer, as well as different points of L-arabinose induction; also, other inducible promoters could be integrated into the synthetic biology toolbox, which could allow a better expression of these cluster of genes.

CHAPTER 4

Directed evolution:

Development of a directed evolution *via* random mutagenesis tool for *Cupriavidus necator* H16 to understand its biotin biosynthesis pathway

Abstract

In **Chapter 4**, evolutionary engineering tools were optimised to engineer *C. necator* H16, the evolutionary engineering tool used was directed evolution *via* random mutagenesis, where an optimised protocol for random mutagenesis with the chemical mutagen EMS for *C. necator* H16 was optimised. The EMS random mutagenesis method provided a survival rate % of 50 to 85 % using different concentrations of EMS from 0.3 to 1.0 % (v/v) EMS. The method was used to try to understand the biotin biosynthesis pathway of *C. necator* H16, where TVA (a biotin analogue) was used as selection pressure to identify beneficial mutants. A total of 696 TVA-resistant mutants were isolated and screened, from which 8 mutants were further characterised. Biotin quantitation was performed for the mutants characterised in order to compare the biotin production with *C. necator* H16. The biotin studies showed no improvement in biotin production compared to *C. necator* H16. The development of the EMS random mutagenesis method for *C. necator* H16 is a potential engineering tool for the strain, as it can be used for engineering *C. necator* H16 for a wide variety of other bioindustry purposes.

4 Chapter 4. Directed evolution: Development of a directed evolution *via* random mutagenesis tool for *Cupriavidus necator* H16 to understand its biotin biosynthesis pathway.

4.1 Introduction

In previous **Chapters 2** and **3**, rational design engineering approaches in *C. necator* H16 were used for the expression of heterologous genes using a synthetic biology toolbox for protein expression in *C. necator* H16. In this **Chapter 4**, a different engineering strategy is described, a tool of evolutionary engineering – directed evolution *via* random mutagenesis- was studied in order to expand the engineering methods useful for chemical production in *C. necator* H16, as this powerful approach remains limited for *C. necator* H16.

As mentioned in **Section 1.3**, rational design engineering (*e.g.*, recombinant strain engineering, metabolic engineering) and evolutionary engineering (*e.g.*, directed evolution, adaptive evolution) are the most common strategies used for chemical production in microorganisms. As reviewed by Nannemann *et al*. (2011), the second strategy is often used when there is no full understanding of the pathways, proteins or genes of interest of the host cell. Evolutionary engineering is a reverse engineering method that generates random mutations and then analyses the impact of the mutations to link phenotype with genotype. Mutants generated by evolutionary engineering sometimes can be used directly as cell factories, and they have an advantage versus plasmid-borne recombinant strains in that these have permanent cellular phenotypes and are not dependent on the stability of the plasmid vector expressing the protein of interest. On the other hand, if mutants generated are not suitable to be used directly for chemical production, whole genome sequencing of the mutant of interest can be performed to understand the changes occurred in their genome, where genome sequences of the wild-type and the mutant strain are compared to find mutations in their DNA sequence. Once beneficial mutations are identified, these can be replicated in wild-type strains by recombinant strain engineering for chemical production.

4.1.1 Directed evolution *via* random mutagenesis tool for *C. necator* H16 to understand its biotin biosynthesis pathway

As reviewed in Kim *et al.* (2005), one of the main tools of evolutionary engineering is directed evolution; it is a powerful approach useful to understand the biological function of genes by producing mutants with altered phenotypes; in this post-genomic era, the use of reverse engineering to understand the function of genes has become widespread. Directed evolution is a term that encompasses a wide variety of methods that are used to improve or alter the functions of genes, enzymes, or pathways. Random mutagenesis is one of the main strategies used in directed evolution experiments to generate a library of mutants; where at the same time different random mutagenesis methods can be used such as epPCR, and chemical or physical mutagenesis. Chemical and physical mutagens can be used for modifying template DNA to foster mutations. Some of the chemical mutagens used for mutation of bacterial strains are ethyl methanesulfonate (EMS), nitrous acid, hydrazine, or formic acid, which generate chemical changes in nucleotide bases; also, physical mutagens such as UV irradiation can be used to generate whole-cell mutations (reviewed by Nannemann *et al.*, 2011). The rationale behind directed evolution *via* random mutagenesis is the incorporation of accelerated random mutagenesis that can create random mutations for instance, in bacterial genomes; the procedure goes through iterative cycles until the desired phenotype is obtained, and a selection pressure is used for the screening of the mutant strains to check for phenotype improvement.

Evolutionary engineering is often used when there is a lack of extensive knowledge on how genes or entire metabolic pathways work, thus, it represented a potential approach to be optimised for *C. necator* H16, which could be used to study one of its unknown metabolic pathways. One of the most intriguing metabolic pathways in many microorganisms to date is biotin biosynthesis, as even when the *bioFADB* biotin operon – which is in charge of the assembly of the fused heterocyclic rings of biotin- is a universal operon in many prokaryote systems, the origin of pimelic acid in bacteria lacking the *E. coli* biotin biosynthesis pathway, remains a mystery (Manandhar *et al*., 2017). This is the

case for *C. necator* H16, where, as mentioned in **Section 1.5.7.5**, the biotin operon in charge of the assembly of the biotin rings in *C. necator* H16 (*bioFADB* biotin operon) has been identified in its genome (Pohlmann *et al*. 2006), although the synthesis of pimelic acid in the strain remains unknown, where only a *bioC* gene has been predicted (Rodionov *et al.*, 2002; Pohlmann *et al.*, 2006), albeit no studies have been reported to date to confirm the function of that or other genes committed to pimelic acid in *C. necator* H16. The understanding of the biotin biosynthesis pathway in *C. necator* H16 could be further used for chemical production in the strain, such as bioproduction of pimelate (which is used as a building block for the synthesis of Nylon 7,7 (Cartman *et al*., 2018)).

Thus, an optimised directed evolution *via* random chemical mutagenesis method for *C. necator* H16 was developed in this chapter, where different chemical or physical mutagens were investigated to choose the mutagen that could be optimal for the strain; also, a biotin analogue was chosen to be used as selection pressure for the identification of rare mutants with an improved phenotype.

4.1.2 EMS random mutagenesis

As mentioned in Section **1.3.2.1**, directed evolution *via* random mutagenesis is a common strategy in bioengineering where mutations rates can be increased with the aid of a chemical or physical mutagen. Chemical mutagens induce high frequencies of base-pair substitutions and some lethality depending on the concentration used. Some of the most widely used chemical mutagens are the alkylating agents *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (NTG) and ethyl methanesulfonate (EMS), which produce transitions at G(guanine)-C(cytosine) sites (Lawrence, 2002). On the other hand, one of the most commonly used physical mutagens is ultraviolet light (UV light, 254 nm). Some strategies of directed evolution consist of combining both mutagens –chemical and physical- which provide outcomes that can satisfy experimental needs.

The external mutagens mentioned in the previous paragraph have been used for random mutagenesis in different experimental approaches of biotin-

overproduction in microorganisms, where all mutagens proved to have induced beneficial mutations, some publications have mentioned that one of the most potent chemical mutagens to induce genetic variability is EMS as it induces high frequencies of base-pair substitutions and little lethality (Coulondre *et al*., 1977; Lawrence, 2002; reviewed in Shahnawaz *et al*., 2014), it has also been described to be even more effective than physical mutagens (Lawrence, 2002; Bhat *et al*., 2005).

As reviewed in Sega (1984), EMS is an alkylating agent that typically produces only point mutations in nucleotides, it owes its biological reactivity to its ethyl group, where G-rich regions are preferred, reacting to form a variety of modified G residues (reviewed in Shahnawaz *et al*., 2014), which results in mispairing and base changes. The strongly biased alkylation G residues forms O^6-ethylguanine, which subsequently can pair only with thymine (T) but not with cytosine (C), then, after DNA repair, the original G/C pair can be substituted by an adenine (A)/thymine (AT pair) (Greene *et al*., 2003). Most of genetic data obtained using microorganisms suggest that mutations induced by EMS will be C-to-T changes (99 %), which result in a G/C to A/T substitution as shown in Error! Reference source not found. (Krieg, 1963; Todd *et al*., 1981; reviewed in Sega, 1984; Cupples *et al*., 1989; Kovalchuk *et al*., 2000; Lawrence, 2002; Greene *et al*., 2003), although at low frequency other mutations can be generated by EMS such as G/C to C/G transversions due to hydrolysis by 7-ethilguanine (Krieg, 1963; reviewed in Kim *et al*., 2005), as reviewed in Sega (1984), there is also some evidence that EMS can induce base-pair deletions or insertions.

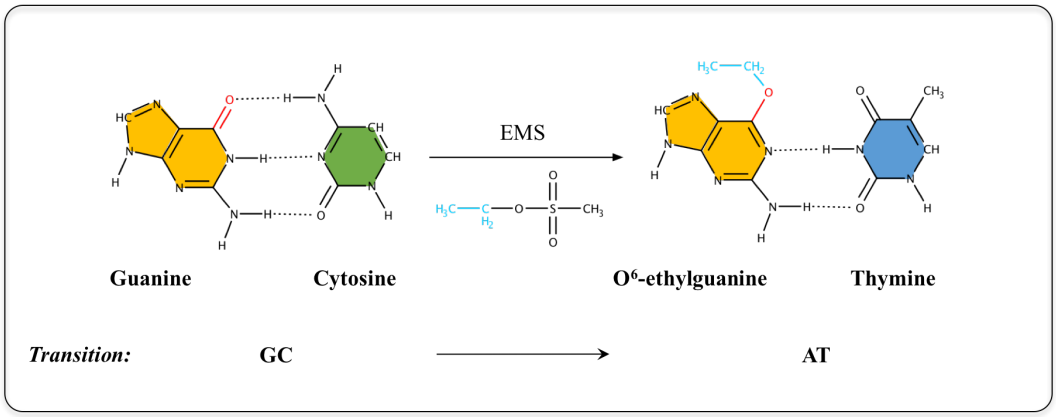


Figure 4.1 EMS mechanism of action as a chemical mutagen. EMS is an alkylating agent that induces predominately GC to AT transitions by alkylating guanine residues to form O^6-ethylguanine which subsequently can only pair with T but not C. Abbreviations: EMS, ethyl methanesulfonate; G, guanine; C, cytosine; A, adenine; T, thymine.

C. necator H16 is a GC-rich microorganism, its two chromosomes have almost identical GC content with > 60 % G + C content (Pohlman *et al*., 2006), and due to EMS has been reported as one of the best external mutagens that prefers G rich regions it is the indicated chemical mutagen to start with for random mutagenesis of its entire genome.

Apart from selecting the external mutagen for random mutagenesis of *C. necator* H16, it is essential to consider in the experimental design the optimal dose of the mutagen, since high concentrations may result in a 100 % killing; a balance between the competing needs for a high mutation frequency and a reasonable survival must be optimised, where usually a 50 % survival rate provides the highest proportion of mutants, it is desirable to avoid doses that would kill more than 95 % of cells (Lawrence, 2002); and, after mutagen treatment, plating dilutions on solid media to select colonies for screening is desirable, as it has the advantage that each specimen of mutation identified is of independent origin (Lawrence, 2002).

The different methodologies used for EMS random mutagenesis in other microorganisms such as *E. coli*, *Bacillus borniphilus*, *S. typhimurium*, and *S. cerevisiae* are described in **Table 4.1.**

Table 4.1 EMS random mutagenesis in microorganisms.

Strain, Cell growth phase previous to mutagenesis	**EMS % (v/v), Incubation time, Survival rate %**	**Cell washes**	**References**
E. coli, early log phase	1.40 % EMS, 30 min, 56 %	KH_2PO_4 buffer	Foster, 1991
E. coli, stationary phase	0.15 % EMS, 90 min, 70 % 0.25 % EMS, 90 min, 40 % 0.30 % EMS, 90 min, 32 % 0.35 % EMS, 90 min, 26 % 0.40 % EMS, 90 min, 22 %	Minimal media	Todd *et al*., 1981
E. coli, stationary phase	1.40 % EMS, 15 min, 71 % 1.40 % EMS, 30 min, 56 % 1.40 % EMS, 60 min, 36 %	KH_2PO_4 buffer	Cupples *et* al., 1989
E. coli, stationary phase	1.00 % EMS, 15 min, 20 % 2.00 % EMS, 10 min, 70 %	Minimal media	Loveless, 1959
B. boroniphilus, early log phase	50.0 % EMS, 60 min, 10 %	KH_2PO_4 buffer	Sen *et al*., 2011
S. typhimurium, stationary phase	4.00 % EMS, 15 min, 30 %	KH_2PO_4 buffer	Loveless, 1959
S. cerevisiae, stationary phase	0.30 % EMS, 30 min, 10 – 50 %	KH_2PO_4 buffer	Lawrence, 2002
S. cerevisiae, early log phase	0.30 % EMS, 40 min, 5 – 30 %	PBS buffer	Sonderegger *et al*., 2003
S. cerevisiae, early log phase	0.30 % EMS, 30 min, 10 %	KH_2PO_4 buffer	Cakar *et al*., 2009

Abbreviations: *E. coli*, *Escherichia coli*; *B. boroniphilus*, *Bacillus boroniphilus*; *S. typhimurium*, *Salmonella typhimurium*; *S. cerevisiae*, *Saccharomyces cerevisiae*; EMS, ethyl methanesulfonate.

The Survival rate % will depend on either the concentration of the mutagen or the time exposure of cells with the mutagen (Foster, 1991), on the other hand, in some other methodologies for random mutagenesis, it is suggested that cells should be mutagenised at the exponential growing log-phase, as according to literature, this greatly enhances the reproducibility of mutagenesis experiments (reviewed in Sega, 1984; Foster, 1991; Sonderegger *et al*., 2003), due to mainly bases in DNA and not in the nucleotide pool are modified (Grzesiuk *et al*., 1993), although some other experiments also mutagenised cells at the stationary phase (Todd *et al*., 1981; Lawrence, 2002).

4.1.3 Biotin analogues

There is a variety of biotin analogues reported in literature that can interfere with the biotin biosynthesis in different microorganisms, such as actithiazic acid

(ACM), α-dehydrobiotin, α-methylbiotin, α-methyldethiobiotin, and amiclenomycin, although some of them are difficult to synthesise at considerable amounts because they are antibiotics, which are produced only in small amounts by actinomycetes (Izumi *et al*., 1977), some others synthetic biotin analogues such as norbiotin, homobiotin, 2-oxo-4-imidazolidinecaproic acid, synthetic amiclenomycin are not commercially available or their synthesis is complicated (Izumi *et al*., 1977; Mann *et al*., 2002; Shi *et al*., 2012); among the synthetic biotin analogues that are easily synthesised or commercially available is 5-(2-thienyl)-valeric acid (TVA). TVA is chemically synthesised from thiophene and glutaric anhydride (Izumi *et al*., 1978).

TVA inhibited the growth of *Rhodotorula glutinis* (fungi), *Bacillus sphaericus* (gram-positive bacteria), and other microorganisms including yeast (*Saccharomyces rouxii*, *Pichia farinose*, *Candida utilis*), bacteria (*Aerobacter* spp., *Agrobaterium radiobacter*, *Bacterium cadaveris*), moulds (*Mucor* spp., *Aspergillus niger*, *Monascus anka*, *Neurospora crassa*), and actinomycetes (*Streptomyces lidicus*) (Izumi *et al*., 1977 and 1978; Yamada *et al*., 1983), where the addition of biotin, dethiobiotin (DTB), and 7,8-diaminopelargonic acid (DAPA) reversed the inhibition of growth in *Rhodotorula glutinis*, while the addition of 7-keto-8-aminopelargonic acid (KAPA) and pimelic acid did not show a recovering effect (Izumi *et al*., 1977), the same effects of cell growth recovery were observed for *B. sphaericus* (Izumi *et al*., 1978); hence Izumi *et al*. (1978) concluded with enzymatic investigations that TVA inhibited the DAPA aminotransferase reaction, which is responsible for DAPA synthesis from KAPA (**Figure 4.2**). These findings suggested that TVA might act as a biotin analogue for a wide variety of microorganisms, which might be linked to the fact that TVA has the same side chain as biotin.

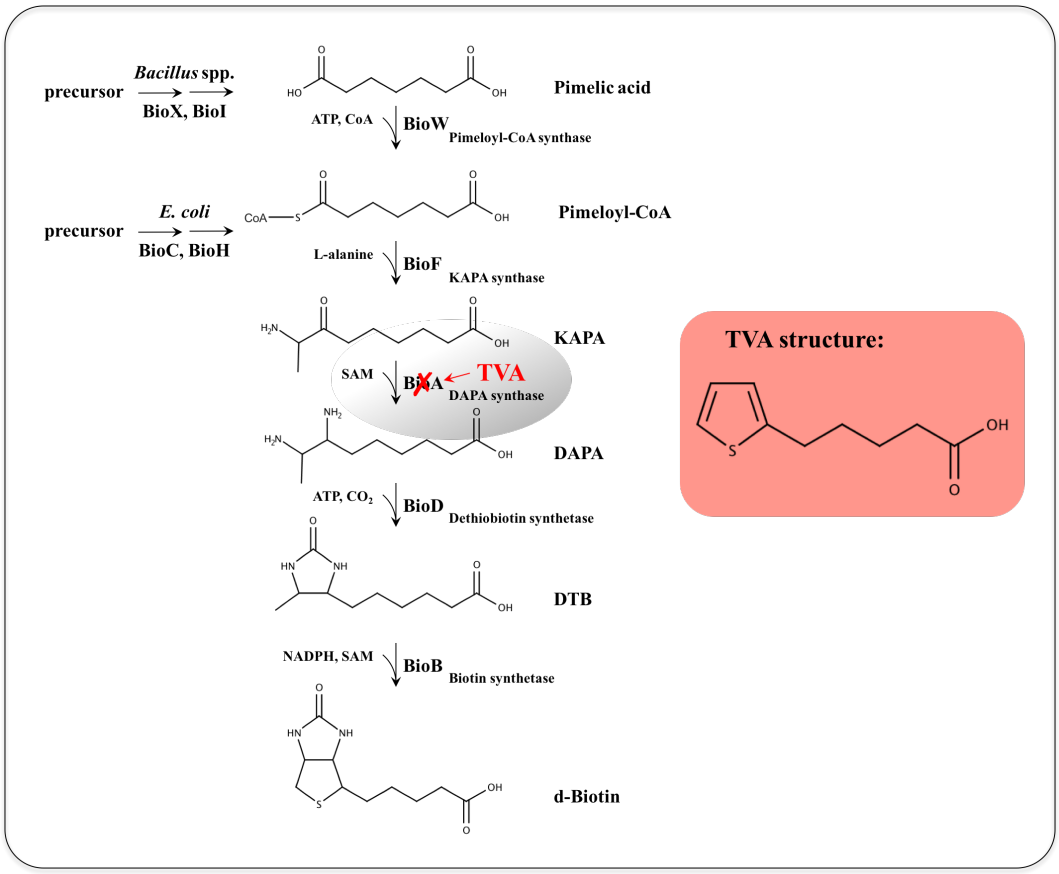


Figure 4.2 Mode of action proposed for the biotin analogue TVA in the orthodox last steps of the biotin biosynthesis pathway. TVA is thought to inhibit BioA, as, after the addition of biotin, DTB, and DAPA, a cell growth recovery has been observed, whereas when KAPA or pimelic acid are added into the media, there is no cell growth recovery. TVA has shown to inhibit cell growth in different microorganisms from yeast, bacteria, moulds, and actinomycetes; the inset shows the chemical structure of TVA. Biotin pathway shown in the illustration corresponds to the biotin synthesis in E. coli and Bacillus spp. Abbreviations: KAPA, 7-keto-8-aminoperlargonic acid; DAPA, 7,8-diaminopelargonic acid; DTB, dethiobiotin; TVA, 5-(2-thienyl)-valeric acid.

A compilation of the biotin analogues that have been used for mutant biotin-overproducing strains is shown in **Table 4.2**, as well as the different microorganisms that have shown cell growth inhibition by the action of these.

Table 4.2 Biotin analogues previously used for biotin studies in different microorganisms and its mechanisms of action.

Biotin analogue	Strains	Description	Comm. avai-lable?	Reference
Acidomycin	*E. coli* *S. marcescens,*	*Other names:* ACM, Actithiazic Acid, 6-(4-oxo-1,3-thiazolidin-2-yl)hexanoic acid *Mechanism of action:* Cell growth recovery after the addition of biotin. ACM might prevent the formation	Yes	Kanzaki *et al.*, 2001; Sakurai *et al.*, 1993

		of D-biotin from dethiobiotin (bioB level)		
Amiclenomycin		*Mechanism of action:* The target of aminoclenomycin is DAPA aminotransferase	No	Mann *et al*., 2002; Izumi *et al*., 1977
Azelaic acid	*B. subtilis*	*Mechanism of action*: It might act at the level of pimelyl-CoA synthetase (*bioW*) as a competitive inhibitor of pimelic acid	Yes	Bower *et al*., 2001
Norbiotin	*S. cerevisiae 139*	*Other names*: 1H-Thieno[3,4-d]imidazole-4-butanoicacid, hexahydro-2-oxo-, (3aS,4S,6aR) *Mechanism of action*: Not investigated	Yes	Izumi *et al*., 1977; Sakurai *et al.*, 1993
α-Dehydrobiotin	*B. subtilis* *E. coli*	*Mechanism of action*: Cell growth recovery after the addition of biotin. It might act like biotin, presumably as a false corepressor for the biotin operon.	No	Eisenberg, 1975; Sakurai *et al*., 1993.
α-methylbiotin	*B. subtilis* *B. cereus,* *Mycobacterium avium* *Mycobacterium phlei*	*Other names*: cis-tetrahydro-a-methyl-2-oxo-thieno[3,4-d]imidazoline-4-valeric acid. *Mechanism of action*: Not investigated	No	Hanka *et al*., 1972; Izumi *et al*., 1977; Sakurai *et al.*, 1993
α-methyldethiobiotin	*B. subtilis* *E. coli* *Mycobacterium avium*	*Other names*: 5-dimethyl-2-oxo-4-imidazolidinehexanoic acid *Mechanism of action*: Not investigated.	No	Hanka *et al*., 1972
5-(2-thienyl)-valeric acid	*B. sphaericus* *B. subtilis* *E. coli* *Rhodotorula glutinis* *S. marcescens*	*Other names*: TVA, 5-(2-thienyl)pentanoic acid *Mechanism of action:* TVA inhibited the DAPA aminotransferase reaction, which is responsible for DAPA synthesis from KAPA (BioA level).	Yes	Bower *et al*., 2001; Izumi *et al*., 1977; Kanzaki *et al.*, 2001; Ifuku *et al*., 1993; Sakurai *et al.*, 1993

Abbreviations: Comm. available, commercially available; ACM, acidomycin; DAPA, 7,8-diaminopelargonic acid; TVA, 5-(2-thienyl)-valeric acid; KAPA, 7-keto-8-aminoperlargonic acid.

4.2 Aim

The aims of **Chapter 4** are:

- To optimise a random mutagenesis method for *C. necator* H16 using EMS as a chemical mutagen.
- To explore the applicability of evolutionary engineering in *C. necator* H16 by using a directed evolution strategy *via* random chemical mutagenesis to generate a library of mutants which could help to understand the biotin pathway in *C. necator* H16.

4.3 Materials and methods

4.3.1 Materials

List of detailed information on the materials used (reagents, kits, equipment, software, media preparation and miscellaneous) can be found in **Appendix 1 - 6**.

4.3.2 Wild-type and mutant strains

All strains that were used in this **Chapter 4** are shown in **Table 4.3**.

Table 4.3 Strains used in this study.

WT and mutant strains	Description	References or source
C. necator H16	Wild-type, Gen^r	DSM 428
m1D9	Mutant strain from first Rd mutagenesis with 0.3 % (v/v) EMS, parent strain *C. necator* H16 WT, Gen^r	This work
m1F2	Mutant strain from first Rd mutagenesis with 0.3 % (v/v) EMS, parent strain *C. necator* H16 WT, Gen^r	This work
m1F5	Mutant strain from first Rd mutagenesis with 0.3 % (v/v) EMS, parent strain *C. necator* H16 WT, Gen^r	This work
m1G6	Mutant strain from first Rd mutagenesis with 0.3 % (v/v)	This work

	EMS, parent strain *C. necator* H16 WT, Gen^r	
m1C3	Mutant strain from first Rd mutagenesis with 0.5 % (v/v) EMS, parent strain *C. necator* H16 WT, Gen^r	This work
m1G9	Mutant strain from first Rd mutagenesis with 1.0 % (v/v) EMS, parent strain *C. necator* H16 WT, Gen^r	This work
m2C2	Mutant strain from second Rd mutagenesis with 0.3 % (v/v) EMS, parent strain *m1F5*, Gen^r	This work
m2C4	Mutant strain from second Rd mutagenesis with 0.3 % (v/v) EMS, parent strain *m1F5*, Gen^r	This work

Abbreviations: Gen^r, gentamicin resistance; Rd, round; WT, wild-type.

4.3.3 Cultivation of *C. necator* H16

C. necator H16 wild-type (WT) and mutant strains were cultivated at 30 ºC using 5 mL of nutrient broth (NB) or mineral salts medium (MSM), and 250 rpm always supplemented with gentamicin [10 μg/mL] when cells were cultivated in falcon tubes. *C. necator* H16 WT and mutant cells were cultivated at 30 ºC using 150 μL of NB or MSM and 1050 rpm always supplemented with gentamicin [10 μg/mL] when cells were cultivated in 96-well microtitre plates. The optical density of all types of cultivation (falcon tubes or microtitre plates) was measured at 595 nm with a microplate reader or at 600 nm with a BioPhotometer.

4.3.4 EMS random mutagenesis for *C. necator* H16

Unless specified otherwise, a pre-culture of *C. necator* H16 was cultivated in 5 mL of MSM with gentamicin for 40 – 44 h at 30 ºC; for random mutagenesis cell preparation, a fresh falcon tube with MSM supplemented with gentamicin was inoculated with the pre-culture at a 1:50 dilution and cultivated at 30 ºC, when cells reached an OD_{600} of 0.9 – 1.0, four millilitres of the cells were then transferred to a sterile 2 mL microcentrifuge tube using two rounds of centrifugation at maximum speed (17,000 x g) for 30 s, the supernatant was removed by pipetting and the cell pellet was resuspended in 1 mL MSM. For chemical mutagen treatment with EMS, three different EMS stock solutions were prepared at different concentrations (0.6 % (v/v), 1.0 % (v/v), and 2.0 % (v/v)), then four different microcentrifuge tubes were prepared with 250 μL of MSM for the control sample (0 % (v/v) EMS), 250 μL of 0.6 % (v/v) EMS stock for a final

concentration of 0.3 % (v/v) EMS, 250 µL of 1.0 % (v/v) EMS stock for a final concentration of 0.5 % (v/v) EMS, and 250 µL of 2.0 % (v/v) EMS stock for a final concentration of 1.0 % (v/v) EMS before 250 µL of the concentrated cells were transferred to each of the four sample tubes to achieve 0 %, 0.3 %, 0.5 % and 1.0 % (v/v) EMS. After exposure to the chemical mutagen for 60 min, cells were centrifuged at maximum speed for 2 min, and the supernatant was removed by pipetting, then, cells were washed thrice by resuspending the cell pellet in 0.5 mL of MSM, after the last washing step, supernatant was removed by pipetting and cell pellet was resuspended in 1 mL of MSM and further diluted into a 2 x 10^4 fold. Then, 100 µL of the diluted cells were plated in MSM with gentamicin agar plates and incubated at 30 ºC for 40 – 48 h. The neat EMS solution was always used in a fume cupboard, and it was always prepared before use. The EMS stock solutions were diluted with 50 mM KH_2PO_4 buffer and then sterilised with a 0.2 µm filter unit. The remaining EMS stock solution and the used EMS waste solution were always treated before discarding with 10 % (w/v) of sodium thiosulphate and 1 M NaOH.

4.3.5 Screening of *C. necator* H16 mutants in TVA

Unless specified otherwise, after EMS mutagenesis of *C. necator* H16, single colonies of mutant strains were picked from MSM agar plates supplemented with gentamicin, and pre-cultured in a 96-well microtitre plate with MSM supplemented with gentamicin; the pre-culture was grown for 24 h or until all cells had reached saturation. Then, cells were transferred with a pin replicator to a fresh microtitre plate with MSM supplemented with different concentrations of TVA and gentamicin for mutant screening, OD_{595} was measured to check for cell growth in the presence of TVA with a microtitre plate reader. Once the microtitre plate of the pre-culture had been used to inoculate the microtitre plate of the outgrowth culture, 50 µL of 50 % (v/v) Glycerol were added to the microtitre plate of the pre-culture and cells were stored at -80 ºC. The screening of mutants was carried out by monitoring cell growth to select the best mutants, cell growth of mutants was always compared against cell growth of the WT and parent strain. After the best mutant of the first round of mutagenesis was characterised and its

phenotype resistant to TVA had been confirmed, the selected mutant was used for the next round of EMS random mutagenesis.

4.3.6 Characterisation of *C. necator* H16 mutants in TVA

A pre-culture of *C. necator* H16 mutant strains was cultivated in 5 mL of MSM with gentamicin for 40 – 44 h at 30 ºC, then, the characterisation of mutants of *C. necator* H16 was performed in 5 mL of MSM supplemented with TVA in falcon tubes with a 1:50 dilution. The growth of the mutant strains was monitored until cells reached stationary phase, optical density was measured at 600 nm in a BioPhotometer. The inoculum of the mutants was always adjusted to an initial OD_{600} of 0.5 using the following formulae:

$$\textbf{Mutant inoculum (µL)} = \frac{OD_{600}\ WT}{OD_{600}\ mutant} \times 100\ \mu L$$

4.3.7 Determination of biotin content in *C. necator* H16 and mutant strains

A pre-culture of *C. necator* H16 WT and mutant strains was cultivated in 5 mL of MSM with gentamicin for 40 h at 30 ºC for biotin quantitation. Standard procedures described in the Biotin Quantitation Assay Kit were followed for determination of biotin content of cells; briefly, the fundamentals of the assay are based on a Biotective Green reagent that consists of avidin labelled with a fluorescent dye and quencher dye ligands, the quencher dye ligands occupy the biotin binding sites, and it is through fluorescence resonance energy transfer (FRET) that the ligand quenches the fluorescence, thus when biotin displaces the quencher dye from the Biotective Green reagent, it yields fluorescence, which is proportional to the amount of biotin. The assay has a sensitivity of 4 to 80 pmol of biotin in a sample. For biotin content determination, a standard curve using biocytin (biotinylated lysine) was performed as described by the manufacturers, and used as a standard for the assay, in the same experiment, 50 µL of the samples were taken directly from the pre-cultures and placed into separate empty wells in a fluorescence microtitre plate, each sample was assayed in triplicates, then, 50

μL of 2x Biotective Green reagent were added to each microtitre well containing either a sample or a standard. The fluorescence microtitre plate was left at room temperature for 5 minutes in the dark before the fluorescence was measured in a fluoroskan microplate reader, where the excitation/emission used for determination of the biotin content was set at 485/538 nm.

4.4 Results and Discussion

4.4.1 Development of EMS random mutagenesis method for *C. necator* H16

Due to a method of EMS random mutagenesis for *C. necator* H16 has not been reported to date, different methodologies used for EMS random mutagenesis in other microorganisms were analysed to design the random mutagenesis method for *C. necator* H16 based on publications described for other microorganisms such as *E. coli*, *Bacillus borniphilus*, *S. typhimurium*, and *S. cerevisiae* (**Table 4.1**).

Hence, after analysing the data shown in **Table 4.1** on how EMS random mutagenesis methods have been developed in other microorganisms, the following variables were taken into consideration for the proposed EMS random mutagenesis method developed for *C. necator* H16 in order to optimise the method: 1) Survival rate %: the target survival rate was set to be within the 50 % survival rate range; 2) Cell growth phase previous to mutagenesis: the cells for chemical mutagenesis was set to be both at the early log-phase and stationary phase; 3) Incubation time of EMS treatment: from the ranges used in different EMS methods shown in **Table 4.1**, it was decided to test mutagenesis with an incubation time of 60 min; 4) Cell wash: before and after mutagenesis treatment, three different buffers were tested, a) sodium thiosulphate, b) buffer solution (KH_2PO_4), and c) MSM; 5) Outgrowth after mutagenesis treatment: two different cell cultivations for the outgrowth can be performed, a) an overnight culture of mutagenised cells before plating treated cells in solid medium, and b) plate mutagenised cells directly in solid medium after mutagen treatment. The first option is recommended to allow cells to stabilise before they are subjected to selective conditions, which can help to enhance the production and expression of mutations (Lawrence, 2002), while the second option is suggested to avoid that the progeny of a small number of mutants could dominant the resultant population which occurs during the overnight culture, as this would give a large proportion of

siblings and could make the process of screening more laborious (Foster, 1991); both strategies were considered to be tested in the experimental design.

Once the variables of the experimental design for random mutagenesis in *C. necator* H16 were defined, the different variables were tested in order to optimise the EMS random mutagenesis method for *C. necator* H16; the first method served as a basis to design the second, and so forth depending on the results obtained, where all the variables specified in the previous paragraph were tested. The experimental designs performed are shown in **Table 4.4**.

Table 4.4 Variables tested for optimisation of EMS random mutagenesis for C. necator H16.

Cell cultivation, Vol. of cells, Cell washes before mutagenesis	**EMS % (v/v), Incubation$_t$**	**Cell washes after mutagenesis**	**Cell outgrowth**
Method 1			
Early-log phase, 0.5 mL cells, KH_2PO_4 wash (2x)	0 % EMS, 60 min, (-)ve 0.30 % EMS, 60 min 0.50 % EMS, 60 min 0.75 % EMS, 60 min 1.00 % EMS, 60 min	$Na_2S_2O_3$ (2x) MSM (2x)	Plate immediately
Method 2			
Stationary phase, 0.5 mL cells, KH_2PO_4 wash (2x)	0 % EMS, 60 min, (-)ve 0.50 % EMS, 60 min 1.00 % EMS, 60 min	**KH_2PO_4 (2x)**	Plate immediately
Method 3 (*a* and *b*)*			
Stationary phase, 0.5 mL cells, KH_2PO_4 wash (2x)	0 % EMS, 60 min, (-)ve 1.00 % EMS, 60 min **2.00 % EMS, 60 min**	KH_2PO_4 (2x)	Plate immediately (*Method 3a*)* **Plate after O.N. cultivation** (*Method 3b*)*
Method 4			
Early-log phase, 1.0 mL cells,	0 % EMS, 60 min, (-)ve 0.30 % EMS, 60 min 0.50 % EMS, 60 min 1.00 % EMS, 60 min	**MSM (3x)**	Plate immediately

Notes: Main variables changed between one experiment an another are shown in green; 50 mM KH_2PO_4, pH 7.0; and 10 % (w/v) $Na_2S_2O_3$ concentrations were used for cell washes. Abbreviations: Vol., volume; Incubation$_t$, incubation time; KH_2PO_4, potassium phosphate monobasic; (-)ve, negative control; $Na_2S_2O_3$, sodium thiosulphate; O.N., overnight.

The results of the protocols tested out for EMS random mutagenesis in *C. necator* H16 described in **Table 4.4** are shown in **Figure 4.3**, where the survival rate % was calculated using the formulae below:

Survival rate % = # of EMS treated survivors (MSM) × 100 %
# of EMS control survivors (MSM)

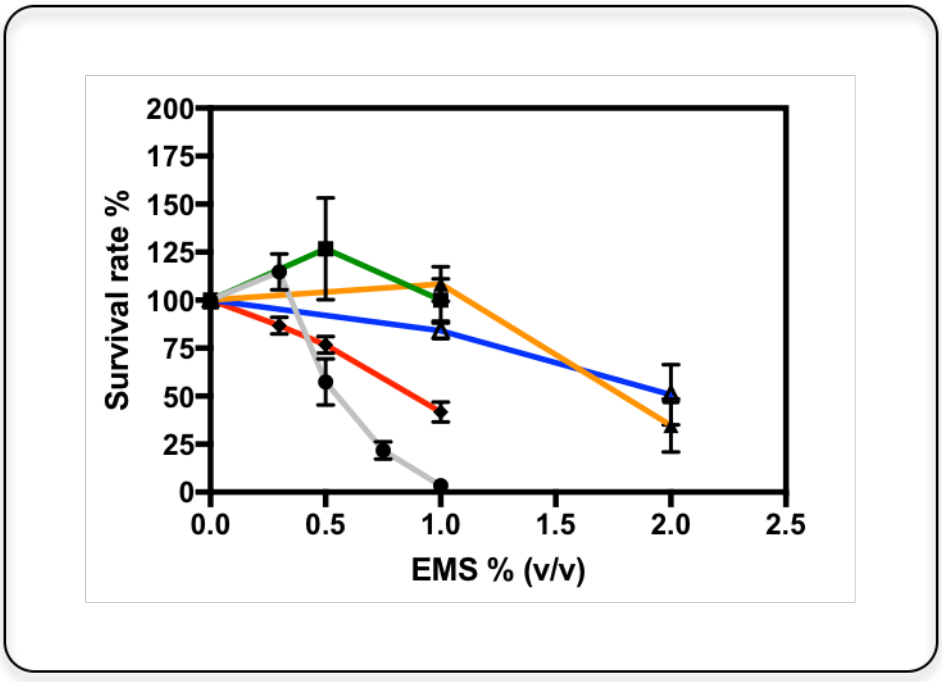


Figure 4.3 Survival rate % of the different EMS random mutagenesis methods tested for C. necator H16. Method 1 (grey line): cells at early-log phase, buffers KH_2PO_4, $Na_2S_20_3$, and MSM, no overnight cultivation after mutagenesis; Method 2 (green line): cells at stationary phase, buffer KH_2PO_4, no overnight cultivation after mutagenesis; Method 3a (orange line): cells at stationary phase, buffer KH_2PO_4, no overnight cultivation after mutagenesis; Method 3b (blue line): cells at stationary phase, buffer KH_2PO_4, overnight cultivation after mutagenesis; Method 4 (red line): cells at early log phase, buffer MSM, no overnight cultivation after mutagenesis. Error bars are ± SEM of three independent determinations. Abbreviations: EMS, ethyl methanesulfonate.

From **Figure 4.3**, it was observed that the Method 1 (grey line) effectively provided a survival rate % decrease as the concentration of EMS increased, where there was > 100 % survival when 0.3 % (v/v) EMS was used to mutagenise cells, while there was 0 % survival when 1.0 % was used. Method 2 (green line), on the other hand, showed a higher survival rate when 0.5 % and 1.0 % (v/v) EMS were used compared to Method 1 (grey line), this due to the fact that the same volume of cells was taken but at different growth phases in both methods, where in Method 2, cells were withdrawn from the culture at the stationary phase and not at the early-log phase as in Method 1 (grey line), hence, cell concentration was higher in Method 2 than in Method 1; From Method 1 and Method 2, it was

observed that all the different buffers used for cell washes worked well for the mutagenesis process. In Method 2, it was observed as well that the different EMS concentrations used (0.5 % and 1.0 % (v/v) EMS) did not provide a survival rate close to 50 %, thus in Method 3a and 3b (orange and blue lines respectively), the EMS concentration was increased.

Methods 3a and 3b have the same experimental design, the only difference was the condition of the outgrowth, whereby in Method 3a (orange line) cells were plated right after EMS mutagenesis and in Method 3b (blue line) cells were plated after overnight culture. For both methods, a higher concentration of EMS (2.0 % (v/v)) was tried out, since previous observations in Method 2, showed no significant survival rate % decrease when they were mutagenised with up to 1.0 % (v/v) EMS. Effectively, doubling the EMS concentration provided a lower survival rate % of ~ 50 % for both methods. Another observation noted from Method 3a and 3b, was that the conditions of cell outgrowth after mutagenesis provided similar results in terms of survival rate %, hence, for the next experimental design only the cell growth condition of Method 3a (cells plated right after mutagenesis) was considered, since this could help to avoid to have a large population of siblings during the screening process.

The last Method 4 (red line) was designed based on previous results obtained in Method 1, 2, 3a and 3b, where the variables with the best outcomes were considered to design the Method 4. The variables considered were: 1) mutagenise cells at the early-log phase due to from literature it is known that this greatly enhances the reproducibility of mutagenesis experiments, 2) increase the volume of cells in order to have a higher concentration of cells (in the early-log phase), 3) use MSM as buffer for cell washes, which could help to make the process more straightforward, and 4) for the cell outgrowth conditions, cells would be plated immediately after EMS mutagenesis (*i.e.*, no overnight culture before plating the cells). Thus, Method 4 (red line), was performed with the previous variables mentioned, where it was observed that compared to Method 1 (grey line) (which also was tried out with cells in the early-log phase but with a lower concentration of cells), a higher survival rate % when 0.5 % and 1.0 % of EMS was used was observed, this might be due to the cell concentration was higher in the samples of

Method 4. Method 4 provided the best results concerning reproducibility, where the standard error of mean (SEM) (error bars shown in **Figure 4.3**) was within the 10 % range. The survival rates of Method 4 were: 87 % survival rate when cells were mutagenised with 0.3 % (v/v) EMS, 43 % survival rate when cells were mutagenised with 1.0 % (v/v) EMS, and 75 % survival rate when 0.5 % EMS was used to mutagenise cells. Once the EMS random mutagenesis method for *C. necator* H16 was optimised (Method 4) for a ~50 % survival rate, the selection pressure to be used was analysed.

4.4.2 Selection pressure to understand *C. necator H16* biotin biosynthesis pathway

In **Section 1.5.7.6**, some examples of biotin-overproducing bacterial mutants are mentioned, whereby when directed evolution *via* random mutagenesis was used to mutagenise whole-cells, a selection pressure (biotin analogue) was used in order to select mutants with an improved phenotype. The biotin analogues previously used for biotin studies in different microorganisms and its mechanisms of action is shown in **Table 4.2**. Thus, after *C. necator* H16 strains have been subjected to mutagen treatment these can be screened in the presence of a biotin analogue, where it would be expected that only the mutants capable of growing under those conditions would be the mutants that have gone through genetic mutations in regions that could be related to the biotin biosynthesis pathway.

Due to the characteristics in terms of 1) commercial availability, 2) microorganisms which have a similar pathway to *C. necator* H16 (*i.e.*, *bioC*, *bioH* genes involved in the biotin synthesis such as *E. coli*) for biotin synthesis, and 3) the mechanism of inhibition of biotin biosynthesis studied in more detail (*i.e.*, that it is known how the biotin analogue inhibits the synthesis of biotin), TVA biotin analogue was chosen to be used as a selection pressure for the study of the biotin biosynthesis pathway in *C. necator* H16.

4.4.2.1 MIC of TVA biotin analogue for *C. necator* H16

The first experiment after selecting the biotin analogue as a selection pressure for *C. necator* H16 mutants was to prove cell growth inhibition of the *C. necator* H16 WT in the presence of TVA in synthetic medium (MSM). In order to do so, a growth inhibition curve was performed, where the minimum inhibitory concentration (MIC) of TVA for *C. necator* H16 was the lowest concentration of TVA at which no growth was visible after 24 h. The growth inhibition curve is shown in **Figure 4.4**.

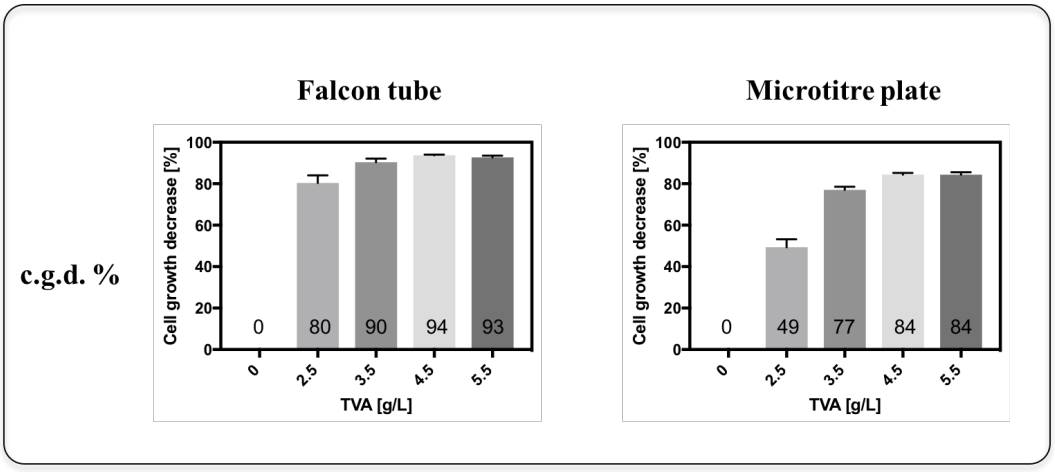


Figure 4.4 Growth inhibition of C. necator H16 by the biotin analogue TVA. C. necator H16 was cultivated at 30 ºC for 24 h in 5 mL MSM for Falcon tube experiments and 150 µL MSM for Microtitre plate experiments containing the indicated concentrations of TVA. Error bars are ± SEM of three independent determinations. Abbreviations: c.g.d. %, cell growth decrease %; TVA, 5-(2-thienyl)-valeric acid.

From **Figure 4.4** it was observed that TVA effectively inhibits cell growth of *C. necator* H16, as reported in literature for other microorganisms such as *E. coli* and *B. subtilis*, this suggests that TVA might act as a biotin analogue as well for *C. necator* H16. On the other hand, the growth inhibition of *C. necator* H16 by TVA was tested in two different conditions, falcon tubes (5 mL medium) and microtitre plates (150 µL medium), since these were the two conditions where all further experiments would be carried out. In these sets of experiments both optical densities were measured with the same microplate reader at 595 nm (OD_{595}).

The MIC of *C. necator* H16 when cells were cultivated in falcon tubes, was observed when TVA had concentrations > 4.5 g/L, while the MIC of *C. necator* H16 when cells were cultivated in microtitre plates was observed when TVA had concentrations > 5.5 g/L, as there was no further growth even after 48 h cultivation at these concentrations. From the results obtained of both cultivation conditions (falcon tubes and microtitre plates), it was noticeable that the cell growth decrease % (c.g.d. %) was higher when the TVA concentration was increased in both cultivation conditions, although it was observed as well that at the same concentrations of TVA, different c.g.d. % resulted among the different cultivation conditions, *e.g*., when 2.5 g/L of TVA was added into the medium, an 80 % c.g.d. was observed when cells were grown in falcon tubes, while a 49 % c.g.d. was observed when cells were grown in microtitre plates at the same TVA concentration; the impact of cell growth inhibition was always higher in falcon tubes than in microtitre plates. The fact that this was reproducible confirms the that cells grow differently in different reactor cultivation conditions, and hence, the effect of TVA inhibition was also different depedending on the reactor used. The difference in the impact of cell growth decrease in the presence of TVA was constantly observed in screening and characterisation studies when different reactors were used.

The reason behind the difference in bacterial cell growth using different reactors (*i.e.*, falcon tubes or microtitre plates in these set of experiments) is the different conditions provided in each of them, such as the shaking, since this parameter not only serves to ensure that cells are uniformly suspended before each OD measurement but also to provide aeration by providing adequate oxygen for cell growth (reviewed in Hall *et al*., 2013). The type of culture vessel used to cultivate cells affects cell growth at different cell growth phases, where one of the main factors that contribute to the cell growth in different reactors is oxygenation provided by the headspace of the vessel (Kram *et al*., 2014). Kram *et al*. (2014) hypothesize that exposure to oxygen allow more efficient respiration. Due to in the bioindustry usually large culture reactors are used to scale up experiments, it is important to consider the fact that the physiology of the bacterial cells is quite different in various types of reactors, the reactors could impact directly in growth, gene expression, protein production, resistance to stress, etc.

4.4.3 Genetic variation of *C. necator* H16 obtained after EMS random mutagenesis – First Round

Once the EMS random mutagenesis method for *C. necator* H16 was developed, the selection pressure determined, and the c.g.d. % at different concentrations of the biotin analogue TVA analysed, the first round of random mutagenesis consisted on mutating *C. necator* H16 WT cells with different concentrations of EMS (0.3 %, 0.5 %, and 1.0 % (v/v) EMS). The procedure followed for the generation and selection of mutants is shown in **Figure 4.5**.

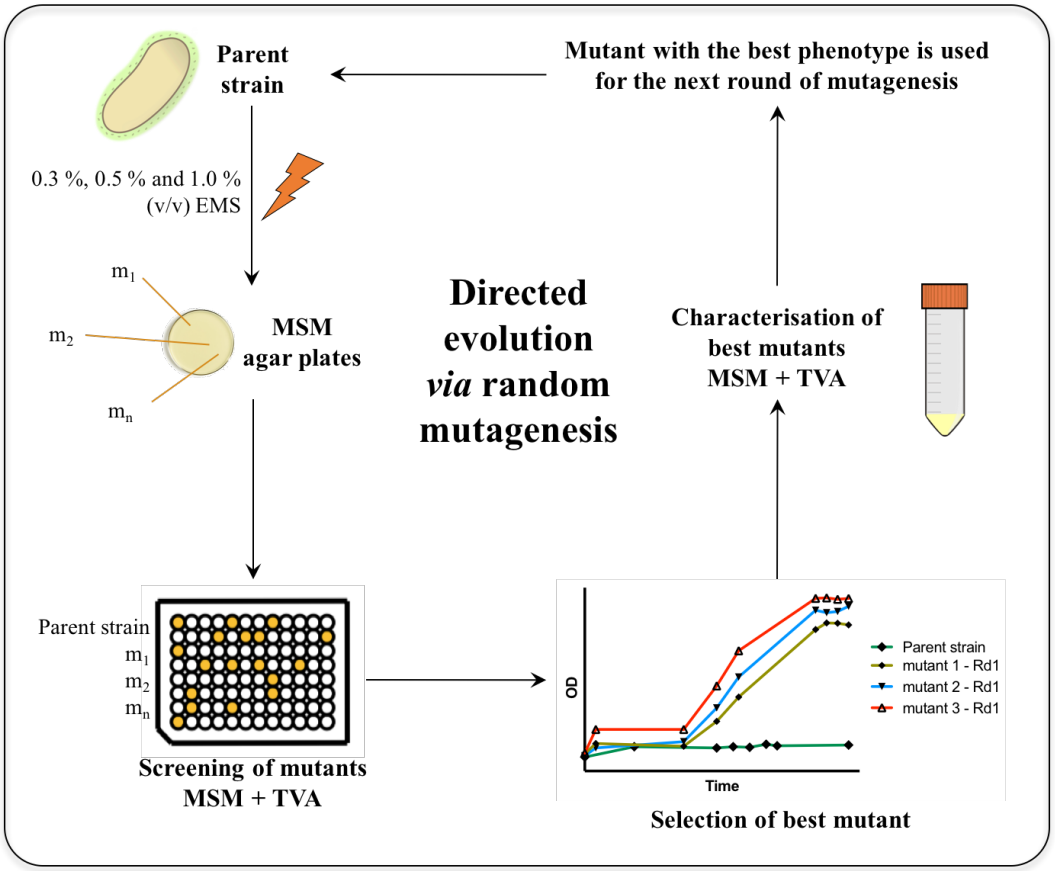


Figure 4.5 Diagram of directed evolution via random mutagenesis: screening, and characterisation scheme for C. necator H16 WT cells. C. necator H16 WT cells (parent strain) were exposed to random mutagenesis with 0.3 %, 0.5 %, and 1.0 % (v/v) EMS. Mutant colonies were then screened in MSM supplemented with different concentrations of TVA in microtitre plates to cover a higher number of mutants. The best mutants were then characterised in falcon tubes to confirm the desired phenotype, and then the best mutant was chosen as the parent strain for the next round of mutagenesis. Abbreviations: EMS, ethyl methanesulfonate; m, mutants; TVA, 5-(2-thienyl)-valeric acid; Rd, round.

The survival rate % of the first round of mutagenesis with 0.3 % (v/v) EMS was 84 % survival rate, with 0.5 % (v/v) EMS was 72 % survival rate, and with 1.0 % (v/v) EMS was 46 % survival rate as observed previously in the development of the EMS random mutagenesis for *C. necator* H16 method. A total of 360 mutant strains from the first round of mutagenesis mutated with different concentrations of EMS were screened throughout the entire study in MSM supplemented with gentamicin and different concentrations of TVA. Each microtitre plate contained control samples to compare cell growth and find beneficial mutants, control samples were the parent strain *C. necator* H16 WT in MSM, and *C. necator* H16 in MSM supplemented with TVA, which were always included in triplicates in all experiments. The best mutant strains for each of the experiments performed for the first round of random mutagenesis were chosen as follows: for each OD_{595} measurement taken during the cell growth curve, the two highest OD_{595} values corresponding to mutant strains were selected to be followed through the entire cell growth study, then, all the mutants selected were compared against each other and against the parent strain to identify the mutant strains with the best cell growth performance in the presence of TVA. The mutant strains were named as "m" for mutant, followed by a "number" from 1 to 2, which corresponds to the round of random mutagenesis, and a "letter + number" corresponding to the well of the microtitre plate where the mutant strain was grown.

Screening analysis was performed in microtitre plates with 150 μL of MSM with or without TVA and OD was measured at 595 nm, while the characterisation analysis was performed in 5 mL of MSM with or without TVA and OD was measured at 600 nm. Through all the experiments performed in this **Chapter 4**, the OD readings of samples measured with cuvettes appear to reach higher ODs (*e.g.* cells in MSM without TVA, cells could reach a max $OD_{600} \approx 14$) than the samples measured with microtitre plates (*e.g.* cells in MSM without TVA, cells could reach a max $OD_{595} \approx 2$), this is due to the method of photometry differs between a standard spectrophotometer and a microtitre plate reader. The OD readings of the samples measured in a cuvette is determined by the light sent

through the sample in a horizontal manner, while the OD readings of the samples measured with the microtitre plate is determined by the light sent through the sample in a vertical manner. With the cuvette, the path length is usually 1 cm, while the path length in a microtitre plate reader depends on the volume of the sample. Due to according to Beer's Law ($A = \varepsilon BC$), the concentration of a solution (C) depends on its absorbance (A), path length (b), and molar absorptivity (ε), an identical sample that is measured in a standard spectrophotometer and in a microtitre plate reader will provide different values due to the difference in path length.

4.4.3.1 Analysis of mutant strains of *C. necator* H16 obtained from EMS random mutagenesis with 0.3 % (v/v) EMS – First Round

When *C. necator* H16 WT cells were mutagenised with 0.3 % (v/v) EMS, the screening of the mutants was performed in two different microtitre plates with different concentrations of TVA. A total of 180 mutant strains were screened in two different 96-well microtitre plates with MSM supplemented with gentamicin and TVA; one microtitre plate was set with 1.0 g/L of TVA and the other one with 3.5 g/L of TVA, these two different concentrations of TVA were used as it was unknown how would mutants grow in the presence of the biotin analogue, hence a low concentration (1.0 g/L TVA) was used in case mutants were not capable of growing at high concentrations of TVA, while a higher concentration (3.5 g/L TVA) was used in case the lower concentrations would not allow detecting a difference in growth performance between the WT and the mutant strains (**Figure 4.6**).

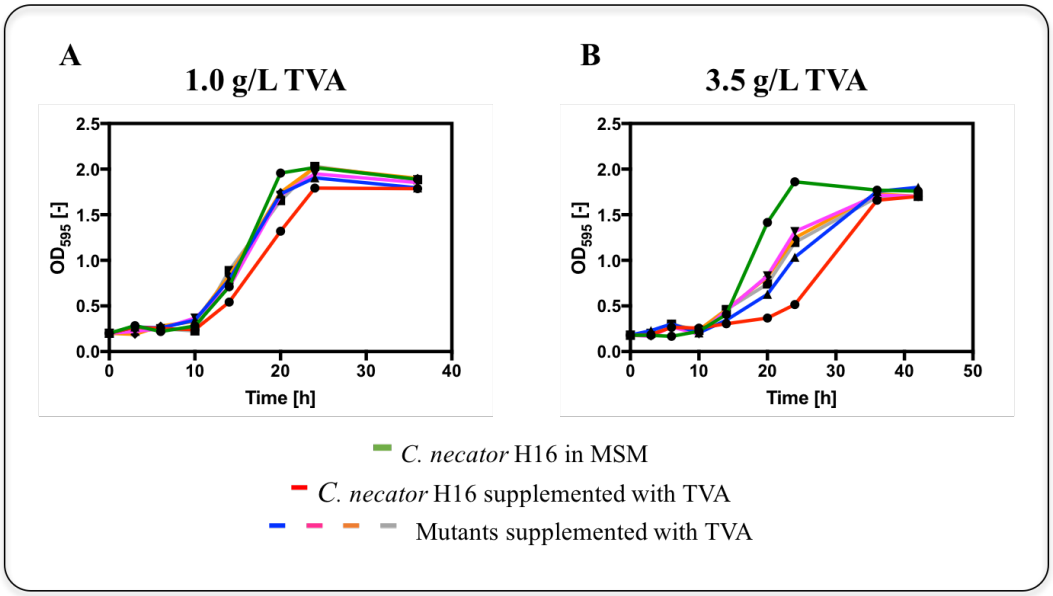


Figure 4.6 Screening of mutant strains mutagenised with 0.3 % (v/v) EMS. Screening analysis was performed in microtitre plates with 150 µL of MSM with or without TVA. ***A)*** *C. necator H16 WT in MSM (circle, green line), and C. necator H16 WT in MSM supplemented with 1.0 g/L TVA (circle, red line); m1A6 in MSM supplemented with 1.0 g/L TVA (triangle, blue line), m1B11 in MSM supplemented with 1.0 g/L TVA (inverted triangle, pink line), m1D9 in MSM supplemented with 1.0 g/L TVA (diamond, orange line), and m1F2 in MSM supplemented with 1.0 g/L TVA (square, grey line).* ***B)*** *C. necator H16 WT in MSM (circle, green line), and C. necator H16 WT in MSM supplemented with 3.5 g/L TVA (circle, red line); m1D9 in MSM supplemented with 3.5 g/L TVA (triangle, blue line), m1F5 in MSM supplemented with 3.5 g/L TVA (inverted triangle, pink line), m1G6 in MSM supplemented with 3.5 g/L TVA (diamond, orange line), and m1G7 in MSM supplemented with 3.5 g/L TVA (square, grey line). Error bars are ± SEM of three independent determinations. Abbreviations: TVA, 5-(2-thienyl)-valeric acid; m, mutant.*

The screening of this first round of random mutagenesis showed mutants with an improved phenotype in cell growth in the presence of the TVA biotin analogue. All the best mutants shown in **Figure 4.6** of both microtitre plates showed less cell growth decrease in the presence of TVA than the WT parent strain. *C. necator* H16 showed an 11 % c.g.d. when grown in 1.0 g/L of TVA, and a 72 % c.g.d. when grown in 3.5 g/L of TVA after 24 h cultivation as observed in previous MIC studies. Although all mutants grew very similarly in both microtitre plates, the two mutants of each microtitre plate with the highest fold in growth compared to the WT were chosen as the best mutants; the fold comparison was analysed at 24 h of cultivation. The best mutants were the *m1D9* and *m1F2* from the 1.0 g/L of TVA microtitre plate, where growth was 1.1-fold faster compared to the WT in TVA for both mutants, and *m1F5* and *m1G6* from the 3.5 g/L of TVA microtitre

plate, where growth was 2.5 and 2.4-fold faster compared to the WT in TVA respectively.

After selecting the best mutants from both microtitre plates, the characterisation of all four mutants was performed in 5 mL of MSM supplemented with TVA in falcon tubes, a higher concentration of TVA was used to analyse the performance of the mutant strains. The cell growth curve for the characterisation of mutant strains is shown in **Figure 4.7**, where 4.5 g/L of TVA was used to confirm their phenotype. Mutant strains were also grown in MSM without TVA to check if mutant strains could still grow like the WT in sodium gluconate.

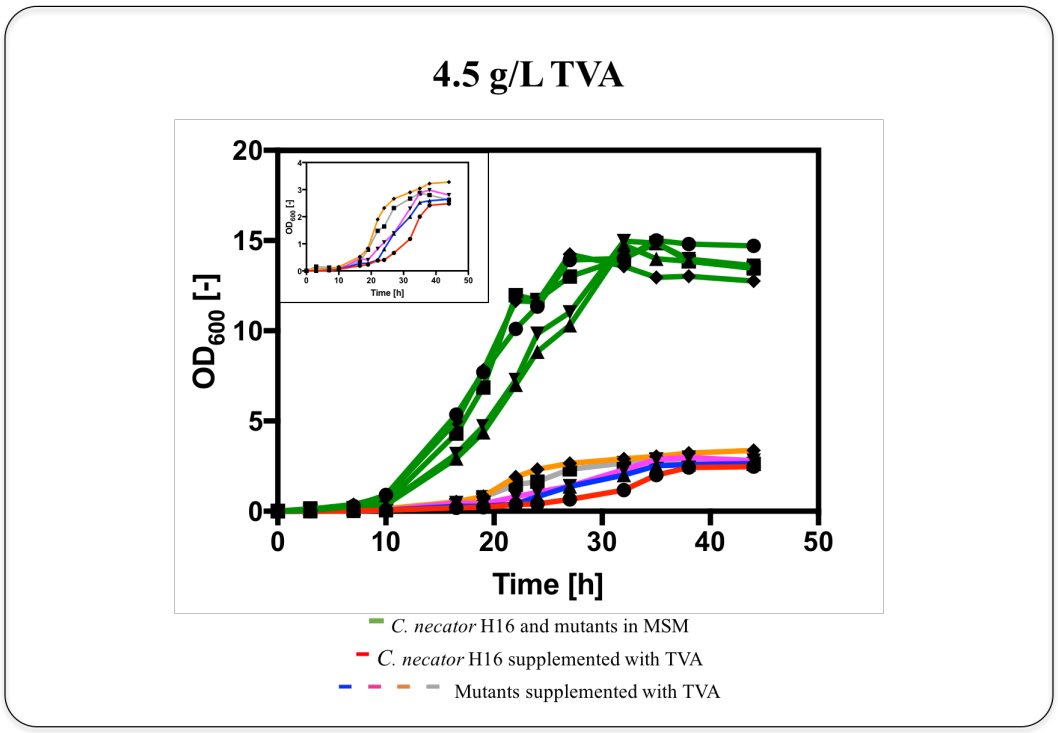


Figure 4.7 Characterisation of mutant strains mutagenised with 0.3 % (v/v) EMS. Characterisation analysis was performed in falcon tubes with 5 ml of MSM with or without TVA. C. necator H16 WT in MSM (circle, green line), and C. necator H16 WT in MSM supplemented with 4.5 g/L TVA (circle, red line); m1D9 mutant strain in MSM (triangle, green line), m1F2 mutant strain in MSM (inverted triangle, green line), m1F5 mutant strain in MSM (diamond, green line), and m1G6 mutant strain in MSM (square, green line); m1D9 mutant strain in MSM supplemented with 4.5 g/L TVA (triangle, blue line), m1F2 mutant strain in MSM supplemented with 4.5 g/L TVA (inverted triangle, pink line), m1F5 mutant strain in MSM supplemented with 4.5 g/L TVA (diamond, orange line), and m1G6 mutant strain in MSM supplemented with 4.5 g/L TVA (square, grey line). The inset shows the magnified cell growth curve of C. necator H16 WT and mutant strains in 4.5 g/L TVA. Abbreviations: TVA, 5-(2-thienyl)-valeric acid; m, mutant.

From **Figure 4.7** it was confirmed that the cell growth of the mutant strains screened in the presence of 1.0 g/L of TVA (*m1D9* and *m1F2*), and the mutants screened in the presence of 3.5 g/L of TVA (*m1F5* and *m1G6*) was effectively improved compared the cell growth of *C. necator* H16 WT in TVA. On the other hand, it was observed as well that mutant strains were also capable of growing in MSM as the WT strain. *C. necator* H16 showed a 96 % c.g.d. when grown in 4.5 g/L of TVA after 24 h cultivation as observed in previous MIC studies.

The best mutant from the characterisation studies was selected by analysing the performance in cell growth of all mutants in the presence of TVA, where the highest fold in growth compared to the WT was chosen as the best mutant, the fold comparison was analysed at 24 h of cultivation. The best mutant from the screening was the *m1F5* mutant (diamond, orange line), which showed to be 5.8-fold faster than the WT strain in the presence of TVA, followed by the *m1G6* mutant (square, grey line), which showed to be 4.1-fold faster than the WT strain, and then by the *m1F2* (inverted triangle, pink line) and *m1D9* (triangle, blue line) mutants; which showed to be 2.6 and 2.0-fold faster than the WT strain respectively when cells were grown at 4.5 g/L of TVA.

4.4.3.2 Analysis of mutant strains of *C. necator* H16 obtained from EMS random mutagenesis with 0.5 % (v/v) EMS – First Round

When *C. necator* H16 WT cells were mutagenised with 0.5 % (v/v) EMS, the screening of the mutants was set in a microtitre plate with 4.5 g/L of TVA for this first round of mutagenesis, where a total of 90 mutant strains were screened (**Figure 4.8**).

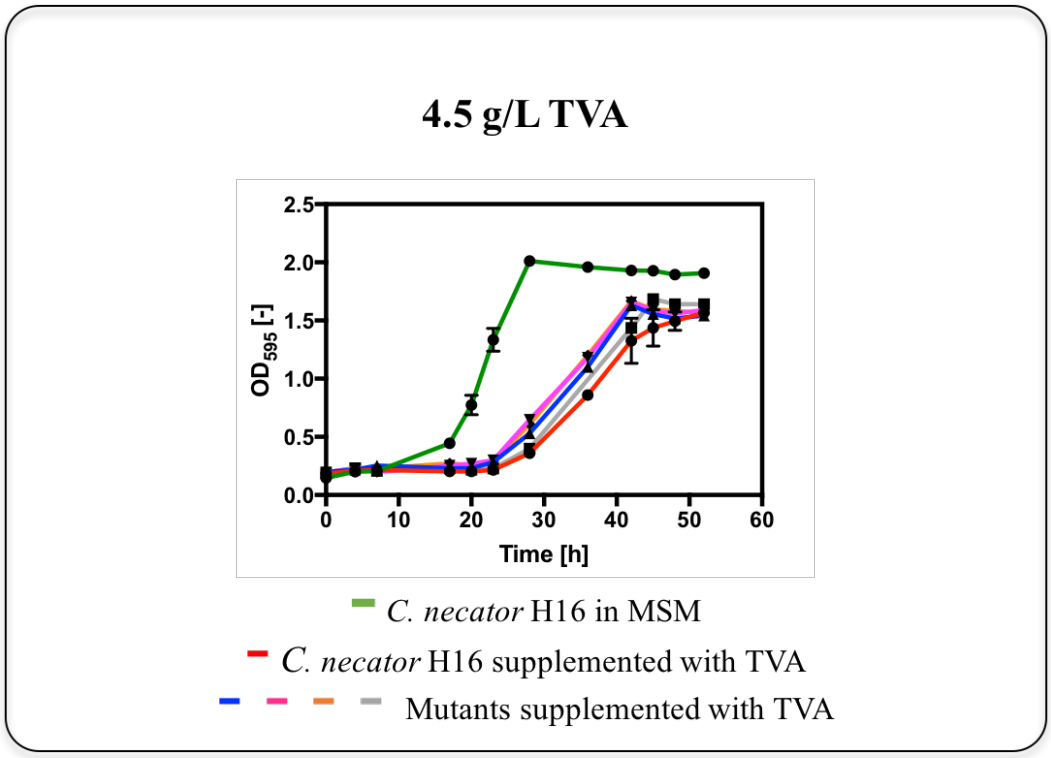


Figure 4.8 Screening of mutant strains mutagenised with 0.5 % (v/v) EMS. Screening analysis was performed in microtitre plates with 150 µL of MSM with or without TVA. C. necator H16 WT in MSM (circle, green line), and C. necator H16 WT in MSM supplemented with 4.5 g/L TVA (circle, red line); m1B8 in MSM supplemented with 4.5 g/L TVA (triangle, blue line), m1C3 in MSM supplemented with 4.5 g/L TVA (inverted triangle, pink line), m1C4 in MSM supplemented with 4.5 g/L TVA (diamond, orange line), and m1G7 in MSM supplemented with 4.5 g/L TVA (square, grey line). Error bars are ± SEM of three independent determinations. Abbreviations: TVA, 5-(2-thienyl)-valeric acid; m, mutant.

The screening of this first round of random mutagenesis showed mutants with an improved phenotype in cell growth in the presence of the TVA biotin analogue. All the best mutants shown in **Figure 4.8** showed less cell growth decrease in the presence of TVA than the WT parent strain. *C. necator* H16 showed an 85 % c.g.d. when grown in 4.5 g/L of TVA after 24 h cultivation as observed in previous MIC studies. The mutant with the highest fold in growth compared to the WT was chosen as the best mutant, the fold comparison was analysed at 24 h of cultivation. The best mutant was the *m1C3*, where growth was 1.4-fold faster compared to the WT in TVA.

After selecting the best mutant, the characterisation of *m1C3* was performed in 5 mL of MSM supplemented with TVA in falcon tubes. The cell growth curve for

the characterisation of the mutant strain is shown in **Figure 4.9**, where 4.5 g/L of TVA was used to confirm its phenotype. The mutant strain was also grown in MSM without TVA to check if it could still grow like the WT in sodium gluconate.

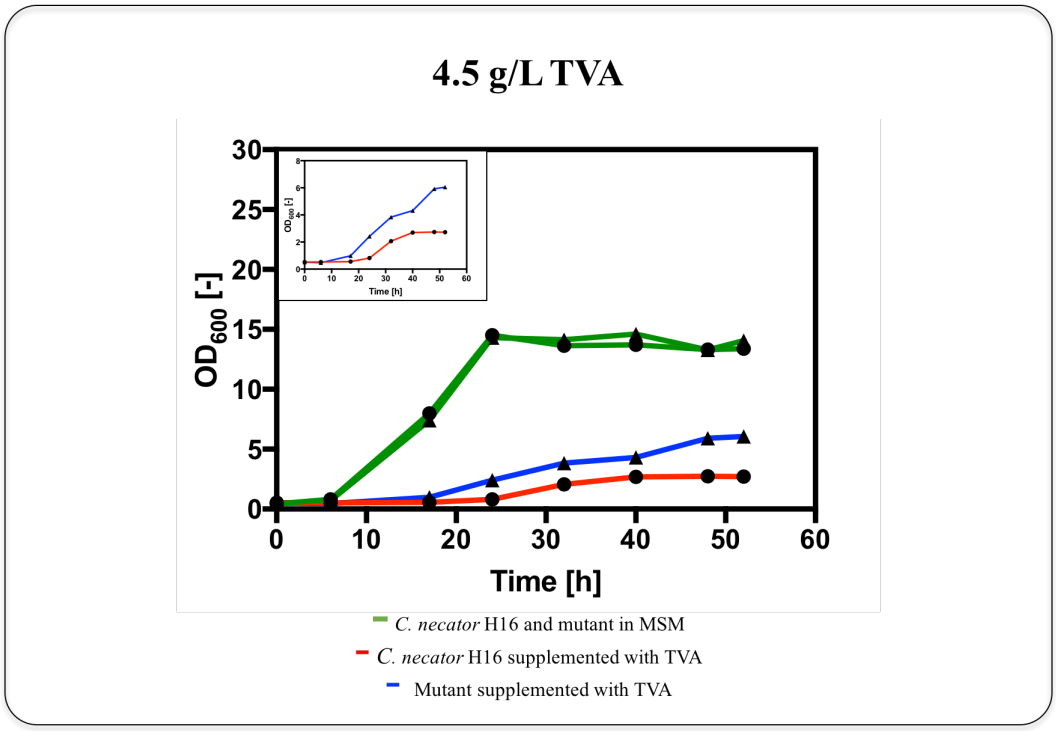


Figure 4.9 Characterisation of mutant strains mutagenised with 0.5 % (v/v) EMS. Characterisation analysis was performed in falcon tubes with 5 ml of MSM with or without TVA. C. necator H16 WT in MSM (circle, green line), and C. necator H16 WT in MSM supplemented with 4.5 g/L TVA (circle, red line); m1C3 mutant strain in MSM (triangle, green line), and m1C3 mutant strain in MSM supplemented with 4.5 g/L TVA (triangle, blue line). The inset shows the magnified cell growth curve of C. necator H16 WT and m1C3 in 4.5 g/L TVA. Abbreviations: TVA, 5-(2-thienyl)-valeric acid; m, mutant.

From **Figure 4.9** it was confirmed that the cell growth of the mutant strain screened in the presence of 4.5 g/L of TVA (*m1C3*) was effectively improved compared the cell growth of *C. necator* H16 WT in TVA. On the other hand, it was observed as well that the mutant strain was also capable of growing in MSM as the WT. *C. necator* H16 showed a 94 % c.g.d. when grown in 4.5 g/L of TVA after 24 h cultivation as observed in previous MIC studies. The performance in cell growth of the mutant strain in the presence of TVA was analysed at 24 h of

cultivation. The *m1C3* mutant (triangle, blue line) showed to be 3.0-fold faster than the WT strain in the presence of 4.5 g/L of TVA.

4.4.3.3 Analysis of mutant strains of *C. necator* H16 obtained from EMS random mutagenesis with 1.0 % (v/v) EMS – First Round

When *C. necator* H16 WT cells were mutagenised with 1.0 % (v/v) EMS, the screening of the mutants was set in a microtitre plate with 4.5 g/L of TVA, where a total of 90 mutant strains were screened (**Figure 4.10**).

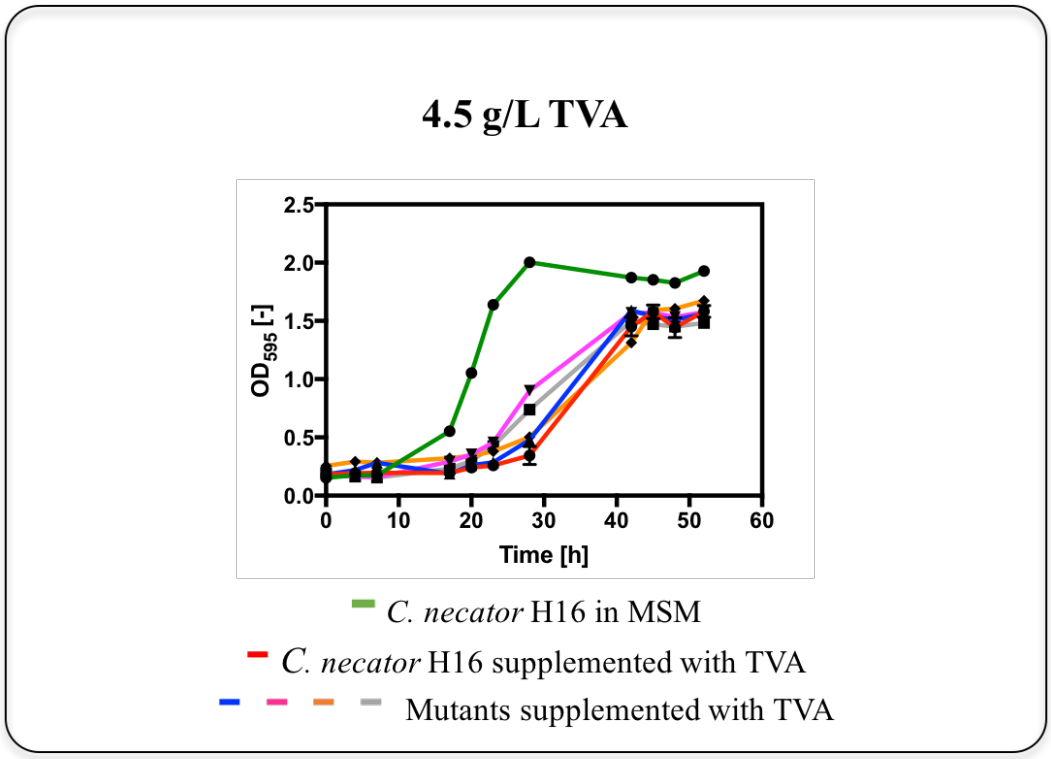


Figure 4.10 Screening of mutant strains mutagenised with 1.0 % (v/v) EMS. Screening analysis was performed in microtitre plates with 150 µL of MSM with or without TVA. C. necator H16 WT in MSM (circle, green line), and C. necator H16 WT in MSM supplemented with 4.5 g/L TVA (circle, red line); m1B11 in MSM supplemented with 4.5 g/L TVA (triangle, blue line), m1G9 in MSM supplemented with 4.5 g/L TVA (inverted triangle, pink line), m1H3 in MSM supplemented with 4.5 g/L TVA (diamond, orange line), and m1H7 in MSM supplemented with 4.5 g/L TVA (square, grey line). Error bars are ± SEM of three independent determinations. Abbreviations: TVA, 5-(2-thienyl)-valeric acid; m, mutant.

The screening of this first round of random mutagenesis showed mutants with an improved phenotype in cell growth in the presence of the TVA biotin analogue. All the best mutants shown in **Figure 4.10** showed less cell growth decrease in the presence of TVA than the WT parent strain. *C. necator* H16 showed an 84 % c.g.d. when grown in 4.5 g/L of TVA after 24 h cultivation as observed in previous MIC studies. The mutant with the highest fold in growth compared to the WT was chosen as the best mutant, the fold comparison was analysed at 24 h of cultivation. The best mutant was the *m1G9*, where growth was 1.8-fold faster compared to the WT in TVA.

After selecting the best mutant, the characterisation of *m1G9* was performed in 5 mL of MSM supplemented with TVA in falcon tubes. The cell growth curve for the characterisation of the mutant strain is shown in **Figure 4.11**, where 4.5 g/L of TVA was used to confirm its phenotype. The mutant strain was also grown in MSM without TVA to check if it could still grow like the WT in sodium gluconate.

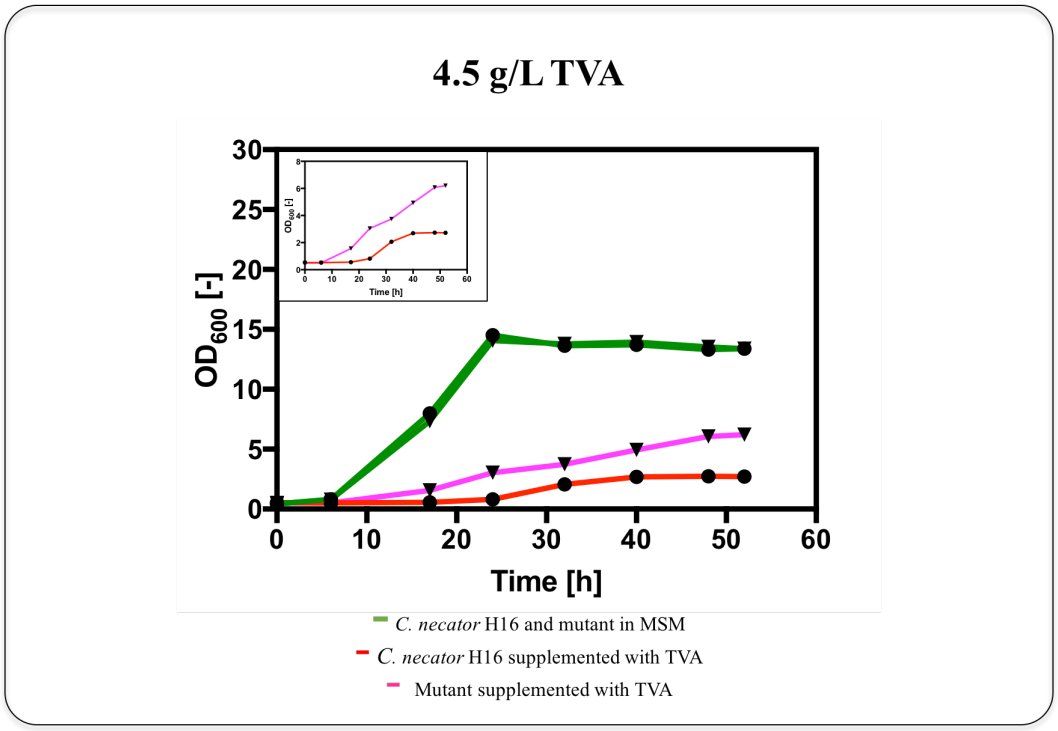


Figure 4.11 Characterisation of mutant strains mutagenised with 1.0 % (v/v) EMS. Characterisation analysis was performed in falcon tubes with 5 ml of MSM with or without TVA. C. necator H16 WT in MSM (circle, green line), and C. necator H16 WT in MSM supplemented with 4.5 g/L TVA (circle, red line); m1G9 mutant strain in MSM (triangle, green line), and m1G9 mutant strain in MSM supplemented with 4.5 g/L TVA (inverted triangle, pink line). The inset shows the magnified cell growth curve of C. necator H16 WT and m1G9 in 4.5 g/L TVA. Abbreviations: TVA, 5-(2-thienyl)-valeric acid; m, mutant.

From **Figure 4.11** it was confirmed that the cell growth of the mutant strain screened in the presence of 4.5 g/L of TVA (*m1G9*) was effectively improved compared the cell growth of *C. necator* H16 WT in TVA. On the other hand, it was observed as well that the mutant strain was also capable of growing in MSM as the WT. *C. necator* H16 showed a 94 % c.g.d. when grown in 4.5 g/L of TVA after 24 h cultivation as observed in previous MIC studies. The performance in cell growth of the mutant strain in the presence of TVA was analysed at 24 h of cultivation. The *m1G9* mutant (inverted triangle, pink line) showed to be 3.7-fold faster than the WT strain in the presence of 4.5 g/L of TVA.

4.4.3.4 *C. necator* H16 mutant library from the first round of mutagenesis

A total of 360 mutant strains were screened after the first round of EMS random mutagenesis with different EMS concentrations, where all the mutants characterised showed to have a better cell growth performance in the presence of the biotin analogue TVA than its parent strain *C. necator* H16, this could mean that these mutants have gone through mutations in specific regions that could be related to the biotin biosynthesis pathway, and these mutations allowed the mutant strains to survive under different concentrations of TVA better than the WT. **Table 4.5** shows the different concentrations of EMS used to mutagenise *C. necator* H16 WT cells, and the survival rates % obtained after mutagenesis, the c.g.d. % obtained from *C. necator* H16 WT in screening and characterisation analyses, and the mutant strains generated from the WT with its corresponding growth fold compared to the WT at 24 h cultivation in characterisation analyses.

Table 4.5 Library of mutants of the first round of mutagenesis.

Parent strain	EMS % (v/v), Survival rate %	Screening/Characterisation in [TVA] and c.g.d. % of WT	Mutant strains	Fold in characterisation
C. necator H16	0.3 % EMS, 84 %	Screening in Mp: [1.0 g/L] TVA, 11 % [3.5 g/L] TVA, 72 % Characterisation in Ft: [4.5 g/L] TVA, 96 %	*m1F2* *m1D9* *m1F5* *m1G6*	2.6 2.0 5.8 4.1
C. necator H16	0.5 % EMS, 72 %	Screening in Mp: [4.5 g/L] TVA, 85 % Characterisation in Ft: [4.5 g/L] TVA, 94 %	*m1C3*	3.0
C. necator H16	1.0 % EMS, 46 %	Screening in Mp: [4.5 g/L] TVA, 84 % Characterisation in Ft: [4.5 g/L] TVA, 94 %	*m1G9*	3.7

Abbreviations: EMS, ethyl methanesulfonate; m, mutants; TVA, 5-(2-thienyl)-valeric acid; c.g.d. %, cell growth decrease %; Mp, microtitre plates; Ft, falcon tubes.

After analysing the results obtained from the first round or mutagenesis, the best mutants were used to proceed with the second round of mutagenesis with the two mutant strains with the highest fold in growth from different EMS concentrations used, hence, the *m1F5* and *m1G9* mutant strains were selected for the second

round of mutagenesis. These new parent strains were mutagenised in the second round of mutagenesis with 0.3 % (v/v) EMS and 1.0 % (v/v) EMS respectively, as it was done in the first round of mutagenesis.

4.4.4 Genetic variation of *m1F5* and *m1G9* mutant strains obtained after EMS mutagenesis – Second Round

Once the best mutants from the first round of mutagenesis were selected, the second round of mutagenesis was performed using *m1F5* and *m1G9* mutants as parent strains. In the second round of mutagenesis 0.3 % (v/v) EMS was used to mutagenise *m1F5* cells, while 1.0 % (v/v) EMS was used to mutagenise *m1G9* cells. The procedure followed for the generation and selection of mutants was the same followed for the first round of mutagenesis (**Figure 4.5**), although this time the parent strains were *m1F5* and *m1G9* and not *C. necator* H16 WT.

The survival rate % of the second round of mutagenesis with 0.3 % (v/v) EMS for *m1F5* mutant was 83 % survival rate, and with 1.0 % (v/v) EMS for *m1G9* mutant was 48 % survival rate as observed previously in the development of the EMS random mutagenesis method. A total of 336 mutant strains from the second round of mutagenesis mutated with different concentrations of EMS were screened throughout the entire study in MSM supplemented with gentamicin and different concentrations of TVA. Each microtitre plate contained control samples to compare cell growth and find beneficial mutants, control samples were *C. necator* H16 WT in MSM, and *C. necator* H16 in MSM supplemented with TVA, and the parent strains *m1F5* and *m1G9* with and without TVA as well, which were always included in triplicates in all experiments. The best mutant strains for each of the experiments performed for the second round of random mutagenesis were chosen as follows: for each OD_{595} measurement taken during the cell growth curve, the two highest OD_{595} values corresponding to mutant strains were selected to be followed through the entire cell growth study, then, all the mutants selected were compared against each other and against the parent strain to identify the mutant strains with the best cell growth performance in the presence of TVA. The mutant strains were named as “m” for mutant, followed by a “number” from 1 to 2, which

corresponds to the round of random mutagenesis, and a "letter + number" corresponding to the well of the microtitre plate where the mutant strain was grown.

4.4.4.1 Analysis of mutant strains of *m1F5* obtained from EMS random mutagenesis with 0.3 % (v/v) EMS – Second Round

When *m1F5* cells were mutagenised with 0.3 % (v/v) EMS, the screening of the mutants was set in a microtitre plate with 4.5 g/L of TVA, where a total of 84 mutant strains were screened (**Figure 4.12**).

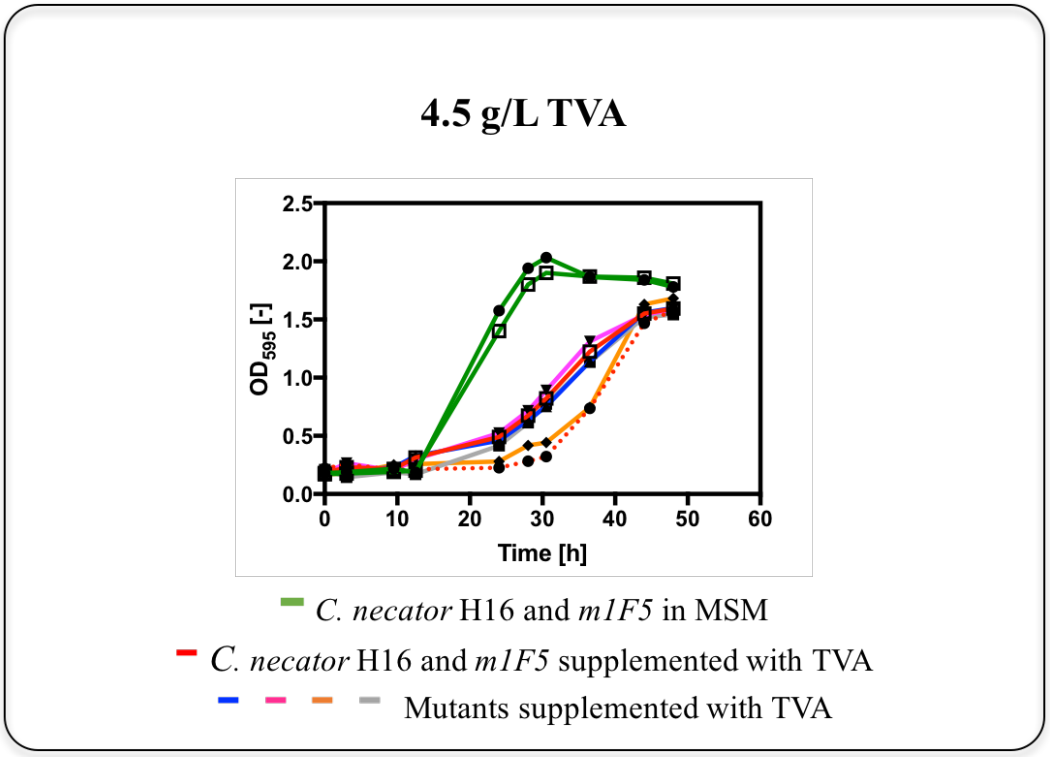


Figure 4.12 Screening of mutant strains mutagenised with 0.3 % (v/v) EMS. Screening analysis was performed in microtitre plates with 150 µL of MSM with or without TVA. C. necator H16 WT in MSM (circle, green line), m1F5 in MSM (open square, green line), C. necator H16 WT in MSM supplemented with 4.5 g/L TVA (circle, red dotted line), and m1F5 in MSM supplemented with 4.5 g/L TVA (open square, red line); m2C2 in MSM supplemented with 4.5 g/L TVA (triangle, blue line), m2C4 in MSM supplemented with 4.5 g/L TVA (inverted triangle, pink line), m2C8 in MSM supplemented with 4.5 g/L TVA (diamond, orange line), and m2H5 in MSM supplemented with 4.5 g/L TVA (square, grey line). Error bars are ± SEM of three independent determinations. Abbreviations: TVA, 5-(2-thienyl)-valeric acid; m, mutant.

The screening of this second round of random mutagenesis showed mutants with an improved phenotype in cell growth in the presence of the TVA biotin analogue. Although all the best mutants shown in **Figure 4.12** showed less cell growth decrease in the presence of TVA than the WT strain (circle, red dotted line), its improvement in cell growth was not better than its parent strain *m1F5* (open square, red line), unless for the mutant *m2C4* (inverted triangle, pink line), which showed to have an slight better cell growth performance than the parent strain *m1F5* in the presence of TVA. *C. necator* H16 showed an 86 % c.g.d. when grown in 4.5 g/L of TVA after 24 h cultivation as observed in previous MIC studies.

Since at least one of the mutants (*m2C4*) showed to have a slightly better cell growth performance in the presence of TVA than the parent strain *m1F5*, the characterisation of the best two mutants with the highest fold in growth compared to the WT was performed. The best mutants were the *m2C2* and *m2C4*, where growth was 2.0 and 2.3-fold faster compared to the WT in TVA, and only 1.1-fold faster than the parent strain *m1F5* for the mutant *m2C4*; the fold comparison was analysed at 24 h of cultivation.

After selecting the best mutants, the characterisation of *m2C2* and *m2C4* was performed in 5 mL of MSM supplemented with TVA in falcon tubes to confirm if effectively *m2C4* showed a better performance in cell growth than *m2C2*, or if they grew similarly in the presence of TVA, and to confirm also if they had a better cell growth performance than its parent strains *m1F5*. The cell growth curve for the characterisation of mutant strains is shown in **Figure 4.13**, where 4.5 g/L of TVA was used to confirm the phenotype of the mutant strains. The mutant strains were also grown in MSM without TVA to check if they could still grow like the WT in sodium gluconate.

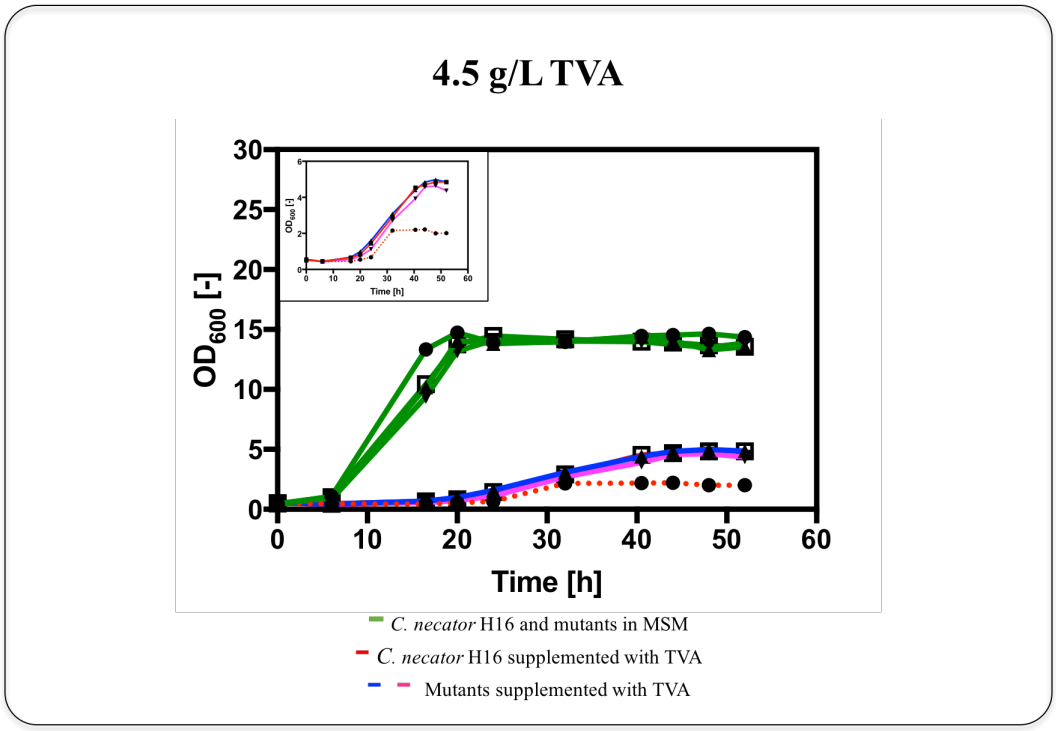


Figure 4.13 Characterisation of mutant strains mutagenised with 0.3 % (v/v) EMS. Characterisation analysis was performed in falcon tubes with 5 ml of MSM with or without TVA. C. necator H16 WT in MSM (circle, green line), m1F5 in MSM (open square, green line), C. necator H16 WT in MSM supplemented with 4.5 g/L TVA (circle, red dotted line), and m1F5 in MSM supplemented with 4.5 g/L TVA (open square, red line); m2C2 in MSM (triangle, green line), and m2C4 in MSM (inverted triangle, green line); m2C2 mutant strain in MSM supplemented with 4.5 g/L TVA (triangle, blue line), and m2C4 mutant strain in MSM supplemented with 4.5 g/L TVA (inverted triangle, pink line). The inset shows the magnified cell growth curve of C. necator H16 WT and mutant strains in 4.5 g/L TVA. Error bars are ± SEM of three independent determinations. Abbreviations: TVA, 5-(2-thienyl)-valeric acid; m, mutant.

From **Figure 4.13** it was confirmed that the cell growth of the mutant strains screened in the presence of 4.5 g/L of TVA (*m2C2,* and *m2C4*) was effectively improved compared the cell growth of *C. necator* H16 WT in TVA. On the other hand, it was observed as well that the mutant strains were also capable of growing in MSM as the WT strain. *C. necator* H16 showed a 95 % c.g.d. when grown in 4.5 g/L of TVA after 24 h cultivation as observed in previous MIC studies. The performance in cell growth of the mutant strains in the presence of TVA was analysed at 24 h of cultivation. The *m2C2* (triangle, blue line) *and m2C4* (inverted triangle, pink line) mutants showed to be 2.4 and 1.7-fold faster than the WT strain in the presence of TVA respectively, and only 1.1-fold faster than the parent strain *m1F5* for the mutant *m2C2*. Thus, the mutants obtained from this second

round of mutagenesis did not improve significantly compared to its parent strain *m1F5*; nevertheless, *m2C2* was chosen as the best mutant strain of this round for further experiments.

4.4.4.2 Analysis of mutant strains of *m1G9* obtained from EMS random mutagenesis with 1.0 % (v/v) EMS – Second Round

When *m1G9* cells were mutagenised with 1.0 % (v/v) EMS, the screening of the mutants was set in a microtitre plate with 4.5 g/L of TVA for this second round of mutagenesis, where a total of 84 mutant strains were screened (**Figure 4.14**).

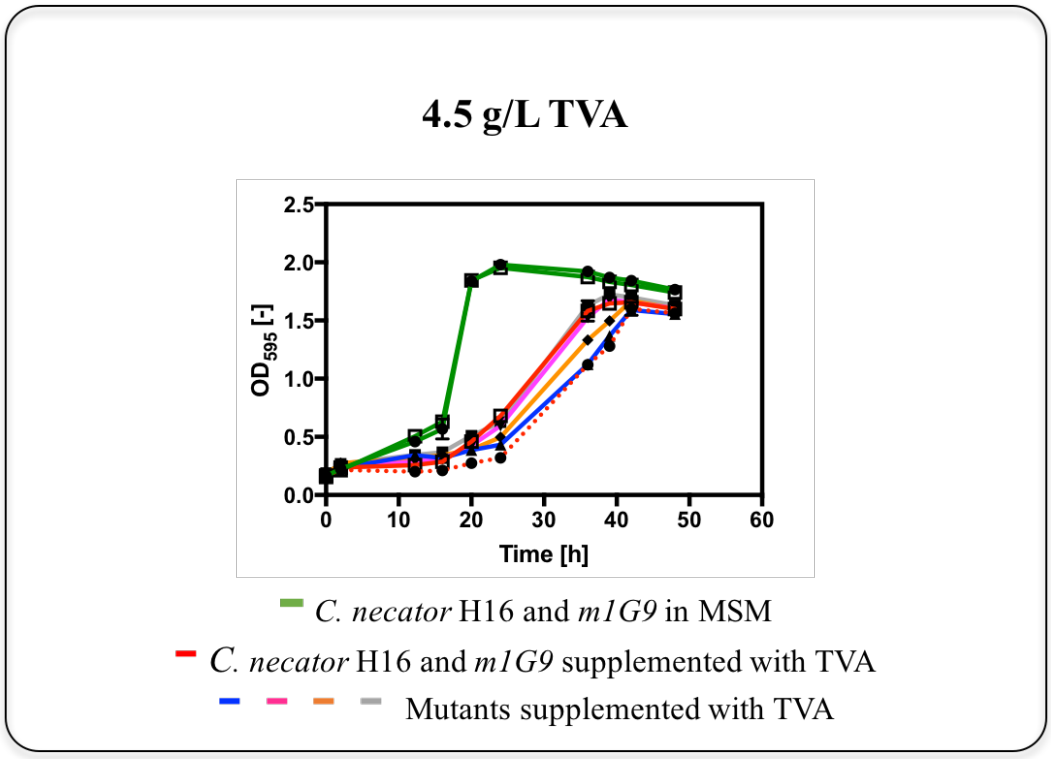


Figure 4.14 Screening of mutant strains mutagenised with 1.0 % (v/v) EMS. Screening analysis was performed in microtitre plates with 150 µL of MSM with or without TVA. C. necator H16 WT in MSM (circle, green line), m1G9 in MSM (open square, green line), C. necator H16 WT in MSM supplemented with 4.5 g/L TVA (circle, red dotted line), and m1G9 in MSM supplemented with 4.5 g/L TVA (open square, red line); m2C7 in MSM supplemented with 4.5 g/L TVA (triangle, blue line), m2E6 in MSM supplemented with 4.5 g/L TVA (inverted triangle, pink line), m2E9 in MSM supplemented with 4.5 g/L TVA (diamond, orange line), and m2G7 in MSM supplemented with 4.5 g/L TVA (square, grey line). Error bars are ± SEM of three independent determinations. Abbreviations: TVA, 5-(2-thienyl)-valeric acid; m, mutant.

The screening of this second round of random mutagenesis showed mutants with an improved phenotype in cell growth in the presence of the TVA biotin analogue. Although all the best mutants shown in **Figure 4.14** showed less cell growth decrease in the presence of TVA than the WT strain (circle, red dotted line), its improvement in cell growth was not better than its parent strain *m1G9* (open square, red line) as observed previously with *m1F5* derived mutants. *C. necator* H16 showed an 84 % c.g.d. when grown in 4.5 g/L of TVA after 24 h cultivation as observed in previous MIC studies.

Due to no mutants were found to be better than the parent strain *m1G9*, no characterisation was performed for this experiment, instead, a separate trial of second round of mutagenesis of *m1G9* with 1.0 % (v/v) EMS was performed, but the same results were found, whereby there were mutants with an improvement in cell growth in the presence of TVA compared to *C. necator* H16 WT (data not shown) but the cell growth improvement was not better than the *m1G9* parent strain; another trial was performed by modifying the selection system, which consisted on increasing throughput for higher probability of identifying rare mutants by plating mutated cells in MSM agar plates supplemented with different concentrations of TVA rather than plating mutated cells in MSM agar plates without the selection pressure. The incubation of agar plates was left for 24 h more than when cells had been plated in agar plates without TVA (where only 48 h was sufficient to observe growth of colonies in agar plates) (**Figure 4.15**), but again, no better mutants than *m1G9* were found after screening analyses even at longer incubation times.

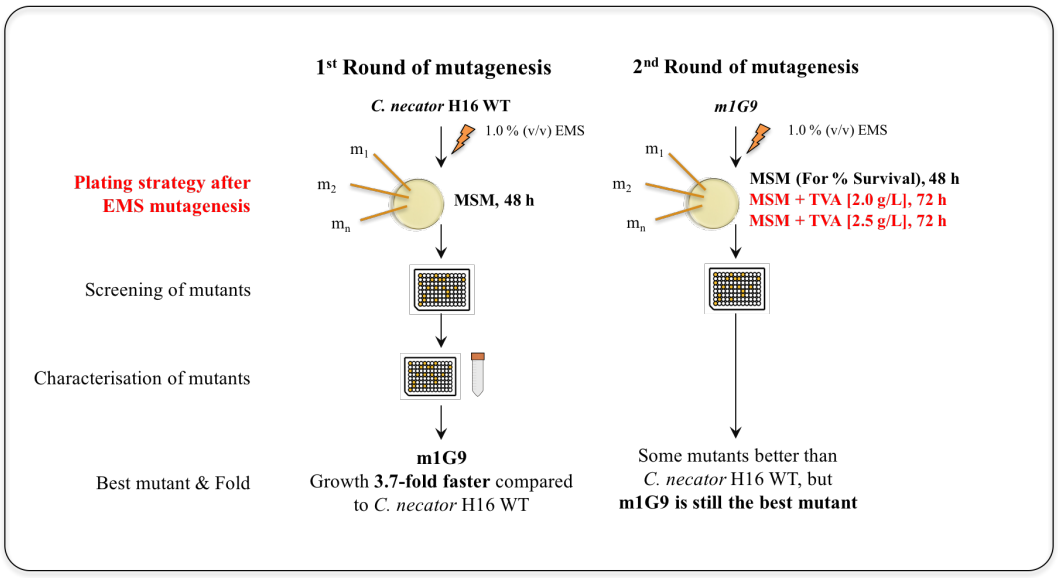


Figure 4.15 Diagram of improvements in the selection system to find rare mutants. The diagram highlights in red the main differences in the selection system on how mutants were plated after EMS mutagenesis in the first and the second round of mutagenesis. In order to select mutants with a better cell growth performance in TVA than m1G9, the mutagenised m1G9 was plated in MSM agar plates supplemented with 2.0 g/L and 2.5 g/L of TVA after EMS treatment in order to increase throughput for a higher probability of identifying rare mutants. Abbreviations: TVA, 5-(2-thienyl)-valeric acid; m, mutant.

4.4.4.3 *C. necator* H16 mutant library from the second round of mutagenesis

A total of 336 mutant strains were screened after the second round of EMS random mutagenesis with different EMS concentrations, where not all the mutants characterised showed to have a better cell growth performance in the presence of the biotin analogue TVA than its parent strains isolated from the first round of mutagenesis, even when improvements in the selection system shown in **Figure 4.15** were performed. **Table 4.6** shows the different concentrations of EMS used to mutagenise *m1F5* and *m1G9* mutant strains, and the survival rates % obtained after mutagenesis, the c.g.d. % obtained from *C. necator* H16 WT in screening and characterisation analyses, and the mutant strains of the second round of mutagenesis with its corresponding growth fold compared to the WT at 24 h cultivation in characterisation analyses.

Table 4.6 Library of mutants of the second round of mutagenesis.

Parent strain	EMS % (v/v), Survival rate %	Screening/Characterisation in [TVA] and c.g.d. % of WT	Mutant strains	Fold in characteri-sation
m1F5	0.3 % EMS, 83 %	Screening in Mp: [4.5 g/L] TVA, 83 % Characterisation in Ft: [4.5 g/L] TVA, 95 %	*m2C2* *m2C4*	2.4 1.7
m1G9	1.0 % EMS, 48 %	Screening in Mp: [4.5 g/L] TVA, 84 %	NA	NA

Abbreviations: EMS, ethyl methanesulfonate; m, mutants; TVA, 5-(2-thienyl)-valeric acid; c.g.d. %, cell growth decrease %; Mp, microtitre plates; Ft, falcon tubes; NA, not applicable.

4.4.5 Determination of biotin content in *C. necator* H16 and mutant strains

In order to corroborate if effectively the improvement in cell growth in the presence of the TVA biotin analogue was due to an overproduction of biotin by the mutant strains, the biotin produced in *C. necator* H16 WT strain was compared to the biotin produced in some of the mutant strains previously isolated from the first and second round of mutagenesis, where it was expected that mutant strains would produce higher quantities of biotin since they had been able to grow better than the WT in the presence of the biotin analogue. The mutants isolated from previous rounds of EMS random mutagenesis are shown in **Figure 4.16**.

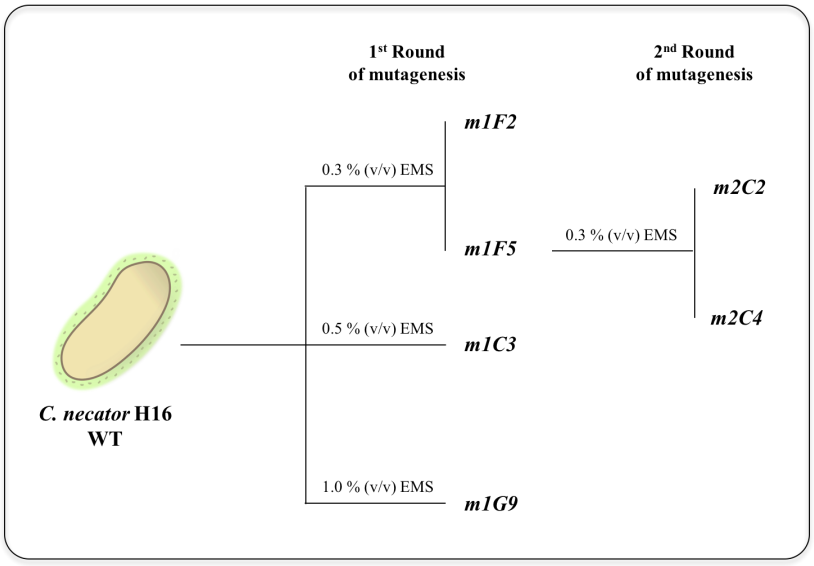


Figure 4.16 Library of mutants generated by directed evolution via EMS random mutagenesis from the first and second round of mutagenesis.

The determination of the biotin content was performed with a FluoReporter Biotin Quantitation Assay Kit, which is a sensitive fluorometric assay for biotin quantitation. A biocytin (biotinylated lysine) standard curve was generated in order to be able to determine the biotin content of cells. Triplicates of each of the samples for the standard curve as well as for cell samples were set in order to determine the biotin content. Once the standard curve was generated, only the mutants with the highest fold in growth compared to the WT were chosen to determine their biotin content: *m1F5*, *m1C3*, *m1G9*, and *m2C2*.

Initially, the biotin content of *C. nec*ator H16 and mutant strains was quantified when cells were grown only in MSM (**Figure 4.17A**), where cells were cultivated in MSM for 40 h before an aliquot of the culture was withdrawn for biotin quantitation assays. After analysing biotin content, from **Figure 4.17A**, it could be observed that the difference in biotin production between the WT and the mutants was not very significant, where the biotin production was only 1.1-fold higher for *m1C3* and *m1G9* compared to the biotin production of the WT strain; due to the biotin production among the WT and the mutant strains was not as high as expected, a further experiment was designed, where the determination of biotin content was analysed when cells were grown in MSM and the selection pressure (MSM supplemented with 2.5 g/L of TVA) (**Figure 4.17B**), the rationale behind this experimental design was that probably the mutant strains would only produce higher amounts of biotin when they were grown under the selection pressure. Nevertheless, the results of the determination of the biotin content when cells were grown in MSM supplemented with TVA were very similar to the results obtained when cells were grown only in MSM, whereby the biotin production was very similar among all samples (WT and mutants), where the highest fold obtained was 1.1 for the *m2C2* compared to the WT.

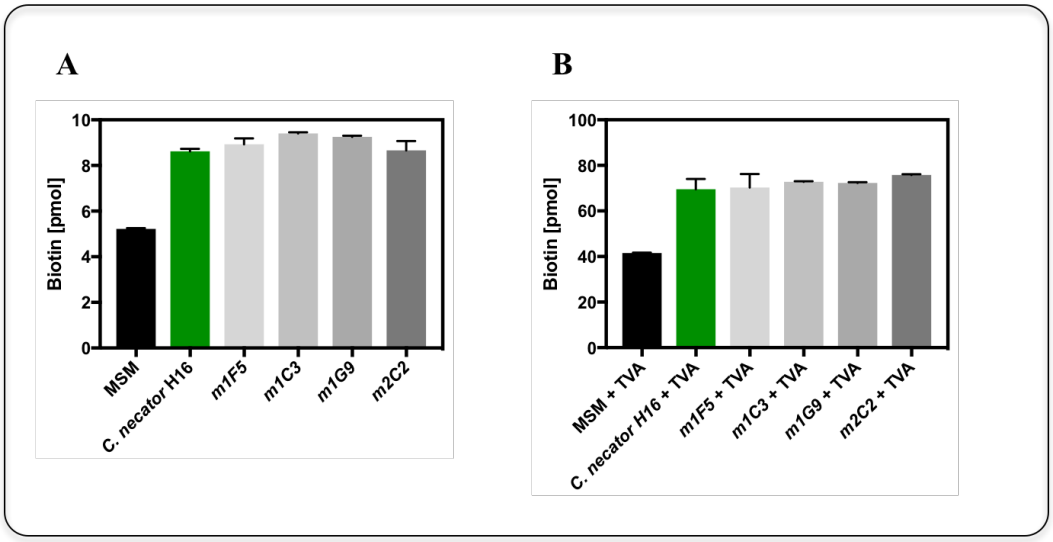


Figure 4.17 Determination of biotin content in C. necator H16 and mutant strains. ***A)*** *Biotin quantitation of samples grown only in MSM, where control samples were MSM medium (no cells) (black column), and C. necator H16 WT strain in MSM (green column).* ***B)*** *Biotin quantitation of samples grown in MSM supplemented with 2.5 g/L TVA, where control samples were MSM medium supplemented with 2.5 g/L TVA (no cells) (black column), and C. necator H16 WT strain in MSM supplemented with 2.5 g/L TVA (green column). Error bars are ± SEM of three independent determinations. Abbreviations: TVA, 5-(2-thienyl)-valeric acid; m, mutant.*

As observed in **Figure 4.17B**, the concentration of biotin was higher by ≈10-fold when cells were grown in the presence of TVA than when cells were grown only in MSM, this might be due to TVA generates a high background, as can be observed with the control sample which only contained MSM medium supplemented with TVA (no cells) (black column). Nevertheless, the outcome was the same than the one obtained in cells grown without TVA, where no significant difference in biotin production among the WT and the mutant strains was found.

The fact that the biotin quantitation results demonstrated that the biotin production was about the same in the WT strain and the mutant strains raise the question on whether TVA is leading to find mutants that are related with the biotin biosynthesis pathway of *C. necator* H16 or not. It is possible that TVA might not be leading to finding mutants that are related with the genes of the biotin biosynthesis but to find mutants that have mutations in other non-biotin related genes, and hence the mutants do not show a higher biotin production.

Also, it could be possible that there is no significant increase in biotin production in the mutant strains compared to the WT due to the difference in cell growth in the presence of TVA between the WT and the mutant strains isolated from the first and second round of mutagenesis is still not significant enough to show a higher biotin production. Further studies would be required to find mutants with a higher fold in growth in TVA compared to the WT, to confirm if a higher fold in growth in the presence of TVA would provide a significant increase in biotin content between the mutants and the WT strain. On the other hand, a different selection pressure, such as the biotin analogues described in Error! Reference source not found., could be tested also to find rare mutants.

4.5 Conclusions

A method of EMS random mutagenesis for *C. necator* H16 was developed and optimised in this **Chapter 4**, which provided a survival rate % from ~ 50 to 85 %. This method represents a powerful tool for *C. necator* H16, since no EMS random mutagenesis method has been reported to date for the strain. This tool can be used for engineering *C. necator* H16 for a wide variety of purposes, as along with other combinatorial approaches, directed evolution *via* random mutagenesis is becoming increasingly important for metabolic engineering of complex phenotypes.

In order to explore the applicability of directed evolution *via* random mutagenesis for *C. necator* H16, a library of mutants was generated to try to understand the biotin pathway in *C. necator* H16. The chemical mutagen EMS was used for random mutagenesis; the mutagen was used to speed up the selection process of *C. necator* H16 mutants that were expected to overproduce biotin in the presence of the biotin analogue TVA. A total of 696 TVA-resistant mutants of *C. necator* H16 were isolated and screened, from which 8 mutants were further characterised at different concentrations of TVA. All these 8 mutants showed an improved performance in cell growth in the presence of TVA, these mutants were obtained either in the first or the second round of EMS random mutagenesis, being *m1F5* (from the first round of mutagenesis, mutated with 0.3 % (v/v) EMS) the mutant that showed the highest cell growth fold compared to the WT (5.8-fold faster than the WT after 24 h cultivation). Nevertheless, when the mutants isolated from the first and second round of mutagenesis were analysed for biotin quantitation, studies showed no improvement in biotin production compared to *C. necator* H16 WT.

The results of the determination of biotin content could suggest that a higher fold in cell growth must be achieved (*e.g.*, a mutant that grows at least 10-fold faster than the WT strain) in order to observe a higher difference in biotin production between the WT and the mutant strains. Nevertheless, the optimisation of the EMS random mutagenesis method developed in this study for *C. necator* H16

proved to be a useful tool, as mutants with an improved phenotype to grow in the presence of TVA were obtained after directed evolution *via* random mutagenesis. This method is a potential tool that could be used for exploring other unknown metabolic pathways of *C. necator* H16, or to continue exploring the biotin biosynthesis pathway of *C. necator* H16 with a different biotin analogue.

CHAPTER 5

Rational design and Evolutionary engineering:

Engineering *Cupriavidus necator* H16 to convert waste stream into useful chemicals by enhancing glycerol assimilation

Abstract

Cupriavidus necator H16 has a versatile metabolism that can utilise a wide-range range of carbon sources, including renewable carbon sources that can be used for chemical production. Glycerol (crude glycerol) is a by-product mainly obtained from biodiesel production and fat splitting processes, and, although *C. necator* H16 can utilise glycerol to produce different bioproducts such as PHB, the growth of the wild-type strain of *C. necator* H16 in glycerol is very slow compared to other preferable carbon sources such as gluconate, this fact limits the use of glycerol as a feedstock for chemical production. In this **Chapter 5**, the knowledge and strategies studied in previous chapters and proven to work in *C. necator* H16 supported the approaches selected to construct recombinant and evolved strains of *C. necator* H16 for chemical production (bioplastics) using waste stream (crude glycerol) as a feedstock. This chapter compares rational design and evolutionary engineering approaches for unlocking the glycerol. Specifically, expression of glycerol utilisation enzymes native to *C. necator* H16 or from *E. coli*, and adaptive evolution were investigated. Both strategies were successful and the best *C. necator* H16 variant showed a 8-fold improvement in growth compared to the wild type strain, where the best variant was derived from adaptive evolution. Worthy of note, this best variant also demonstrated superior growth in crude glycerol (sweetwater) obtained from an industrial fat splitting process.

5 Chapter 5. Rational design and Evolutionary engineering: Engineering *Cupriavidus necator* H16 to convert waste stream into useful chemicals by enhancing glycerol assimilation

5.1 Introduction

The exploration and optimisation of rational design and evolutionary engineering tools for *C. necator* H16 investigated in previous chapters are useful tools that can be used to engineer the strain by introducing and tuning metabolic pathways for industrial purposes. The work presented in this **Chapter 5** aims to prove the previous statement by using *C. necator* H16 as cell factory to convert waste stream into useful chemicals such as polyhydorxybutyrate (PHB), which is valuable plastic that has been widely produced in *C. necator* H16 as the strain is a natural PHB producer.

As mentioned in **Section 1.4.3**, crude glycerol –a by-product of the biodiesel industry- is sometimes even considered a waste stream. This waste stream could be used as feedstock and be converted into a valuable co-product for the biodiesel supply chain. In this work, *C. necator* H16 was used as a cell factory to convert this waste stream into a value-added product (PHB, bioplastics). As reviewed in He *et al*. (2017), there are different value-added products such as PHB, 1,3-propanediol, short-chain organic acids, alcohols, among others that are very promising bioproducts that can be obtained from microbial fermentation with crude glycerol.

C. necator H16 can naturally utilise glycerol as a carbon source, and the engineering tools required to tune its metabolic pathway have been optimised in previous chapters of this thesis, and are available to engineer the strain to improve its glycerol utilisation.

The engineering of *C. necator* H16 WT to improve glycerol assimilation is essential, since the WT strain grows slower in glycerol than in other preferred

carbon substrates such as gluconate, hence, in this work, two parallel strategies - rational design and evolutionary engineering- studied and described in previous chapters were used to improve the productivity levels of *C. necator* H16 for a large-scale production using glycerol as its carbon feedstock. The first strategy consisted of expressing native glycerol utilisation enzymes, and the second strategy consisted of evolving *C. necator* H16 wild-type cells in glycerol *via* adaptive evolution. The rational design strategy was used based on the improvement of glycerol utilisation by increasing the intracellular levels of glycerol kinase (GlpK) and glycerol dehydrogenase (GlpD) - both native of the strain-, since these proteins were postulated to participate in *C. necator* H16 glycerol metabolism (Fukui *et al*., 2014); while the adaptive evolution strategy was used due to the underlying reason of slow growth of *C. necator* H16 is not fully understood, and this strategy circumvents this problem by exploiting the linkage between carbon (glycerol) utilisation and survival (growth).

Initially, the studies for the enhancement of glycerol assimilation in *C. necator* H16 were performed in laboratory-grade glycerol (pure glycerol), and once the improvements in pure glycerol assimilation were achieved in recombinant strains of *C. necator* H16 -from the rational design engineering strategy-, and variants strains of *C. necator* H16 -from the evolutionary engineering strategy-, the performance of the above-mentioned strains was verified in crude glycerol (also referred to as sweetwater), which was sourced from a fat splitting process.

5.2 Aim

The aim of **Chapter 5** is:

- To apply rational design and evolutionary engineering tools explored and optimised in this PhD work to engineer *C. necator* H16 to enhance its glycerol assimilation for chemical production using waste stream (crude glycerol) as feedstock.

5.3 Materials and methods

5.3.1 Materials

List of detailed information on the materials used (reagents, kits, equipment, software, media preparation and miscellaneous) can be found in **Appendix 1 – 6**.

5.3.2 Strains, plasmids and primers

All strains, plasmids and primers that were used in this **Chapter 5** are shown in **Table 5.1**.

Table 5.1 Strains, plasmids and primers used in this study.

Strains, plasmids or primers	Description	References or source
Strains		
E. coli DH5α	Standard cloning strain	Lab collection
E. coli BW25113	Wild-type Kan^r	Datsenko *et al*., 2000
C. necator H16	Wild-type Gen^r	DSM 428
v6C6	Variant of *C. necator* H16 from Rd 6 of adaptive evolution	This work
Plasmid backbones		
pPj5c.rbs-RFP	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, P_{j5}, RBS, *rfp*	This work
pBADc.rbs-RFP	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, *araC*, P_{BAD}, RBS, *rfp*	This work
Plasmids constructed		
pPj5'c.rbs- $glpK_{H16}$	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, P_{j5}, RBS, $glpK_{H16}$	This work
pPj5'c.rbs- $glpK_{Ec}$	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, P_{j5}, RBS, $glpK_{Ec}$	
pPj5'c.rbs- $glpD_{H16}$	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, P_{j5}, RBS, $glpD_{H16}$	This work
pPj5'c.rbs1- $glpKD_{H16}$	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, P_{j5}, RBS, $glpKD_{H16}$, native $glpD_{H16}$ RBS preceding $glpD_{H16}$	This work
pPj5'c.rbs2- $glpKD_{H16}$	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, P_{j5}, RBS, $glpKD_{H16}$, synthetic RBS preceding $glpD_{H16}$	This work
pBADc.rbs-$glpK_{H16}$	Cam^r, pBBR1 Rep, pBBR1 oriV, *mob*, *araC*, P_{BAD}, RBS, $glpK_{H16}$	This work
Primers (5' → 3')		
NdeI-H16-A2507-fwd	GGAATTCCATATGAACCAGCCAGCCGCGCAG	This work
XhoI-H16-A2507-rev	AACGCTCGAGTTATCAGGCGTGGCTGCCGCTG	This work
BW25113_glpk-fwd	GCTCCATATGACTGAAAAAAAATATATCG	This work
BW25113_glpk-rev	TATAATCCTCGAGTTATTCGTCGTGTTC	This work
NdeI-glpD-fwd	GGAATTCATATGCAGAGAGTCTCC	This work
XhoI-glpD-rev	TTAACTCGAGTTATCAGCCGAGCATGTA	This work
NdeI-glpK-fwd	GGAATTCCATATGAACCAGCCAGCC	This work

RBS-glpD-fwd	AGATCTTTTAAGAAGGAGATATACATATGCAG AGAGTCTC	
HiFi-fwd	TACATGCTCGGCTGATAACTCGAGTAAGGAT	This work
Hifi-rev	CTTCTTAAAAGATCTTCAGGCGTGGCTGCC	This work

All the RBS noted in all the plasmids in the list correspond to the sequence encoding for the *E. coli* consensus RBS. Restriction enzyme sequences of primers are underlined. Abbreviations: Kan^r, kanamycin resistance; Gen^r, gentamicin resistance; Rd, round; Cam^r, chloramphenicol resistance; RBS, ribosome binding site; *rfp*; red fluorescence protein gene; $glpK_{H16}$, glycerol kinase from *C. necator* H16; $glpK_{Ec}$, glycerol kinase from *E. coli* BW25113; $glpD_{H16}$, glycerol dehydrogenase from *C. necator* H16.

5.3.3 Cultivation of *E. coli* DH5α and *C. necator* H16

E. coli DH5α and *E. coli* BW25113 cells were cultivated at 37 ºC on 2x YT medium and 250 rpm and when required with chloramphenicol [25 μg/mL] or kanamycin [50 μg/mL] according to the application. *E. coli* DH5α was used for all molecular cloning, plasmid propagation, and maintenance; and *E. coli* BW25113 was used for genomic DNA extraction. *C. necator* H16 WT and *C. necator* H16 recombinant and variant strains cells were cultivated at 30 ºC using nutrient broth (NB), or mineral salts medium (MSM) with different carbon sources depending on the experiment performed: MSM with 1.00 % or 0.59 % (w/v) sodium gluconate (for practical purposes, sodium gluconate will be referred to as gluconate from this point onwards), 0.50 % (w/v) laboratory-grade glycerol with purity over 99 % (for practical purposes, laboratory-grade glycerol will be referred to as glycerol from this point onwards), or 4.00 % (v/v) crude glycerol, and 250 rpm supplemented always with gentamicin [10 μg/mL] and when needed with chloramphenicol [25 μg/mL]. Crude glycerol (sweetwater) sample (batch 8/1/18), was obtained from a high-pressure fat splitting process, kindly provided by Croda (Hull). The optical density was measured at 600 nm with a BioPhotometer or at 595 nm with a microtitre plate reader.

For growth comparison, a pre-culture of *C. necator* H16 WT strain, recombinant strain or variant strain, was prepared from a single colony in an agar plate of MSM with 1.00 % (w/v) gluconate, which was further inoculated in 5 mL of MSM with either gluconate, glycerol or crude glycerol at different concentrations, always with a starting OD_{600} of 0.1 using the pre-culture. Cells were cultivated at

30°C, and their optical density monitored at 600 nm with a BioPhotometer or at 595 nm with a microtitre plate reader. Specific growth rate was calculated by fitting the exponential growth to the exponential growth equation ($X = X_o e^{\mu t}$) provided in GraphPad Prism, where X is the number or mass of cells (mass/volume), X_0 the initial number or mass of cells, t is time, and μ is the specific growth rate constant (1/time).

5.3.4 Bacterial transformation of *E. coli* DH5α and *C. necator* H16

*E. coli DH5*α cells were transformed using a standard chemical transformation method. *E. coli* DH5α was used for circularization and propagation of ligation mixtures of all plasmids constructed. *C. necator* H16 cells were transformed by the electroporation method optimised in our laboratory group and described in **Chapter 2**.

5.3.4.1 Transformation using $CaCl_2$ heat-shock method for *E. coli* DH5α

A pre-culture of *E. coli* DH5α was cultivated in 5 mL of 2x YT for 16 h at 37 ºC; for competent cell preparation, a fresh 2x YT falcon tube was inoculated with the pre-culture at a 1:100 dilution and cultivated at 37 ºC, when cells reached an optical density at 600 nm (OD_{600}) of 0.5, 1 mL of cells were transferred to a sterile 1.5 mL microcentrifuge tube. Cells were centrifuged at maximum speed (17,000 x g) for 30 s, and the supernatant was removed by pipetting. The pellet was washed once with 500 μL of pre-chilled 50 mM $CaCl_2$. Then, the cell pellet was resuspended carefully in 500 μL of pre-chilled 50 mM $CaCl_2$ and incubated in ice for 30 min (for transforming intact plasmid, a 10 min incubation was sufficient). Plasmid DNA was added to a concentration of 1 μg or 5 μL of ligation mixture and mixed gently. A second 30 min incubation step was performed after the addition of the plasmid (for transforming intact plasmid, a 10 min incubation was sufficient). After the second incubation, the cells were heat-shocked at 42 ºC for 1

min and further incubated in ice for 2 min. After heat shock and ice incubation time, 800 µL of 2x YT was added, and cells were left to grow for a 1 h outgrowth at 37 ºC. After the outgrowth, cells were centrifuged at maximum speed for 30 s and most of the media removed, the remaining 200 – 300 µL media was used to resuspend the cells gently before plating them on TYE agar plates with the required antibiotic, and incubated overnight at 37 ºC.

5.3.4.2 Transformation by electroporation method for *C. necator* H16

A pre-culture of *C. necator* H16 was cultivated in 5 mL of NB with gentamicin for 40 – 44 h at 30 ºC; for electrocompetent cell preparation, a fresh NB supplemented with gentamicin falcon tube was inoculated with the pre-culture at a 1:50 dilution and cultivated at 30 ºC, when cells reached an OD_{600} of 0.5 – 0.7, cells were transferred to ice and chilled for 5 min. Two millilitres of the cells were then transferred to a sterile 2 mL microcentrifuge tube. Cells were centrifuged at maximum speed (17,000 x g) for 30 s and the supernatant was removed. Then, the cell pellet was washed once with 1 mL of pre-chilled 50 mM $CaCl_2$ and incubated for 15 min in ice. Then the cell pellet was centrifuged at maximum speed for 30 s followed by two washes by resuspension in 1 mL of pre-chilled 0.2 M sucrose. After the final wash, the cell pellet was resuspended in 100 µL of pre-chilled 0.2 M sucrose. Plasmid DNA was added to a concentration of 0.5 µg to the resuspended cells and mixed gently. The resuspension was then transferred into a pre-chilled 2 mm electroporation cuvette and electroporated at 2.3 kV. After electroporation, 1 mL of NB was added immediately directly to the electroporation cuvette, then, cells were transferred to a new 2 mL microcentrifuge tube for a 2 h outgrowth at 30 ºC. After the outgrowth, typically about 10 % (v/v) cells were plated in NB with gentamicin and chloramphenicol agar plates and incubated at 30 ºC for 40 – 48 h.

5.3.5 Colony PCR verification of transformants in *C. necator* H16

All plasmids constructed and transformed into *C. necator* H16 were verified by colony PCR using Q5 polymerase according to the manufacturer's protocol. Colonies were picked from NB agar plates with a pipette tip and then were swirled vigorously in 20 μL of nuclease-free water in 1.5 mL microcentrifuge tubes. The liquid suspension was used as DNA template for colony PCR. Primers specific to the insert present in the plasmids constructed were used for colony PCR verification. The PCR product was then analysed using a 1.0 % (w/v) agarose gel electrophoresis.

5.3.6 DNA preparation

Standard procedures were used for isolation of plasmids, restriction enzyme digestions, polymerase chain reaction (PCR) with Q5 DNA polymerase, *DpnI* digestions, DNA gel extraction, PCR purification, and T4 DNA ligation (Sambrook *et al*., 2011) and recommendations by the manufacturers. All primers were synthesised by Eurofins Genomics.

5.3.7 DNA Gel electrophoresis

An agarose gel electrophoresis was prepared to analyse DNA; 0.7 % or 1.0 % (w/v) agarose gels were prepared in 1x TBE buffer (prepared from a 5x TBE buffer composed of 54.0 g of Tris base, 27.5 g of Boric acid, and 20 mL of 0.5 M EDTA per 1 L, pH 8.3) by dissolving 0.35 g or 0.50 g of agarose respectively in 50 mL buffer. The percentage of gel used was dependent on the purpose of the gel, if only an analysis of the DNA was required, then 1.0 % (w/v) gels were used, but if gel extraction was required, then 0.7 % (w/v) agarose gels were used to excise the DNA band from the gel. To ensure that the agarose is fully dissolved in the buffer, the solution was microwaved until all the agarose was dissolved, then, when the gel had cooled down, 2 μL of ethidium bromide was added and the gel

was cast using a gel caster, then the comb was inserted and the gel was left to room temperature until it solidified. 6 μL of DNA Ladder 1 kb and appropriate volume of DNA samples were loaded into the gel. The electrophoresis was run at a constant voltage of 100 V for 60 min. The gel image was captured with a gel documentation system.

5.3.8 Plasmid construction

All plasmid constructed in this **Chapter 5** were based on the plasmid backbones pBADc.rbs-RFP and pPj5c.rbs-RFP plasmids. The plasmid backbones are illustrated in **Figure 3.1** and **Figure 5.3** respectively. The primers and *in silico* analyses of plasmid construction were generated using the SnapGene software tool. Correct construction of all plasmids was confirmed by restriction enzyme analysis (ReA) and DNA gene sequencing by Eurofins Genomics.

5.3.8.1 Construction of plasmids with a constitutive promoter

The constitutive plasmids driven by the P_{j5} promoter were constructed by first extracting total genomic DNA from *E. coli* BW25113 and *C. necator* H16, where the $glpK_{Ec}$, $glpK_{H16}$, $glpD_{H16}$, and $glpKD_{H16}$ genes were amplified from their respective genomic DNA (where "$_{Ec}$" stands for *E. coli* BW25113, and "$_{H16}$" for *C. necator* H16) by PCR with the primers BW25113_glpk-fwd and BW25113_glpk-rev for plasmid pPj5c.rbs-$glpK_{Ec}$, NdeI-H16-A2507-fwd and XhoI-H16-A2507-rev for plasmid pPj5c.rbs-$glpK_{H16}$, NdeI-glpD-fwd and XhoI-glpD-rev for plasmid pPj5c.rbs-$glpD_{H16}$, and NdeI-glpK-fwd and XhoI-glpD-rev for plasmid pPj5c.rbs1-$glpKD_{H16}$. Subsequently, the amplified fragments were subjected to *DpnI* digestion, PCR purification and restriction enzyme digestion with *NdeI* and *XhoI*, followed by a further PCR purification for ligation with T4 DNA ligase with the *NdeI* and *XhoI* digested and gel extracted fragment containing the pBBR1 Rep, pBBR1 oriV origin of replication, *mob* gene, Cam^r cassette, P_{j5} promoter, and *E. coli* consensus RBS from pPj5c.rbs-RFP. The plasmid pPj5c.rbs-RFP was constructed based on the pPj5c.T7rbs-RFP from the synthetic biology toolbox described in **Chapter 2**, where the T7 mRNA stem-

loop was removed from pPj5c.T7rbs-RFP by site-directed mutagenesis using the NEBaseChanger protocol.

The standard restrictive digestion and ligation were used for all plasmids unless for pPj5c.rbs2-glpKD$_{H16}$, which was constructed using the HiFi DNA Assembly cloning method with primers RBS-glpD-fwd, HiFi-fwd and HiFi-rev. After DNA sequencing, mutations in the P_{j5} promoter region of pPj5c.rbs-glpK$_{H16}$ were found, for this reason, all genes were subsequently re-cloned to have the same mutated $P_{j5'}$ promoter for a fair comparison (**Table 5.2**), resulting in the plasmids pPj5'c.rbs-glpK$_{Ec}$, pPj5'c.rbs-glpK$_{H16}$, pPj5'c.rbs-glpD$_{H16}$, pPj5'c.rbs1-glpKD$_{H16}$, and pPj5'c.rbs2-glpKD$_{H16}$. After transformation of *E. coli* DH5α, a single colony of the resulting plasmids was isolated and analysed by restriction enzyme analysis and DNA sequencing to confirm the construct before they were transformed into *C. necator* H16.

Table 5.2 Comparison of P_{j5} and $P_{j5'}$ promoter sequences.

Promoter	Sequence (5' → 3')
P_{j5}	agcggatataaaaaccgttattgacacaggtggaaatttagaatatact**t**gttagtaaacctaatggatcgacc**tt**
$P_{j5'}$	agcggatataaaaaccgttattgacacaggtggaaatttagaatatac**g**gttagtaaacctaatggatcgacc**t**

Note: The nucleotides in red, highlight the differences between the 2 promoters.

5.3.8.2 Construction of plasmid with an inducible promoter

The inducible pBADc.rbs-glpK$_{H16}$ was constructed by first extracting total genomic DNA from *C. necator* H16, where the *glpK* gene was amplified from genomic DNA by PCR with NdeI-H16-A2507-fwd (forward primer) and XhoI-H16-A2507-rev (reverse primer) primers. Subsequently, the amplified fragment was subjected to *DpnI* digestion, PCR purification and restriction enzyme digestion with *NdeI* and *XhoI*, followed by a further PCR purification for ligation with T4 DNA ligase with the *NdeI* and *XhoI* digested and gel extracted fragment containing the pBBR1 Rep, pBBR1 oriV origin of replication, *mob* gene, Cam^r cassette, L-arabinose-inducible system (containing *araC* and P_{BAD}), and *E. coli* consensus RBS from pBADc.rbs-RFP resulting in the 7.6 kb plasmid pBADc.rbs-

$glpK_{H16}$. After transformation of *E. coli* DH5α, a single colony of the resulting plasmid was isolated and analysed by restriction enzyme analysis and DNA sequencing to confirm the construct before it was transformed into *C. necator* H16.

5.3.8.3 Gene Sequencing Analysis

Eurofins Genomics sequencing service was used to check the sequence quality of the plasmids and GlpK proteins of *C. necator* H16 WT and *C. necator* H16 variant strains (*v6C6*, *v6F8*, and *v6G7*) from the Round 6 of adaptive evolution. The DNA fragments of GlpK and its 500-bp upstream element were amplified by PCR with Q5 polymerase from genomic DNA of their corresponding strain in order to check for gene modifications in that region. The DNA sequences were then analysed using the SnapGene software tool.

5.3.9 Adaptive evolution of *C. necator* H16

Adaptive evolution of *C. necator* H16 was performed by serial cultivation in 5 mL of MSM with 0.50 % (w/v) glycerol as a sole carbon source in falcon tubes for six rounds of adaptive evolution. In each round, cells were grown to their early stationary phase before using them to inoculate the subsequent round with a 1:100 dilution. Serial cultivation was continued until there was no significant improvement in glycerol assimilation (sixth round). An aliquot of 0.5 mL was withdrawn from each round of adaptive evolution and stored as glycerol stock of the cell population of each round at -80°C for further analysis.

5.3.10 Screening of *C. necator* H16 variant strains

After six rounds of adaptive evolution, the population of variants of *C. necator* H16 from the last three rounds of adaptive evolution were screened for isolation of single variants from the population. The population of the last three rounds was

grown in 5 mL of MSM with 0.50 % (w/v) glycerol and gentamicin [10 μg/mL] to an OD_{600} of 1.0 before cells were plated on MSM agar plate with the same glycerol and gentamicin concentrations. Agar plates were incubated at 30°C for 48 h before 90 single colonies from each round were transferred to 96-well microtitre plates and screened in 150 μL of MSM with 0.50 % (w/v) glycerol. The growth of the variant strains was monitored until cells reached stationary phase at optical density was measured at 595 nm with a microtitre plate reader.

5.3.11 Characterisation of *C. necator* H16 variants in glycerol

A pre-culture of *C. necator* H16 variant strains was cultivated in 5 mL of MSM with gentamicin for 40 – 44 h at 30 ºC, then, the characterisation of variants of *C. necator* H16 was performed in 150 μL of MSM with 0.50 %, 1.00 % and 2.00 % (w/v) glycerol in microtitre plates with a 1:100 dilution. The growth of the variant strains was monitored until cells reached stationary phase at optical density was measured at 595 nm with a microtitre plate reader. The inoculum of the variants was always adjusted to an initial OD_{600} of 0.1.

5.3.12 Confirmation of improved glycerol-utilising phenotype

Improved glycerol-utilising phenotype in the identified variants was further verified by five rounds of cultivation in the absence of glycerol, in either MSM with 1.00 % (w/v) gluconate or NB. In each of these rounds, cells were cultivated for 48 h at 30°C. Cells from the fifth round were re-transferred into MSM with 0.50 % (w/v) glycerol and gentamicin [10 μg/mL]. True variants were expected to grow as quickly as when they were first isolated, in contrast to those that exhibited transient adaptation.

5.3.13 Protein model of GlpK of *C. necator* H16 and *v6C6*

The protein models of GlpK of *C. necator* H16 WT and *v6C6* variant strain were generated using SWISS-MODEL (Biasini *et al*., 2014), where the $GlpK_{Ec}$ from *E.*

coli (PDB code 1BOT) was used as a template protein. The graphics were generated using PyMOL software.

5.3.14 Analysis of PHB content using Nile red assay

A pre-culture of *C. necator* H16 WT and *v6C6* variant was cultivated in 5 mL of MSM with 1.00 % (w/v) gluconate supplemented with gentamicin [10 µg/mL] at 30 ºC for 40 h. This preculture was then used to inoculate fresh 300 mL Erlenmeyer flasks (in a 1:100 dilution) with 50 mL of four different media: 1) MSM with 0.59 % (w/v) gluconate and 0.10 % (w/v) NH_4Cl (for nutrient-balanced conditions), 2) MSM with 0.59 % (w/v) gluconate and 0.05 % (w/v) NH_4Cl (for nitrogen-limiting conditions), 3) MSM with 0.50 % (w/v) glycerol and 0.10 % (w/v) NH_4Cl (for nutrient-balanced nitrogen conditions), and 4) MSM with 0.50 % (w/v) glycerol and 0.05 % (w/v) NH_4Cl (for nitrogen-limiting conditions), all four media were supplemented also with gentamicin [10 µg/mL] and grown at 30 ºC and 250 rpm. The growth of samples was monitored by measuring the optical density at 600 nm with a BioPhotometer. A 1 mL aliquot of all samples was harvested at the exponential, early stationary, and late-stationary phases for PHB quantitation. When the aliquot of the samples had an $OD_{600} > 2$, the cell samples were diluted to an $OD_{600} \leq 2$ with MSM in order to keep the fluorescence detection within the linear range. Subsequently, the aliquots of the samples were centrifuged, and the spent media was removed by pipetting, and the pellet was stored at -20ºC until all samples of all growth phases had been collected. For the determination of PHB content with Nile Red, all samples were thawed and resuspended in 0.5 mL of 50 % (v/v) ethanol, then 50 µL of cell suspension of each of the samples was mixed with 50 µL of 10 µg/mL of Nile Red in a microtitre plate, and the fluorescence was measured (E_x 552 nm, E_m 600 nm; bottom read) in a microtiter plate/cuvette reader for 1 h at 10 min intervals, where the fluorescence values were stable after 1 h for all samples, hence, comparisons between samples were made at this time point. The comparison of PHB production among samples was based on the fluorescence per 100 µL of cell culture.

5.3.15 Fluorescence microscopy of PHB stained with Nile Red

A pre-culture of *C. necator* H16 WT and *v6C6* variant was cultivated in 5 mL of MSM with 0.59 % (w/v) gluconate or 0.50 % (w/v) glycerol under nitrogen-limiting conditions supplemented with gentamicin [10 µg/mL] at 30 ºC for 40 h. This preculture was then used to inoculate fresh 50 mL falcon tube (in a 1:100 dilution) with 5 mL of two different media: 1) MSM supplemented with 0.59 % (w/v) gluconate and 0.05 % (w/v) NH_4Cl (for nitrogen-limiting conditions), and 2) MSM supplemented with 0.50 % (w/v) glycerol and 0.05 % (w/v) NH_4Cl (for nitrogen-limiting conditions), all media were supplemented also with gentamicin [10 µg/mL] and grown at 30 ºC and 250 rpm until cells reached early stationary phase. Then, samples were adjusted to an OD_{600} of 7.0 before 200 µL of the samples were centrifuged and the spent media removed. The cell pellet was then resuspended in 50 µL of PBS buffer and used for fluorescence microscopy; cells were visualised using the filter G-1 ((E_x 560 nm, E_m 645 nm).

5.4 Results and discussion

5.4.1 Glycerol assimiliation in *C. necator* H16

As reviewed in Volodina *et al.* (2016), *C. necator* H16 can grow in a wide variety of carbon sources that ranges from carbohydrates, organic acids, and glycerol among other carbon sources. In order to improve the specific growth rate and hence the PHB productivity of *C. necator* H16 from glycerol, the first experiment designed was the study of the natural cell growth of the strain at different concentrations of glycerol, which were compared to cell growth of the strain in a preferred carbon source (gluconate). Glycerol was added as a carbon source in MSM at 0.50 %, 1.00 % and 2.00 % (w/v) of glycerol, and the reference sample was inoculated in MSM with 0.59 % (w/v) of gluconate. The concentrations of glycerol and gluconate at 0.50 % (w/v) and 0.59 % (w/v) respectively, were

chosen to maintain the same carbon molarity (0.16 M) among those two carbon sources. As depicted in **Figure 5.1**, the specific growth rate is significantly slower when cells are grown in glycerol than in gluconate.

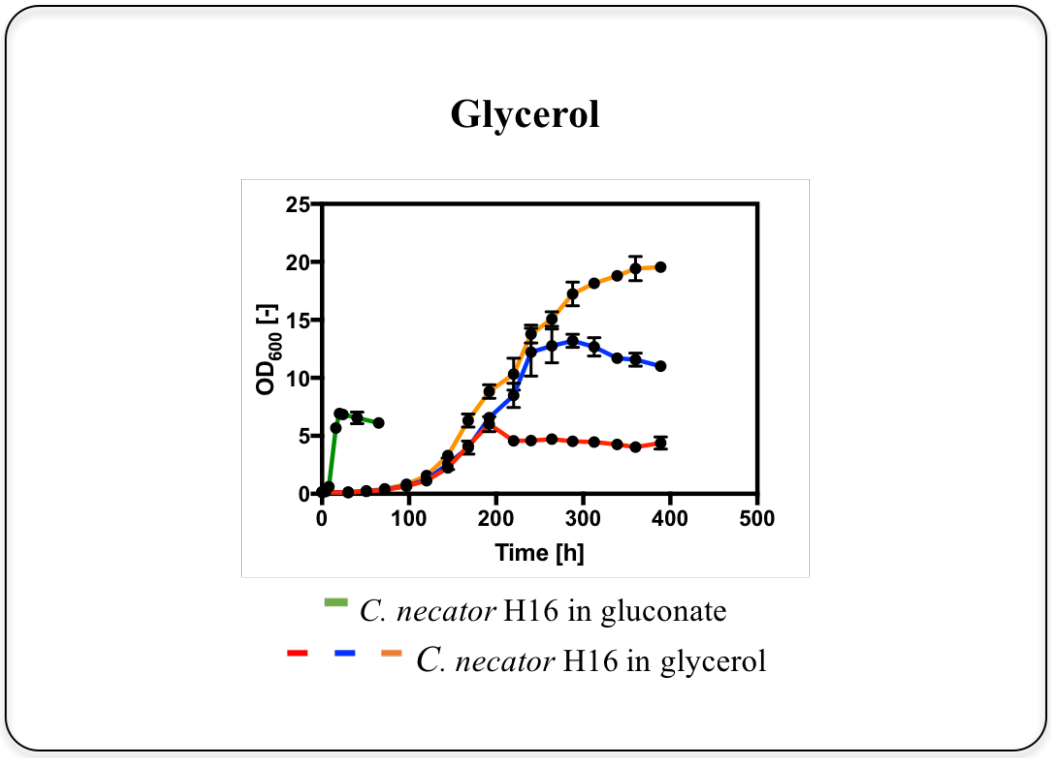


Figure 5.1 Cell growth study of C. necator H16 WT in gluconate and glycerol. C. necator H16 in MSM with 0.59 % (w/v) gluconate (circle, green line), C. necator H16 in MSM with 0.50 % (w/v) glycerol (circle, red line), C. necator H16 in MSM with 1.00 % (w/v) glycerol (circle, blue line), C. necator H16 in MSM with 2.00 % (w/v) gluconate (circle, orange line). Error bars are ± SEM of two independent determinations.

From **Figure 5.1**, it could be observed that effectively, *C. necator* H16 cell growth in 0.50 % (w/v) glycerol and in 0.59 % (w/v) gluconate showed a similar final OD_{600} of 6.0 to 7.0 as expected, since both have the same carbon concentration. The cell growth of *C. necator* H16 in glycerol was characterised by a long lag-phase which took 110 h to reach an OD_{600} of 1.0, while *C. necator* H16 in gluconate reached the same OD_{600} after only 9 h. The specific growth rate of *C. necator* H16 in 0.50 % (w/v) glycerol was 0.024 h^{-1}, while the specific growth rate of *C. necator* H16 in 0.59 % (w/v) gluconate was 0.290 h^{-1}, which is 12-fold faster compared to the specific growth rate of *C. necator* H16 in glycerol. On the other hand, the specific growth rate of *C. necator* H16 in the three different

glycerol concentrations shown in **Figure 5.1** were similar, which ranged from 0.024 to 0.027 h^{-1} (**Table 5.3**), although the final OD_{600} was different among the glycerol samples, this fact corroborated a positive correlation with the increasing glycerol concentrations used. The highest concentrations of glycerol used did not inhibit *C. necator* H16 cell growth, which suggests that the slow growth of the strain might be attributed to slow metabolism of glycerol in *C. necator* H16.

Table 5.3 Specific growth rate of C. necator H16 in gluconate and glycerol.

Strain	Carbon source	Specific growth rate [h^{-1}]
H16_WT	0.59 % (w/v) gluconate	0.290
H16_WT	0.50 % (w/v) glycerol	0.024
H16_WT	1.00 % (w/v) glycerol	0.024
H16_WT	2.00 % (w/v) glycerol	0.027

5.4.2 Rational design strategy for improvement of glycerol assimilation in *C. necator* H16

As mentioned in **Section 1.4.3.5**, the genome of *C. necator* H16 has two pairs of putative glycerol genes: glycerol kinase and glycerol-3-phosphate dehydrogenase. The first pair has the gene loci of H16_A2507 (which encodes for the glycerol kinase gene), and H16_A2508 (which encodes for the glycerol-3-phosphate dehydrogenase); and the second pair has the gene loci of H16_B1198 (which encodes for the glycerol-3-phosphate dehydrogenase gene), and H16_B1199 (which encodes for the glycerol kinase gene) (**Figure 5.2**). The glycerol kinase (GlpK) phosphorylates glycerol by ATP, which yields ADP and glycerol-3-phosphate, then, glycerol-3-phosphate dehydrogenase (GlpD) is in charge of the oxidation of glycerol-3-phosphate to dihydroxyacetone phosphate (**Figure 1.1**). As shown in **Table 1.2**, the gene H16_A2507 has the highest protein sequence identity compared to the known $GlpK_{Ec}$ of *E. coli* with a protein sequence identity of 52 %; also previous work by Shimizu *et al*. (2013) showed that the first pair of genes had a higher transcriptional level. Hence, based on this information available about glycerol metabolism for *C. necator* H16 glycerol-related proteins, the $GlpK_{H16}$ and $GlpD_{H16}$ from the pair of genes H16_A2507 and H16_A2508 were chosen to be studied and cloned for protein expression in *C. necator* H16

using the previously constructed synthetic biology toolbox tailored for *C. necator* H16 described in **Chapter 2**.

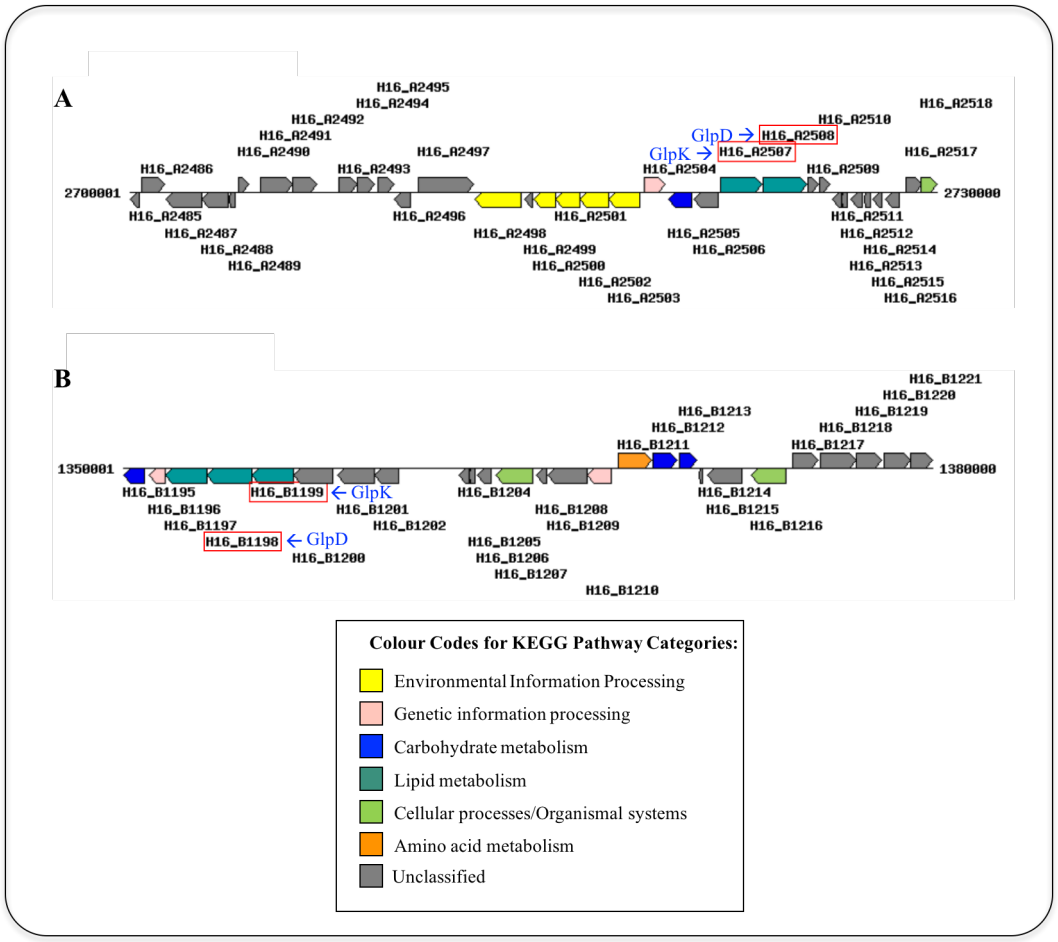


Figure 5.2 Two pairs of putative genes of glycerol metabolism in C. necator H16 (glycerol kinase and glycerol-3-phosphate dehydrogenase). The genes encoding glycerol kinase and glycerol dehydrogenase are boxed in red. The inset shows the colour codes used by KEGG to describe the function of the open reading frames. ***A)*** *The first pair has the gene loci of H16_2507 and H16_A2508.* ***B)*** *The second pair has the gene loci H16_B1198 and H16_B1199.*

5.4.2.1 Construction of plasmids

The plasmids with a constitutive promoter pPj5'c.rbs-glpK$_{H16}$, pPj5'c.rbs-glpK$_{Ec}$, pPj5'c.rbs-glpD$_{H16}$, pPj5'c.rbs1-glpKD$_{H16}$, and pPj5'c.rbs2-glpKD$_{H16}$ were constructed using pPj5c.rbs-RFP plasmid as plasmid backbone. The pPj5c.rbs-RFP is illustrated in **Figure 5.3**, the digested fragment of pPj5c.rbs-RFP harboured the pBBR1 Rep, pBBR1 oriV origin of replication, *mob* gene,

chloramphenicol resistance (Cam^r) cassette (containing P_{cat} and *Cam^r*), P_{j5} promoter, and the *E. coli* consensus RBS. The construction of all plasmids with the P_{j5}, constitutive promoter was successfully achieved by replacing the *rfp* gene from pPj5c.rbs-RFP with the PCR amplified glycerol-related protein genes either from *C. necator* H16 or *E. coli* BW25113.

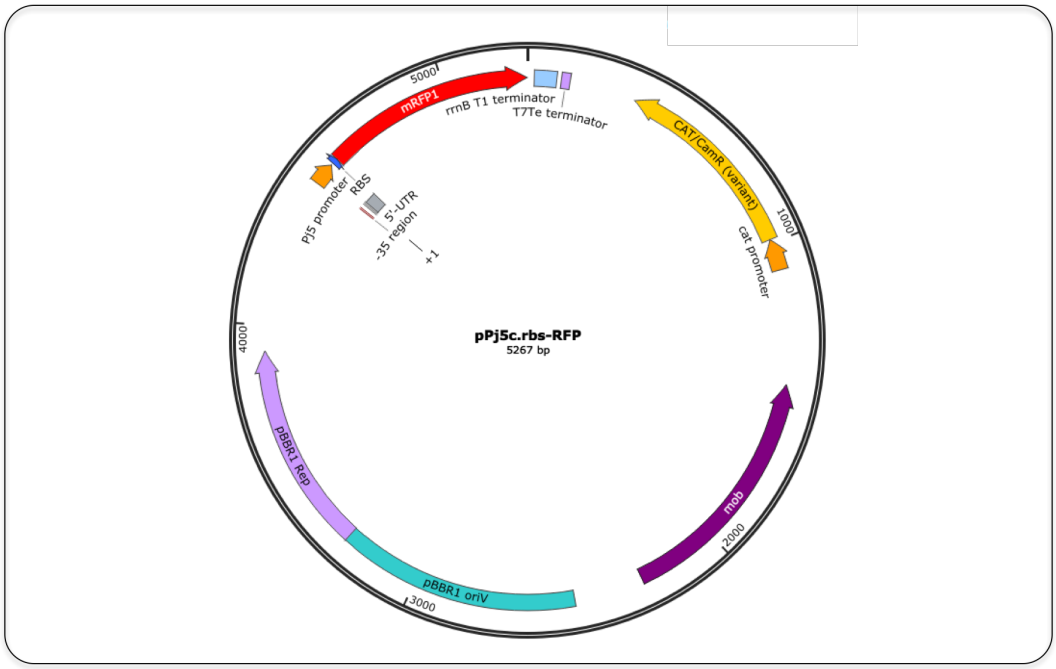


Figure 5.3 Illustration of pPj5c.rbs-RFP map (5.2 kb). The plasmid encodes for pBBR1 Rep, pBBR1 oriV origin of replication, mobilisation sequence mob, CAT promoter, chloramphenicol resistance (Camr), P_{j5} promoter, E. coli consensus RBS, and rfp gene.

The plasmid with an inducible promoter pBADc.rbs-glpK$_{H16}$ was constructed using pBADc.rbs-RFP plasmid as plasmid backbone from the synthetic biology toolbox described in **Chapter 2**. The pBADc.rbs-RFP is illustrated in **Figure 3.1**, the digested fragment of pBADc.rbs-RFP harboured the pBBR1 Rep, pBBR1 oriV origin of replication, *mob* gene, chloramphenicol resistance (Cam^r) cassette (containing P_{cat} and *Cam^r*), L-arabinose-inducible system (containing *araC* and P_{BAD}), and the *E. coli* consensus RBS. The construction of the arabinose-inducible pBADc.rbs-glpK$_{H16}$ plasmid was successfully achieved by replacing the *rfp* gene from pBADc.rbs-RFP with the PCR amplified *glpK$_{H16}$* gene from *C. necator* H16.

The structure of the plasmids constructs used in this work is shown in **Figure 5.4**. The construction of the plasmids with the inducible and constitutive promoters was designed in order to study the effects of the glycerol-related proteins either individually or in combination, where all plasmids were pBBR1-based plasmids (**Figure 5.4A**). The plasmids pPj5'c.rbs1-glpKD$_{H16}$ and pPj5'c.rbs2-glpKD$_{H16}$ which harbour *glpK$_{H16}$* and *glpD$_{H16}$* genes in combination, differed in the ribosome binding site (RBS) preceding the *glpD$_{H16}$* gene, while pPj5'c.rbs1-glpKD$_{H16}$ used the native *glpD$_{H16}$* RBS (RBS in light blue in **Figure 5.4B**), pPj5'c.rbs2-glpKD$_{H16}$ was constructed with the same synthetic RBS (RBS in dark blue in **Figure 5.4B**) preceding both *glpK$_{H16}$* and *glpD$_{H16}$* genes.

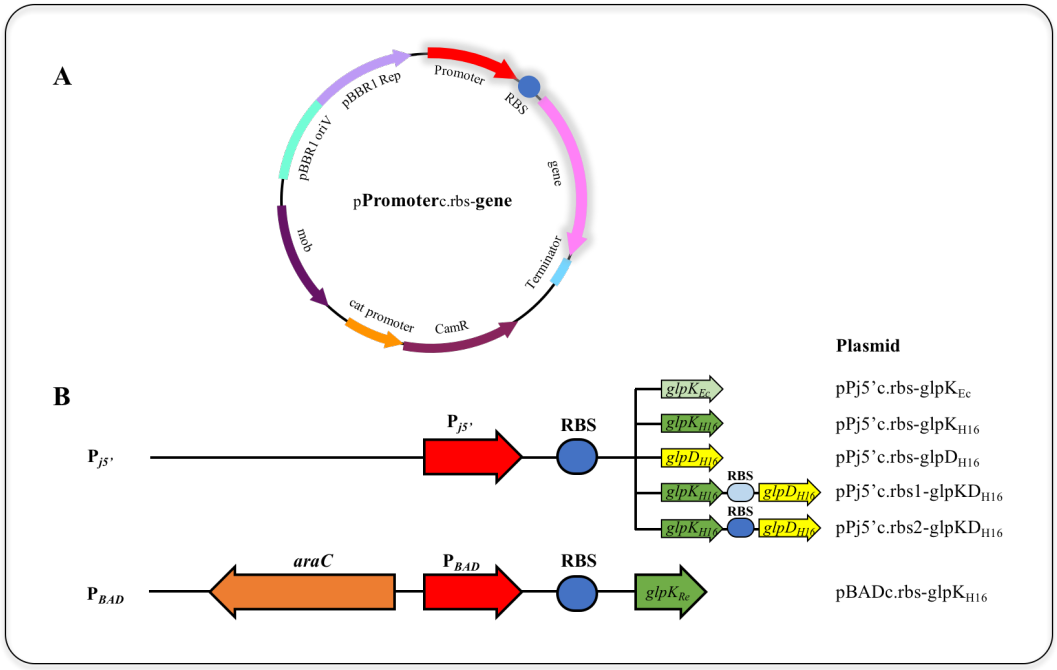


*Figure 5.4 Schematic drawing of the recombinant plasmids used in this work for engineering C. necator H16. **A)** Overall plasmid construct. **B)** Various promoter and gene combinations*

5.4.2.2 Cell growth studies of glycerol metabolic pathway expression in *C. necator* H16

Once all the plasmids were successfully constructed and transformed into *C. necator* H16, cell growth studies were performed in order to see the effects on

glycerol assimilation of the expression of the glycerol-related proteins. The specific growth rate of *C. necator* H16 carrying all plasmids was compared to *C. necator* H16 WT to check for improvement in glycerol assimilation (**Figure 5.5**).

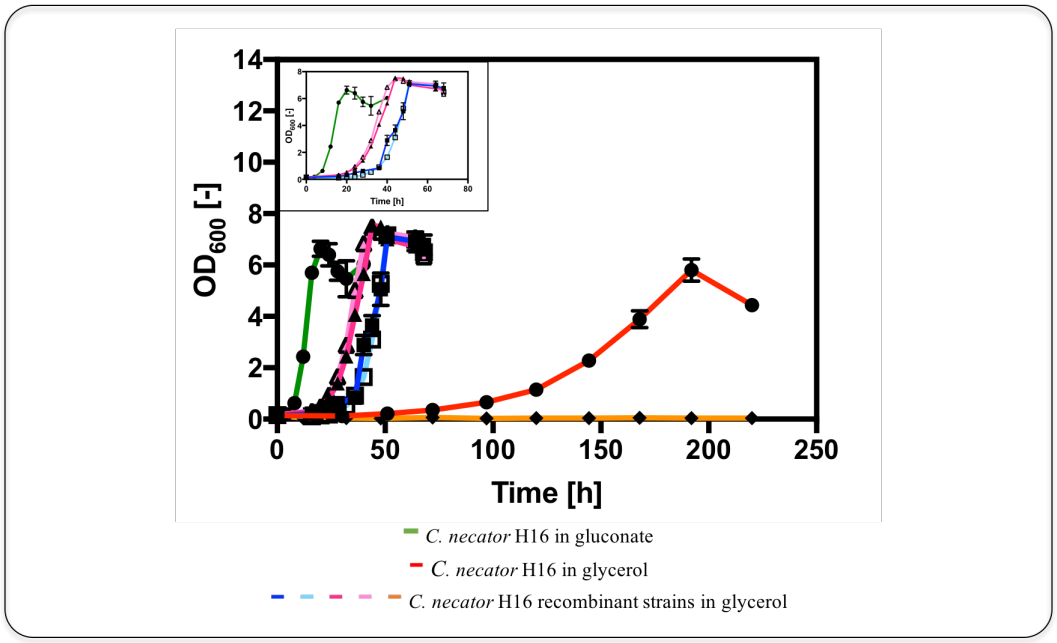


Figure 5.5 Cell growth curve of C. necator H16 harbouring $glpK_{H16}$, $glpK_{Ec}$, $glpD_{H16}$, and $glpKD_{H16}$ plasmids cultivated in 0.50 % (w/v) glycerol. Cell growth study of H16_WT (circle, green line) in MSM with 0.59 % (w/v) gluconate, and H16_WT (circle, red line) in MSM with 0.50 % (w/v) glycerol; H16_$glpK_{H16}$ (square, dark blue line) in MSM with 0.50 % (w/v) glycerol, H16_$glpK_{Ec}$ (open square, light blue line) in MSM with 0.50 % (w/v) glycerol, H16_rbs1.$glpKD_{H16}$ (triangle, dark pink line) in MSM with 0.50 % (w/v) glycerol, H16_rbs2.$glpKD_{H16}$ (open triangle, light pink line) in MSM with 0.50 % (w/v) glycerol, and H16_$glpD_{H16}$ (diamond, orange line) in MSM with 0.50 % (w/v) glycerol. The inset shows the magnified cell growth curve of H16_WT in gluconate and H16 recombinant strains harbouring $glpK_{H16}$, $glpK_{Ec}$, and $glpKD_{H16}$ in glycerol. Error bars are ± SEM of three independent determinations.

From **Figure 5.5**, it could be observed that most of the recombinant strains harbouring the glycerol-related proteins ($glpK_{H16}$, $glpK_{Ec}$, and $glpKD_{H16}$) showed a significant specific growth rate improvement compared to *C. necator* H16 WT, unless for H16_$glpD_{H16}$. The recombinant strains H16_rbs1.$glpKD_{H16}$, which was constructed with a native RBS, and H16_rbs2.$glpKD_{H16}$ constructed with a synthetic RBS, showed also similar cell growth among them, these two recombinant strains showed the best cell growth performance in glycerol compared to the other recombinant strains. The specific growth rate in glycerol improved from 0.024 h^{-1} for *C. necator* H16 WT to 0.120 - 0.140 h^{-1} for

recombinant strains (**Table 5.4**). This suggests that *glpK*$_{H16}$ might be more critical compared to *glpD*$_{H16}$ for improving glycerol assimilation in *C. necator* H16.

Table 5.4 Specific growth rate of C. necator H16 in glycerol.

Strain	Carbon source	Specific growth rate [h^{-1}]
H16_WT	0.59 % (w/v) gluconate	0.290
H16_WT	0.50 % (w/v) glycerol	0.024
H16_glpK$_{H16}$	0.50 % (w/v) glycerol	0.120
H16_glpK$_{Ec}$	0.50 % (w/v) glycerol	0.130
H16_glpD$_{H16}$	0.50 % (w/v) glycerol	-
H16_rbs1.glpKD$_{H16}$	0.50 % (w/v) glycerol	0.130
H16_rbs2.glpKD$_{H16}$	0.50 % (w/v) glycerol	0.140

On the other hand, as observed in **Figure 5.5**, the recombinant strain H16_glpD$_{H16}$ (diamond, orange line) not only did not improve glycerol assimilation, but it was not able to grow in glycerol, even though the presence of plasmid-borne *glpD*$_{H16}$ gene in the transformed cells of *C. necator* H16 was confirmed by colony PCR. As reviewed in Allaman *et al*. (2015), the possible reason of the growth inhibition when GlpD$_{H16}$ was expressed in *C. necator* H16 might be that the increase of GlpD$_{H16}$ could lead to dihydroxyacetone phosphate accumulation, which in turn might be converted into methylglyoxal by the methylglyoxal synthase (MgsA), which is encoded by H16_A0932 in *C. necator* H16; methylglyoxal, is a highly reactive dicarbonyl compound, and it is one of the most potent glycating agents, readily reacting with nucleic acids, proteins and lipids, that could lead to cellular damage.

Once that GlpK was identified as a key enzyme for improving glycerol assimilation in *C. necator* H16, the construction of the pBADc.rbs-glpK$_{H16}$ was performed in order to increase the expression level of the enzyme by the L-arabinose-inducible P$_{BAD}$ promoter, although despite three independent attempts to improve specific growth rate using the pBADc.rbs-glpK$_{H16}$ plasmid, variable results were obtained (data not shown). In contrast to the improvement in glycerol assimilation observed with the recombinant strain H16_glpK$_{H16}$ driven by the constitutive promoter P$_{j5'}$, the recombinant strain H16_glpK$_{H16}$ driven by the inducible promoter P$_{BAD}$ showed either similar or slower growth in glycerol compared to the *C. necator* H16 WT. This variability observed, could be due to

protein expression is sensitive to the point of L-arabinose induction, or due to plasmid instability during cultivation. A recent publication by Aboulnaga *et al*. (2018) also reported that the constitutive expression of $glpK_{Ec}$ was more successful than the inducible expression of the same gene.

Apart from that, it is worth to note that the integration of $glpK_{Ec}$ into *C. necator* H16 chromosome was previously reported by Fukui *et al*. (2014) to improve glycerol assimilation with some success, but the effects of $glpK_{H16}$ and $glpD_{H16}$ genes native of the *C. necator* H16 strain were not investigated. Fukui *et al*. (2014) showed that the plasmid harbouring the $glpK_{EC}$ gene was not successfully transformed into *C. necator* H16, likely due to spontaneous deletion of the gene by *C. necator* H16. In order to understand better the differences between $glpK_{H16}$ and $glpK_{Ec}$, the plasmid pPj5'c.rbs-glpK$_{Ec}$ was constructed; after the analysis of the cell growth study shown in **Figure 5.5**, the specific growth rate of H16_glpK$_{Ec}$ (open square, light blue line) was 0.130 h^{-1}, which was similar to the specific growth rate of H16_glpK$_{H16}$ (square, dark blue line) (0.120 h^{-1}). Thus, it is possible that the difference observed for plasmid-based $glpK_{Ec}$ might be attributed to the use of different promoters that could cause variability in plasmid stability and expression level in *C. necator* H16, which has been reported in other publications (Johnson *et al.*, 2018; Aboulnaga *et al*., 2018).

From these experiments performed in the rational design approach, it could be observed that the specific growth rate of the recombinant strains (≈ 0.130 h^{-1}) in 0.50 % (w/v) glycerol was significantly improved compared to the specific growth rate of *C. necator* H16 (0.024 h^{-1}) in 0.50 % (w/v) glycerol, which was 5.4-fold faster compared to the specific growth rate of *C. necator* H16, and only 2.2-fold slower compared to the specific growth rate of *C. necator* H16 in 0.59 % (w/v) gluconate.

5.4.3 Evolutionary engineering strategy for improvement of glycerol assimilation in *C. necator* H16

A parallel strategy to improve glycerol assimilation in *C. necator* H16 apart from rational design was performed, which was based on evolutionary engineering *via* adaptive evolution since the underlying reason of the slow growth of *C. necator* H16 growth in glycerol is not fully understood. In **Chapter 4**, directed evolution *via* random mutagenesis of the entire genome of *C. necator* H16 was used to evolve the strain, where a population of *C. necator* H16 cells was subjected to chemical mutagenesis in order to create unknown random mutations which were furtherly screened in a defined medium with a selection pressure to select only the mutated cells with the desired improved phenotype. A similar approach of evolutionary engineering was used to evolve *C. necator* H16 cells to enhance its glycerol utilisation, although the conditions of the experiment allowed to evolve a population of *C. necator* H16 WT cells in a defined habitat –glycerol- which would allow to generate unknown mutations but in specific DNA regions related to glycerol assimilation, thus linking phenotype with genotype.

As reviewed in Porcar (2010), adaptive evolution, is a process whereby a microorganism can become better to live in a defined habitat by improving its genetic circuits; the experimental design is straightforward, it involves maintaining cells in the exponential growth or early stationary phase by daily passage of cultures into a fresh medium under a selection pressure, in this case, glycerol. This strategy can be a simpler and faster tool for the selection of desired phenotypes compared to the previous evolutionary engineering tool studied in **Chapter 4**, although it is not always feasible for all studies (as it was the case of the studies of **Chapter 4**). Thus, adaptive evolution was used as a strategy to improve the phenotype of *C. necator* H16 cells for enhancement of glycerol assimilation, where *C. necator* H16 WT cells were evolved by serial cultivation in MSM with 0.50 % (w/v) glycerol.

5.4.3.1 Genetic variation of *C. necator* H16 obtained after adaptive evolution

The initial selective condition in adaptive evolution was MSM with 0.50 % (w/v) glycerol. Serial cultivation was followed for up to six rounds of adaptive evolution, where cells were transferred from one culture to the next one at the early-stationary phase (**Figure 5.6**). The number of rounds performed for this experiment of adaptive evolution was accumulated until no further improvement in glycerol assimilation was found. A significant improvement in glycerol assimilation by the new population of *C. necator* H16 was observed from the second round of adaptive evolution (open circle, grey line).

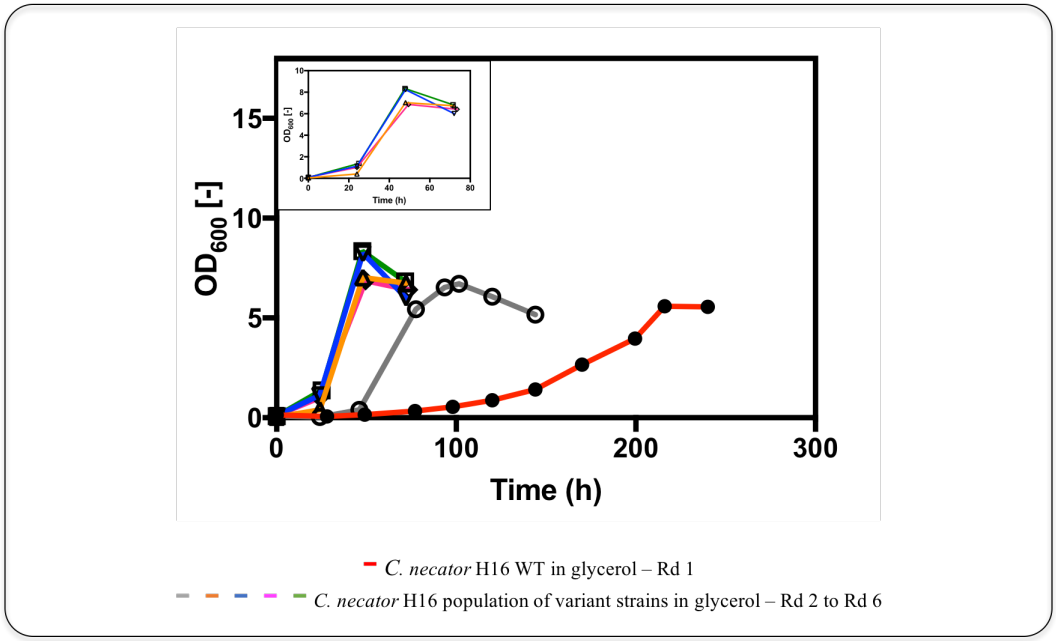


Figure 5.6 Cell growth curve of C. necator H16 in the six serial cultivations of adaptive evolution in MSM with 0.50 % (w/v) glycerol. Round 1 (circle, red line), Round 2 (open circle, grey line), Round 3 (open triangle, orange line), Round 4 (inverted open triangle, blue line), Round 5 (open diamond, pink line), and Round 6 (open square, green line). The inset shows the magnified cell growth curve of the last four rounds of adaptive evolution in glycerol. Abbreviations: Rd, round.

From **Figure 5.6** it could be noted that the cell population adapted quickly, where cells were able to reach their maximum optical density from 9 days in Round 1 to only 4 days in Round 2, and to 2 days in Round 3. The time required to reach maximum optical density did not change significantly from Round 3 to Round 6, although the lag phase (inset **Figure 5.6**) in these rounds seemed to be different. It is perhaps unsurprising that *C. necator* H16 WT was capable of adapting quickly

to improve its growth in glycerol, as this kind of adaptation has been previously observed when *C. necator* H16 was adapted to grow in glucose (Franz *et al*., 2012). After the six rounds of adaptive evolution, the population of variants from Round 4, Round 5, and Round 6 was selected for screening analysis.

5.4.3.2 Analysis of variant strains of *C. necator* H16 obtained after adaptive evolution from Round 4 to 6.

The screening of the population of *C. necator* H16 variant strains was performed in three different microtitre plates in MSM with 0.50 % (w/v) glycerol supplemented with gentamicin (one microtitre plate for each of the rounds of adaptive evolution), where a total of 270 variant strains were screened (**Figure 5.7**). Each microtitre plate contained control samples to compare cell growth and find the variants with an improved phenotype; the control samples were the parent strain *C. necator* H16 WT in MSM with 0.59 % (w/v) gluconate and *C. necator* H16 in MSM with 0.50 % (w/v) glycerol, which were always included in triplicates in all experiments. The best variant strains for each of the rounds of adaptive evolution were chosen as follows: for each OD_{595} measurement taken during the cell growth curve, the two highest OD_{595} values corresponding to variant strains were selected to be followed through the entire cell growth study, then, all the variants selected were compared against each other and against the parent strain to identify the variant strains with the best cell growth performance in glycerol. The variant strains were named as "v" for variant, followed by a "number" from 4 to 6, which corresponds to the round of adaptive evolution, and a "letter + number" corresponding to the well of the microtitre plate where the variant strain was grown.

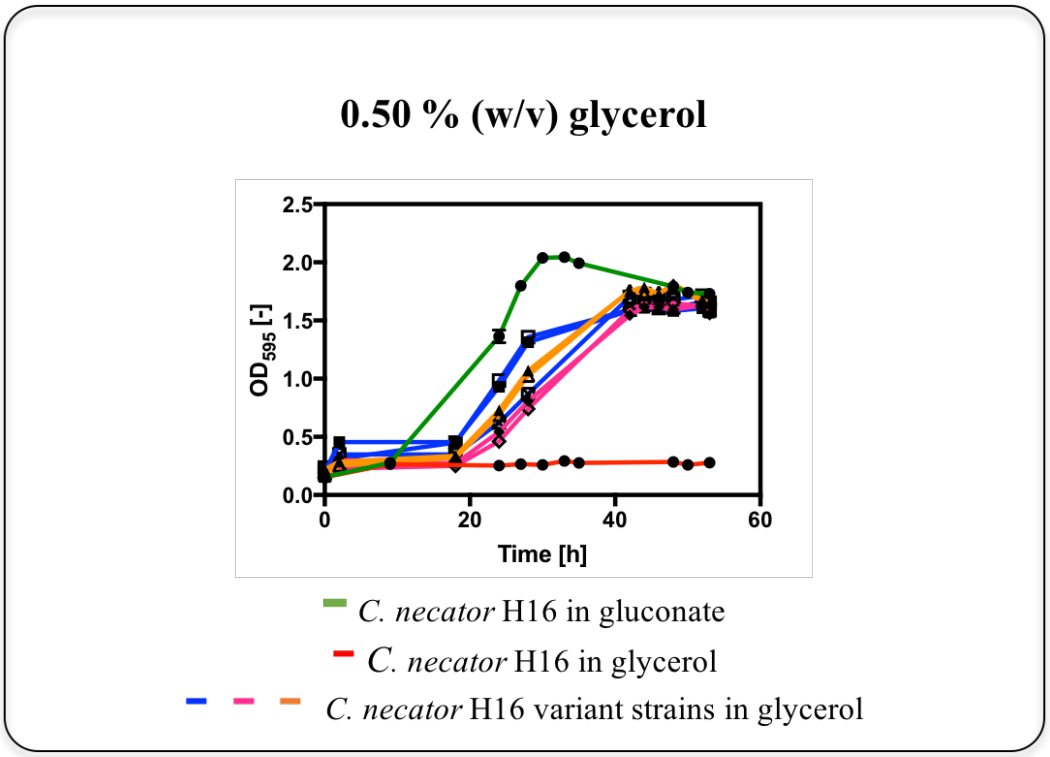


Figure 5.7 Screening of variant strains adapted in 0.50 % (w/v) glycerol from Round 4, Round 5 and Round 6. Screening analysis was performed in microtitre plates with 150 µL of MSM with 0.50 % (w/v) glycerol. C. necator H16 WT in MSM with 0.59 % (w/v) gluconate (circle, green line), and C. necator H16 WT in MSM with 0.50 % (w/v) glycerol (circle, red line); v4C2 in MSM with 0.50 % (w/v) glycerol (triangle, orange line), v4D1 in MSM with 0.50 % (w/v) glycerol (open triangle, orange line), v5A4 in MSM with 0.50 % (w/v) glycerol (diamond, pink line), v5E8 in MSM with 0.50 % (w/v) glycerol (open diamond, pink line), v6C6 in MSM with 0.50 % (w/v) glycerol (square, blue line), v6F8 in MSM with 0.50 % (w/v) glycerol (open square, blue line), and v6G7 in MSM with 0.50 % (w/v) glycerol (open crossed square, blue line). Error bars are ± SEM of three independent determinations. Abbreviations: v, variant.

The screening of the last three rounds of adaptive evolution showed variants with an improved phenotype in glycerol assimilation compared to their parent strain *C. necator* H16 WT. While the parent strain did not show any growth even after 52 h of cultivation, all the best variants depicted in **Figure 5.7** reached the maximum optical density in less than 42 h of cultivation. Although all variants from the three microtitre plates grew similarly, and all of them showed a significant improvement in glycerol assimilation compared to the WT parent strain, two of the variants from the Round 6 of adaptive evolution *v6C6* (square, blue line) and *v6G7* (open square, blue line) in the screening analysis showed a slightly better cell growth performance in glycerol than the rest of the variants of Round 5 and Round 4 of adaptive evolution.

After screening analysis, the characterisation of the three variants from the Round 6 of adaptive evolution (*v6C6*, *v6F8*, and *v6G7*), and only one variant from Round 5 (*v5A4*) and one from Round 4 (*v4C2*) was performed in 150 µL of MSM with 0.50 %, 1.00 % and 2.00 % (w/v) glycerol in microtitre plates to identify the best variant strains, all samples were included in triplicates. The cell growth study for the characterisation of the variant strains is shown in **Figure 5.8**. Variant strains were also grown in MSM with 0.59 % (w/v) gluconate to check if the variant strains could still grow like the WT.

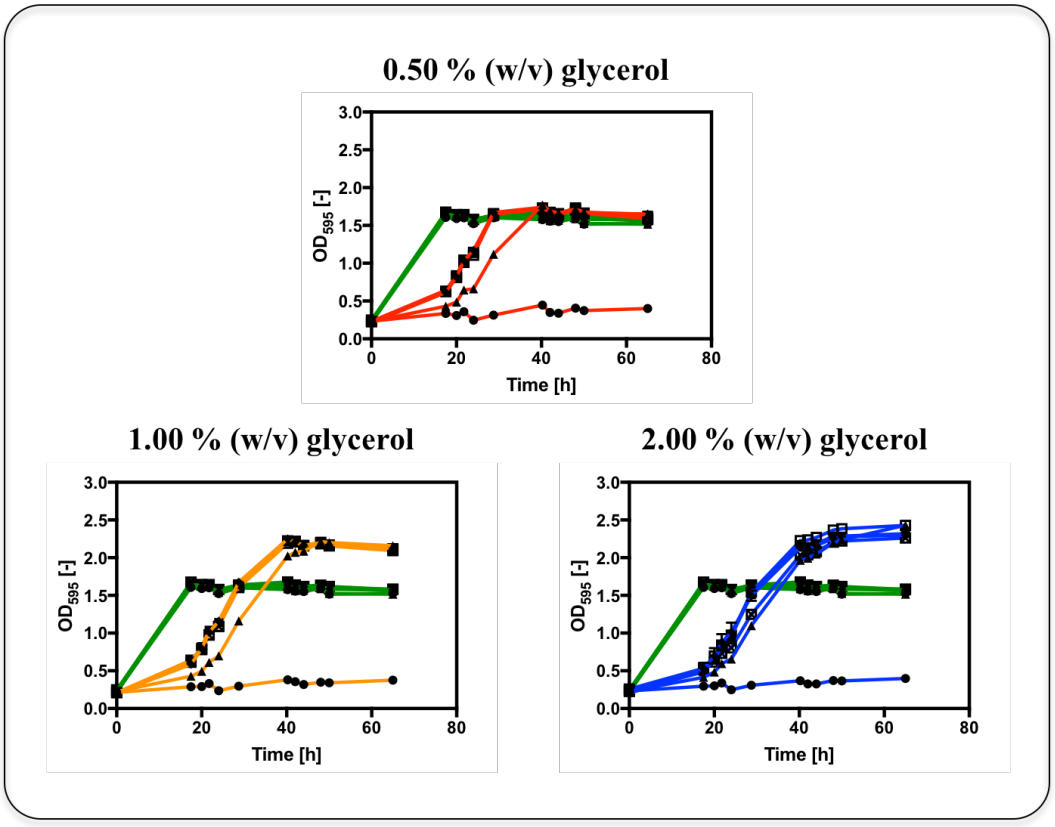


Figure 5.8 Characterisation of variant strains adapted in 0.50 % (w/v) glycerol from Round 4, Round 5 and Round 6 from adaptive evolution. Characterisation analysis was performed in microtitre plates with 150 µL of MSM with 0.50 %, 1.00 % and 2.00 % (w/v) glycerol. C. necator H16 (circles), v4C2 (triangles), v5A4 (diamonds), v6C6 (squares), v6F8 (crossed squares), and v6G7 (open squares) were grown in MSM with 0.59 % (w/v) gluconate (green lines), 0.50 % (w/v) glycerol (red lines), 1.00 % (w/v) glycerol (orange lines), and 2.00 % (w/v) glycerol (blue lines). Error bars are ± SEM of three independent determinations. Abbreviations: v, variant.

From **Figure 5.8** it was confirmed that the cell growth of the variant strains screened from the last three rounds of adaptive evolution was effectively

improved compared the cell growth of *C. necator* H16 WT in glycerol. On the other hand, it was observed as well that variant strains were also capable of growing in gluconate (green lines) as the WT strain, which showed a similar maximum OD_{595} compared to the samples grown in 0.50 % (w/v) glycerol (red lines) as they have the same carbon concentration. The three graphs depicted in **Figure 5.8** showed that all variants grew with a similar growth rate in all different concentrations of glycerol, unless for the variant *v4C2*, which was isolated from the Round 4 of adaptive evolution. Hence, from this point onwards, only the variants from the Round 6 of adaptive evolution (*v6C6*, *v6F8*, and *v6G7*) were used for further investigations.

5.4.3.3 Confirmation of improved glycerol-utilising phenotype

To confirm that the improved phenotype was not transient adaptations but true mutations, the three variants from the Round 6 of adaptive evolution *v6C6*, *v6F8*, and *v6G7* were subjected to five consecutive rounds of cultivation in either a complex media (NB) or synthetic medium (MSM with 1.00 % (w/v) gluconate) in the absence of glycerol. After the five consecutive rounds of cultivation in both media, the variant strains were transferred back to MSM with 0.50 % (w/v) glycerol to confirm if these were still capable of growing in glycerol with the same specific growth rate as observed before. The three variants maintained their growth rates at the same level when they were first isolated after 5 passages in NB or MSM supplemented with gluconate (data not shown); this indicated that the improved growth rate in glycerol was a result of a genetic mutation and not only physical adaptation during adaptive evolution.

5.4.3.4 Cell growth studies of variants of *C. necator* H16

Due to previous studies described in **Section 5.4.3.3** demonstrated that there were true mutations in variants of the Round 6 of adaptive evolution, a cell growth study of one of the variants from Round 6 (*v6C6*) was performed in order to investigate its improvement in specific growth rate compared to *C. necator* H16

WT (**Figure 5.9A**). The specific growth rate of *v6C6* grown in 0.50 % (w/v) glycerol was 8-fold faster than its parent strain *C. necator* H16 WT in the same concentration of glycerol, and it was 69 % the rate in gluconate. Due to the *v6C6* variant strain showed a faster specific growth rate compared to the WT, the variant strain was transformed with the plasmids pPj5'c.rbs-glpK$_{H16}$, pPj5'c.rbs1-glpKD$_{H16}$, and pPj5'c.rbs2-glpKD$_{H16}$ (**Figure 5.9B**). Nevertheless, no improvement in growth was observed for the *v6C6* recombinant strains, on the contrary, the recombinant strains resulted in a reduction of specific growth rate of the variant strain as shown in **Table 5.5**, which was reduced from 0.20 h^{-1} to 0.14 – 0.15 h^{-1}. The specific growth rate of the recombinant strain v6C6_glpK$_{H16}$ (0.21 h^{-1}) was very similar to the specific growth rate of *v6C6* (0.20 h^{-1}) but with a longer lag phase, which might have been due to an additional metabolic burden of protein expression.

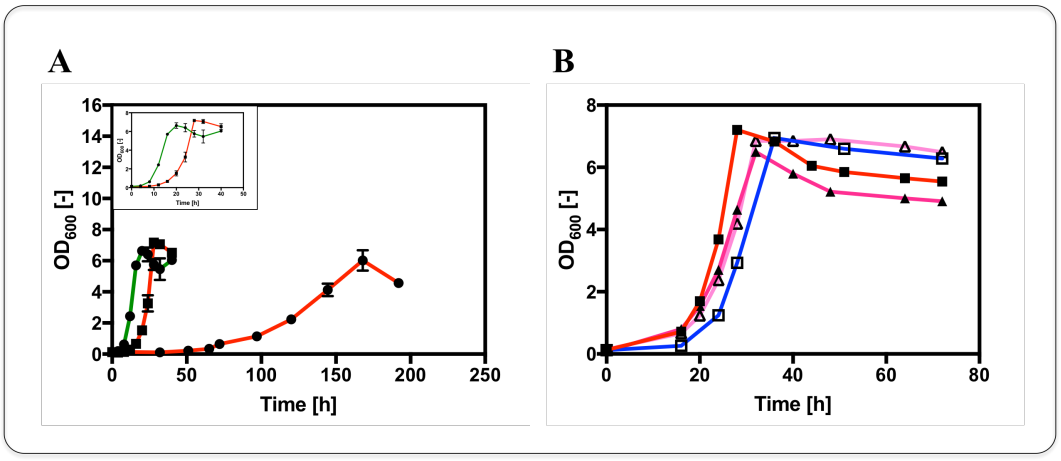


*Figure 5.9 Cell growth curve of C. necator H16 WT, v6C6 and v6C6 recombinant strains harbouring glpK$_{H16}$, and glpKD$_{H16}$ plasmids in 0.50 % (w/v) glycerol. **A)** Growth curve of C. necator H16 in MSM with 0.59 % (w/v) gluconate (circle, green line), v6C6 in MSM with 0.50 % (w/v) glycerol (square, red line), and C. necator H16 in MSM with 0.50 % (w/v) glycerol (circle, red line). The inset shows the magnified cell growth curve of C. necator H16 in gluconate and v6C6 in glycerol. **B)** Growth curve of v6C6 in MSM with 0.50 % (w/v) glycerol (square, red line), v6C6_glpK$_{H16}$ in MSM with 0.50 % (w/v) glycerol (open square, blue line), v6C6_rbs1.glpKD$_{H16}$ in MSM with 0.50 % (w/v) glycerol (triangle, dark pink line), and v6C6_rbs2.glpKD$_{H16}$ in MSM with 0.50 % (w/v) glycerol (open triangle, light pink line). Error bars are ± SEM of three independent determinations.*

The comparison table showing the different specific growth rates of all the recombinant and variant strains developed in this work are shown in **Table 5.5**,

from which is noted that the *v6C6* variant strain remained as the fastest growing glycerol-utilising strain obtained from these studies.

Table 5.5 Specific growth rate of C. necator H16, v6C6 and recombinant strains in glycerol.

Strain	Carbon source	Specific growth rate [h^{-1}]
H16_WT	0.59 % (w/v) gluconate	0.290
H16_WT	0.50 % (w/v) glycerol	0.024
H16_glpK$_{H16}$	0.50 % (w/v) glycerol	0.120
H16_glpK$_{Ec}$	0.50 % (w/v) glycerol	0.130
H16_glpD$_{H16}$	0.50 % (w/v) glycerol	-
H16_rbs1.glpKD$_{H16}$	0.50 % (w/v) glycerol	0.130
H16_rbs2.glpKD$_{H16}$	0.50 % (w/v) glycerol	0.140
v6C6	0.50 % (w/v) glycerol	0.200
v6C6_ glpK$_{H16}$	0.50 % (w/v) glycerol	0.210
v6C6_ rbs1.glpKD$_{H16}$	0.50 % (w/v) glycerol	0.140
v6C6_ rbs2.glpKD$_{H16}$	0.50 % (w/v) glycerol	0.150

Apart from gluconate, fructose is another preferred carbon source for *C. necator* H16, where at 0.50 % (w/v) fructose, *C. necator* H16 has shown to have a specific growth rate of 0.28 h^{-1} (Fukui *et al*., 2014), similar to the specific growth rate with 0.59 % (w/v) gluconate (0.29 h^{-1}). The best engineered strain *v6C6* has a specific growth rate of 0.20 h^{-1} when cells are grown in 0.50 % (w/v) glycerol, although the specific growth rate is lower than when *C. necator* H16 is grown in a preferred carbon source such as gluconate or fructose, the use of these carbohydrates is expensive, and hence, their application is not attractive for bioproduction.

After identifying that cells of *v6C6* showed a faster specific growth rate compared to the recombinant strains of *C. necator* H16 and recombinant strains of *v6C6*, a further adaptive evolution was performed on *v6C6* using different concentrations of glycerol (0.25 %, 2.00 % and 5.00 % (w/v) glycerol), were 267 variants were screened, but this adaptive evolution experiment did not show variants with a further improvement in glycerol assimilation (data not shown).

5.4.3.5 Mutations found in *glpK$_{v6c6}$*

As reviewed in Bloom *et al*. (2009), evolutionary engineering has revealed that a single amino acid mutation is capable of enhancing catalytic activity or stability of

a protein, and that adaptation can generally occur through pathways that consist of sequential beneficial mutations, as might have occurred for with the variants isolated from the last rounds of adaptive evolution, which after adaptive evolution improved glycerol utilisation significantly.

The *v6C6* variant strain was isolated from the Round 6 of adaptive evolution, and even though it was evident that *C. necator* H16 effectively went through beneficial true mutations after adaptive evolution –as reflected by its new phenotype characteristics with an enhancement of glycerol assimilation-, the precise number, location, and nature of the genetic mutation(s) in its genome is unknown. Nevertheless, the rational design studies demonstrated that $glpK_{H16}$ played an important role in the enhancement of glycerol assimilation. Thus, it was decided to sequence the *glpK* gene (gene locus H16_A2507) and its 500-bp upstream element in *v6C6* to check for gene modifications in that targeted region. Even though the genome size of *C. necator* H16 is 7,416,678 bp, it was surprising that after DNA sequencing of the $glpK_{H16}$ gene and its 500-bp upstream element, mutations were found within the $glpK_{H16}$ gene (**Figure 5.10**).

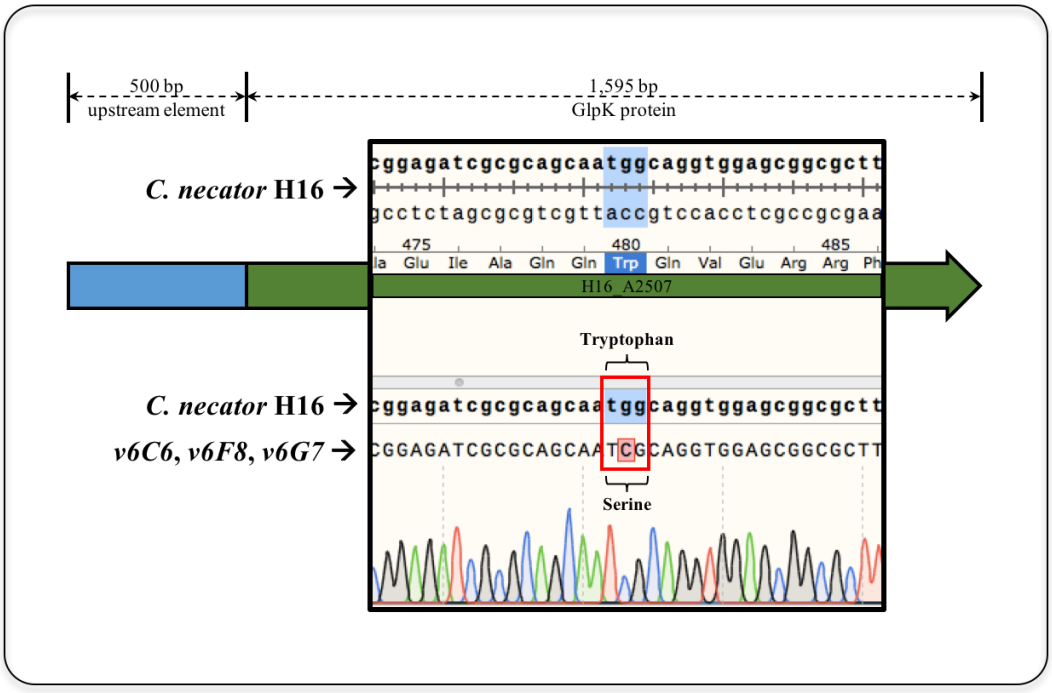


Figure 5.10 Chromatogram and DNA sequence of GlpK in C. necator H16 and variant strains. A single non-synomymous mutation was found after DNA sequencing of the glpK gene (in green) and its 500-bp upstream element (in blue) of v6C6, v6F8, and v6G7. The same W480S substitution was found in the three variants isolated from the Round 6 of adaptive evolution.

The mutation found within the *glpK$_{H16}$* gene was a non-synonymous mutation, where the amino acid position 480 was mutated from tryptophan to a serine (W480S substitution). Thus, after finding this substitution, the *glpK$_{H16}$* gene and its 500-bp upstream element of the other two variants isolated as well from the Round 6 of adaptive evolution -*v6F8*, and *v6G7*- were also analysed, and the same substitution was found. This could indicate that this modification was enriched during the different rounds of adaptive evolution and that this could be the mutation responsible for the enhancement of glycerol assimilation.

5.4.3.6 Protein model of GlpK$_{v6C6}$

A protein model of GlpK$_{H16}$ and GlpK$_{v6C6}$ was created using GlpK$_{Ec}$ (PDB 1BOT) as a template (**Figure 5.11**). From the protein model created, it could be observed that W480 is found on a coiled region that is away from the catalytic centre. Due to W480 is not found in any known glycerol-binding, ATP-binding or activation loop in GlpK (Yeh *et al*., 2004; Anderson *et al*., 2007), it is difficult to predict if the mutation could have altered its catalytic function, further investigations would be required to confirm the effects of this mutation found in the protein due to this substitution; it is also possible that other genetic mutations could be found in other regions of *C. necator* H16 variant strains which could also be related to the enhancement of glycerol assimilation, where also, further analysis must be performed to understand how mutations helped to improve glycerol utilisation. For instance, whole-genome sequencing of the variant strains could be performed, as this would help to identify other possible mutations in the genome that could be responsible for the enhancement of glycerol-utilising phenotype observed in the variant strains.

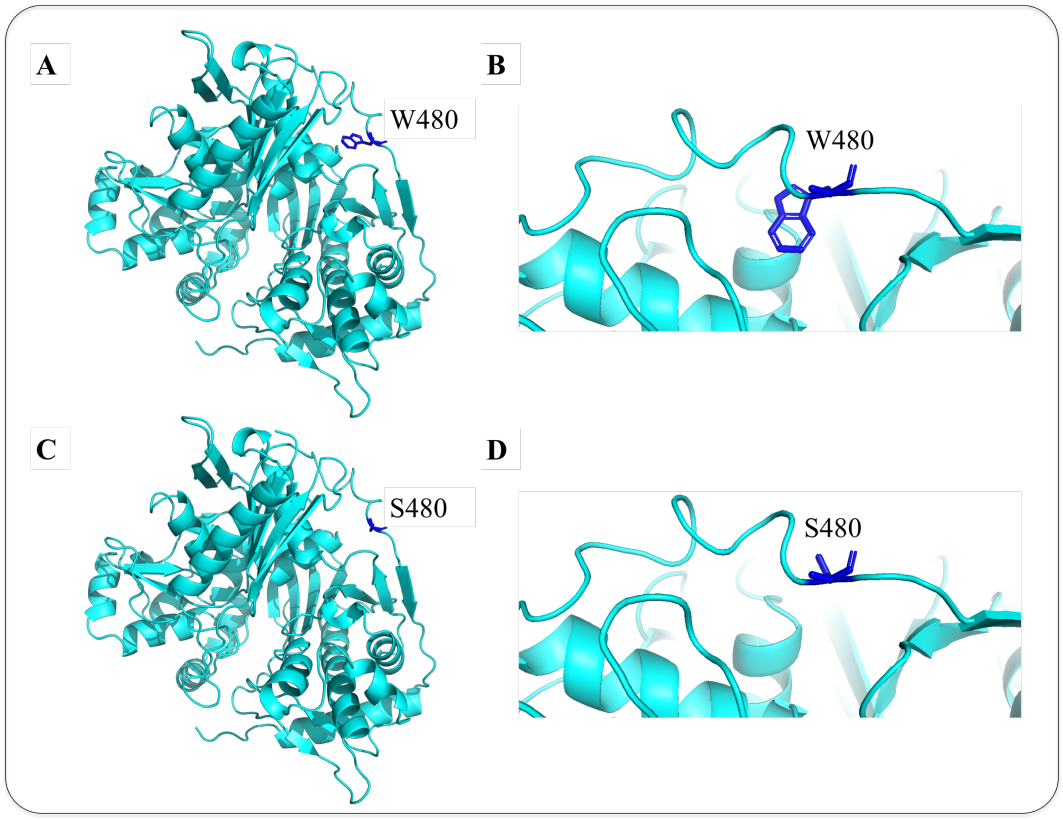


Figure 5.11 Protein models of C. necator H16 GlpK$_{H16}$ (gene locus: H16_A2507). ***A)*** *and* ***B)*** *GlpK$_{H16}$ with Trp480 shown in dark blue.* ***C)*** *and* ***D)*** *GlpK$_{v6C6}$ (GlpK W480S mutant) with Ser480 shown in dark blue.*

On the other hand, a Conserved Domain Search (Marchlet-Bauer *et al*., 2004) in a multiple sequence alignments of GlpK$_{H16}$ against related protein in a variety of organisms, revealed that W480 is a reasonable conserved region (**Figure 5.12**), hence, the impact of this mutation could be of significant importance for the enhancement of glycerol utilisation.

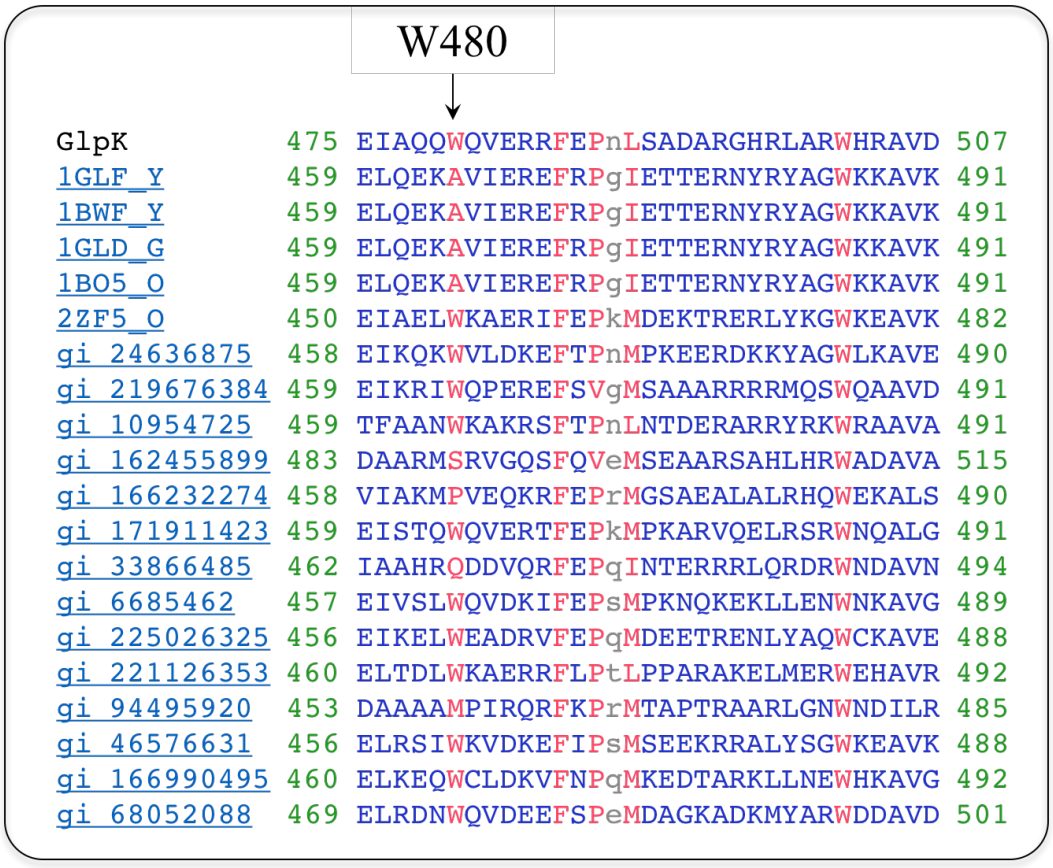


Figure 5.12 Conserved Domain Search (Marchlet-Bauer et al., 2004) of $GlpK_{H16}$. The illustration shows the multiple sequence alignment of $GlpK_{H16}$ and related proteins from a variety of organisms. Sequence alignment is shown using colour bits 3.5 for the amino acids, where red colour indicates highly conserved and blue colour less conserved.

5.4.4 Chemical production in *C. necator* H16 WT and *v6C6* using glycerol as a carbon source

After investigating both strategies –rational design and evolutionary engineering- to engineer *C. necator* H16 for enhancement of glycerol utilisation, PHB was chosen as an example of a chemical that could be produced using glycerol as feedstock, since the strain is a natural PHB producer. The PHB productivity was analysed in *C. necator* H16 WT and the variant strain *v6C6*, which showed to be the best-engineered strain of *C. necator* H16 for utilising glycerol. Two different carbon sources were used for PHB quantitation, a preferred carbon source (gluconate), and glycerol. PHB quantitation was surveyed using a Nile red assay (Tyo *et al.*, 2006). Nile red is a lipophilic stain that is widely used for PHB quantitation, and in order to use it for PHB quantitation, initially, a standard curve

of Nile red was performed in order to keep the fluorescence detection within the linear range (**Figure 5.13**).

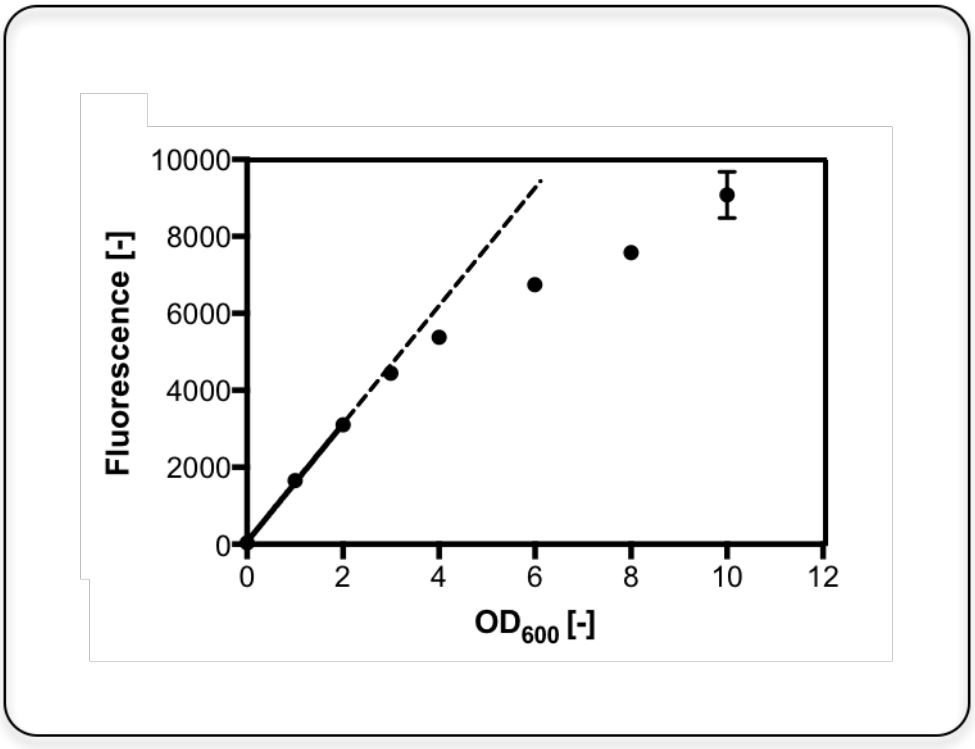


Figure 5.13 Standard curve of Nile red assay for PHB quantitation. Error bars are ± SEM of three independent determinations.

From the standard curve shown in **Figure 5.13**, it was determined that the maximum optical density used for Nile red assays would be OD_{600} of 2.0. The analysis of PHB content using Nile red assay was performed by generating a cell growth curve of *C. necator* H16 and *v6C6* in four different media: 1) MSM with 0.59 % (w/v) gluconate and 0.10 % (w/v) NH_4Cl (for nutrient-balanced nitrogen conditions), 2) MSM with 0.59 % (w/v) gluconate and 0.05 % (w/v) NH_4Cl (for nitrogen-limiting conditions), 3) MSM with 0.50 % (w/v) glycerol and 0.10 % (w/v) NH_4Cl (for nutrient-balanced nitrogen conditions), and 4) MSM with 0.50 % (w/v) glycerol and 0.05 % (w/v) NH_4Cl (for nitrogen-limiting conditions), where an aliquot of all samples was withdrawn from the cultures at the exponential, early stationary and late stationary growth face for PHB quantitation. As observed in **Figure 5.14** for both strains and carbon sources, a higher

fluorescence was obtained when cells were grown in nitrogen-limiting media, which agrees with the fact that *C. necator* H16 produces more PHB under nutrient stress (reviewed in Lee, 1996).

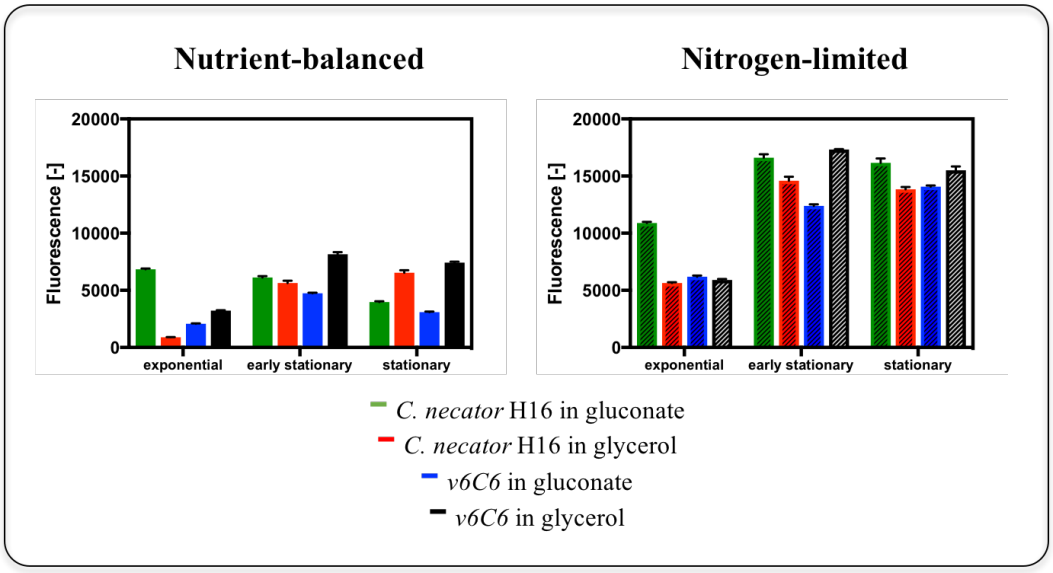


Figure 5.14 PHB production surveyed using Nile red fluorescence assay. PHB production was quantified in 0.59 % (w/v) gluconate and 0.50 % (w/v) glycerol under nutrient-balanced and nitrogen-limited conditions. Nutrient-balanced conditions: C. necator H16 WT (green solid bars), and v6C6 variant (blue solid bars) in 0.59 % (w/v) gluconate, C. necator H16 WT (red solid bars), and v6C6 variant (black solid bars) in 0.50 % (w/v) glycerol. Nitrogen-limited conditions: C. necator H16 WT (green stripe bars), and v6C6 variant (blue stripe bars) in 0.59 % (w/v) gluconate, C. necator H16 WT (red stripe bars), and v6C6 variant (black stripe bars) in 0.50 % (w/v) glycerol. Error bars are ± SEM of three independent determinations.

From **Figure 5.14** it could also be observed that under nitrogen-limiting conditions, *C. necator* H16 WT produced more PHB in gluconate than in glycerol in all growth phases analysed. Contrary to what was observed for *v6C6*, where more PHB was produced in glycerol than in gluconate. The maximum PHB production under nitrogen-limited conditions from glycerol was achieved at the early stationary phase for *v6C6*, the variant strain produced 19 % more PHB than *C. necator* H16 WT, what is more, this maximum PHB production was achieved in 28 h for *v6C6* against the 168 h for the WT strain; this represents a more than 6-fold improvement in the PHB volumetric productivity from glycerol for *v6C6*.

PHB production was also confirmed by staining the PHB granules from samples of *C. necator* H16 and *v6C6* in gluconate and glycerol under nitrogen-limiting conditions withdrawn at the early stationary phase with Nile red; the PHB granules were visualised using fluorescence microscopy (**Figure 5.15**).

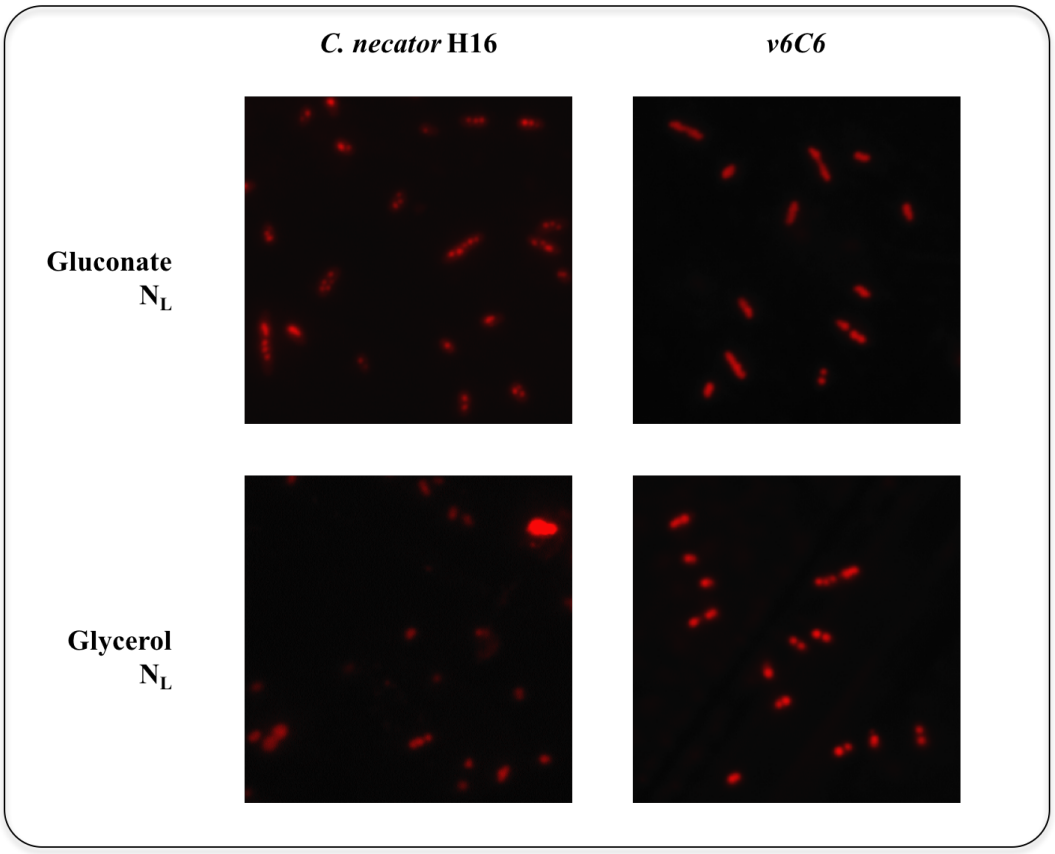


Figure 5.15 Fluorescence microscopy of Nile red-stained PHB granules. C. necator H16 and v6C6 cells samples were grown in 0.59 % (w/v) gluconate and 0.50 % (w/v) glycerol under nitrogen-limiting conditions. Abbreviations: N_L, nitrogen-limited conditions.

5.4.5 Crude glycerol assimilation of recombinant and variant strains of *C. necator* H16

After engineering, characterising and bioproducing PHB from engineered *C. necator* H16 strains using laboratory-grade glycerol (pure glycerol), the potential for industrial applications of the engineered strains was verified using crude glycerol obtained from a high-pressure fat splitting process. Comparative growth

studies of recombinant strains and *v6C6* were performed using crude glycerol as feedstock. Initially, different concentrations of crude glycerol were used (1.00 %, 2.00 %, and 4.00 % (v/v) of MSM with crude glycerol) (**Figure 5.16A**), where, as observed with pure glycerol studies, *v6C6* showed higher specific growth rates and shorter lag phase compared to *C. necator* H16 WT (**Table 5.6**), *v6C6* showed a specific growth rate of 0.220 h^{-1} compared to the specific growth rate of 0.048 h^{-1} for *C. necator* H16 in 4.00 % (v/v) crude glycerol (*i.e.*, 4.6-fold faster than the WT strain). After analysing the cell growth at different concentrations of glycerol, 4.00 % (v/v) crude glycerol was chosen for studying the growth of *C. necator* H16 recombinant strains (**Figure 5.16B**). The results of the cell growth study in crude glycerol with recombinant strains of *C. necator* H16 and *v6C6* showed higher specific growth rates and shorter lag phase compared to *C. necator* H16 WT (**Table 5.6**) as observed in previous results with pure glycerol, although the *v6C6* variant strain remained as the best-engineered strain in glycerol assimilation.

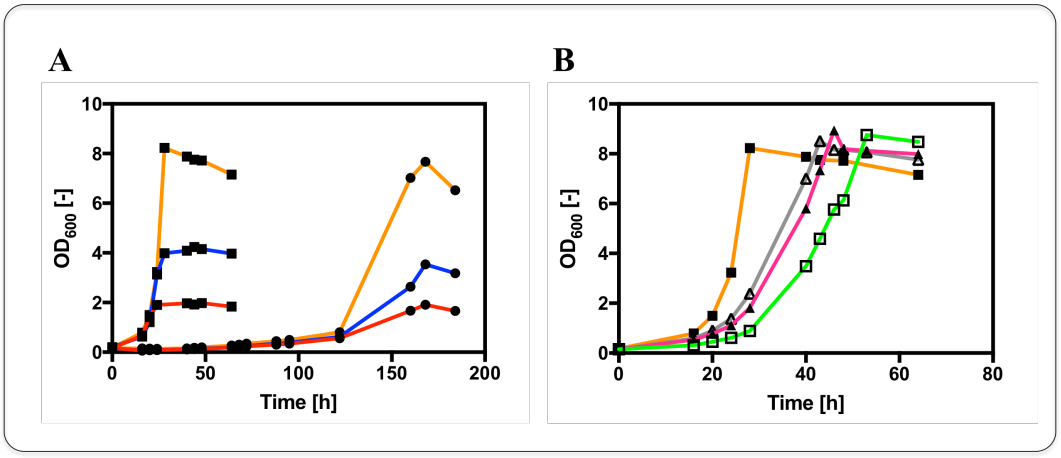


Figure 5.16 Cell growth of C. necator H16, v6C6 and recombinant strains in crude glycerol. ***A)*** *Growth curve of C. necator H16 (circles), and v6C6 (squares) in MSM with 1.00 % (v/v) crude glycerol (red lines), MSM with 2.00 % (v/v) crude glycerol (blue lines), and MSM with 4.00 % (v/v) crude glycerol (orange lines).* ***B)*** *Growth curve of v6C6 (square, orange line) in MSM with 4.00 % (v/v) crude glycerol, H16_glpK$_{H16}$ (open square, green line), H16_rbs1.glpKD$_{H16}$ (triangle, pink line) in MSM with 4.00 % (v/v) crude glycerol, and H16_rbs2.glpKD$_{H16}$ (open triangle, grey line) in MSM with 4.00 % (v/v) crude glycerol.*

The comparison table showing the different specific growth rates of all the recombinant and variant strains grown in crude glycerol are shown in **Table 5.6**, where also it was noted that the *v6C6* variant strain remained as the fastest growing crude glycerol-utilising strain obtained from these studies. The comparison on growth rates of the recombinant and *v6C6* engineered strains was performed only in 4.00 % (v/v) crude glycerol, as this concentration provided a similar OD_{600} of ≈ 8 when cells reached stationary phase to the OD_{600} of ≈ 8 also when cells reached the same growth phase when cells were grown in MSM with 0.59 % (w/v) gluconate or 0.50 % (w/v) glycerol.

Table 5.6 Specific growth rate of C. necator H16, v6C6 and recombinant strains in crude glycerol.

Strain	Carbon source	Specific growth rate [h^{-1}]
H16_WT	1.00 % (v/v) crude glycerol	0.023
H16_WT	2.00 % (v/v) crude glycerol	0.031
H16_WT	4.00 % (v/v) crude glycerol	0.048
v6C6	1.00 % (v/v) crude glycerol	0.170
v6C6	2.00 % (v/v) crude glycerol	0.220
v6C6	4.00 % (v/v) crude glycerol	0.220
H16_glpK$_{H16}$	4.00 % (v/v) crude glycerol	0.096
H16_rbs1.glpKD$_{H16}$	4.00 % (v/v) crude glycerol	0.095
H16_rbs2.glpKD$_{H16}$	4.00 % (v/v) crude glycerol	0.098

5.5 Conclusion

The plasmids described in this **Chapter 5** were constructed using the synthetic biology toolbox with the inducible and constitutive systems described in **Chapter 2** for the expression of glycerol-related proteins in *C. necator* H16, which were successfully transformed and expressed into *C. necator* H16 by electroporation with the optimised protocol method described in **Chapter 2**.

Both strategies –rational design (optimised and studied in **Chapter 2**) and evolutionary engineering (studied in **Chapter 4**)- used in this work were proven successful to unlock the glycerol utilisation ability of *C. necator* H16, where the best variant of *C. necator* H16 –*v6C6*- showed a significant cell growth improvement in glycerol utilisation with an 8-fold faster specific growth rate compared to the WT. The best glycerol-utilising variant *v6C6* also showed at least a 6-fold higher PHB volumetric productivity from glycerol compared to the WT. Although further studies –such as whole-genome sequencing- should be performed in order to fully understand the glycerol-utilising phenotype of the variant strains, the already identified $GlpK_{v6C6}$ represents an attractive candidate to be used for directed evolution to construct a better glycerol-utilising mutant strain. The variants isolated from evolutionary engineering strategy *via* adaptive evolution –*v6C6*, *v6F8*, and *v6G7*- are valuable starting points for developing a sustainable bioprocess that could use crude glycerol as feedstock for chemical production such as the biodegradable PHB bioplastics.

The fact that the best-engineered strain was derived from evolutionary engineering *via* adaptive evolution, even though the key glycerol enzymes were known and targeted in rational design, reflects the potential of evolutionary engineering tools and the challenges that the rational design engineering can have due to biological complexities.

CHAPTER 6

Concluding remarks and Future work

6 Concluding remarks and Future work

6.1 Concluding remarks

In **Chapter 1**, the need for expanding tools for chemical production using *C. necator* H16 as a cell factory, the current and potential bioproducts that can be synthesised in the strain, and the different carbon sources -including sustainable and cheap feedstock- were discussed. The compilation of the literature of *C. necator* H16 also contributed to understanding the natural metabolism of *C. necator* H16, and hence, the design of the experiments in this work was designed based on potential bioproducts that could be synthesised in engineered strains of *C. necator* H16. In order to use *C. necator* H16 as a cell factory, mainly two different strategies were reviewed: Rational design and Evolutionary engineering, which are the most common engineering tools used for chemical production in industry.

The first tool required to engineer *C. necator* H16 based on a rational design engineering approach was the developing of a more straightforward tool to transform *C. necator* H16 cells, which was achieved by the optimisation of the transformation method by electroporation for *C. necator* H16 described in **Chapter 2**. Due to bacterial conjugation –an alternative of transformation in *C. necator* H16- is a different technique compared to electroporation (the method proposed in this work), the comparison among these two methods in terms of efficiency is difficult, although, the key advantage with the method proposed in this work lies mainly in that this is a shorter and simpler method for transformation of *C. necator* H16, which only takes 2 days to be accomplished, while bacterial conjugation requires 4 to 5 days to be accomplished. The maximum transformation efficiency obtained for *C. necator* H16 was 3.86 x 10^5 cfu/μg DNA. Some parameters have been studied in this work have been investigated in previous publications (**Table 6.1**), although the combination of all the optimised parameters has not been published before. As observed in **Table 6.1**, Solaiman *et al*. (2010) previously reported a transformation efficiency of 470

cfu/μg DNA using the pBHR1plasmid, which was also used in this study, this represents a 10^3-fold improvement in transformation efficiency for the electroporation method proposed in this work.

Table 6.1 Transformation efficiency of *C. necator* H16 reported in literature.

Plasmid		**Cells**	**Electroporation condition**			
Name	Size (kb)	OD_{600}	Buffer	Voltage [kV cm^{-1}]	Transformation eff. [cfu//μg DNA]	References
pKT230	11.9	0.8	10 % (v/v) glycerol	11.5	8.00×10^1	Park *et al*., 195
pBHR1	5.3	0.6	0.2 M sucrose	11.5	3.86×10^5	This work (Tee *et al.*, 2017)
pBHR1	5.3	n.r.	0.3 M sucrose	25	4.70×10^2	Solaiman *et al*., 2010
pKK-II 89-6	7.8	n.r.	0.3 M sucrose	25	1.20×10^2	Solaiman *et al*., 2010
pBS29-P2-gfp	8.7	n.r.	0.3 M sucrose	25	4.04×10^3	Solaiman *et al*., 2010

Abbreviations: n.r.: not reported. Adapted from Tee *et al*., 2017.

After optimising the transformation method by electroporation for *C. necator* H16, the construction of a synthetic biology toolbox tailored for *C. necator* H16 with two different promoters (P_{BAD} and P_{j5}) was performed, where all the plasmids constructed were based on a broad host range plasmid vector (pBBR1MCS-1), as this is a *C. necator* H16-compatible vector. The plasmids were constructed for protein expression for metabolic engineering purposes.

Then, in **Chapter 3**, the applicability of the synthetic biology toolbox was tried out for the expression of heterologous biotin operons in *E. coli* BW25113 ΔbioC and *C. necator* H16 for pimelate production. Although most of the plasmids constructed in this work using the synthetic biology toolbox –either with a constitutive or inducible gene expression- mediated successfully the expression of native and heterologous genes in *C. necator* H16, the unsuccessful expression of all the genes of the biotin operons in *C. necator* H16 (and in *E. coli* BW25113 ΔbioC) could be attributed to proteolysis as a mechanism of defence of the strain to protect it from cellular toxicity, although other studies could be tested to corroborate the expression of all the genes of the extended biotin operons (such as

GeLC-MS/MS), as well as study further the inducible system by optimising the induction point and concentration.

In **Chapter 4**, a second strategy to engineer *C. necator* H16 was studied: evolutionary engineering. A method of directed evolution *via* random mutagenesis was developed to introduce random mutations in the genome of *C. necator* H16. The optimised method was then applied to try to understand the unknown biotin biosynthesis pathway of *C. necator* H16 by selecting mutant strains that could grow in the presence of a biotin analogue. The results showed that the selected mutants after screening selection were not overproducing biotin as expected, this might be due to the best mutants obtained after two rounds of directed evolution *via* random mutagenesis still did not show a significantly higher fold in cell growth compared to the WT, and hence the biotin production between *C. necator* H16 and the mutants was relatively the same. Nevertheless, the development of the EMS random mutagenesis method for *C. necator* H16 is a potential engineering tool for the strain, as it can be used for engineering *C. necator* H16 for a wide variety of purposes, as directed evolution *via* random mutagenesis is becoming increasingly crucial for metabolic engineering of complex phenotypes.

The study and optimisation of some of the engineering tools used in rational design and evolutionary engineering studied in previous chapters contributed to engineer *C. necator* H16 strains for chemical production (bioplastics) using waste stream (crude glycerol) as feedstock in **Chapter 5**, where both recombinant and evolved cells of *C. necator* H16 were generated with the aid of recombinant strain engineering and adaptive evolution for improvement of glycerol utilisation. The parallel use of both strategies demonstrated the potential of combinatorial approaches, where the rational design approach showed that the *glpK* gene was essential for improving glycerol utilisation, and the evolutionary engineering approach also corroborated that the GlpK protein was a key enzyme for glycerol assimilation; although by natural adaptive evolution, the *C. necator* H16 strain improved better glycerol utilisation than the recombinant strains; this reflects that cells are complex systems, and thus, rational design engineering sometimes can be limited by this factor.

C. necator H16 is a versatile strain that can grow and naturally produce PHB on gluconate, fructose, fatty acids, and plant oils, but it cannot metabolise glucose, xylose or arabinose, and it shows a very poor growth in glycerol, an abundant waste stream in the bioindustry. Hence, the expansion of the carbon substrate range is vital for wider applications. To date, engineered strains of *C. necator* H16 have been reported to be able to grow in glucose either using strategies of recombinant strain engineering (Sichwart *et al*., 2011) or evolutionary engineering (Franz *et al*., 2012). The same has been reported for related species of *C. necator* to grow in glycerol; it has been reported previously that other wild-type strains apart from *C. necator* H16 (Fukui *et al*., 2014) -*C. necator* JMP134 (DSMZ 4058) and DSM 545- were able to grow and to accumulate PHB on glycerol (Mothes *et al*., 2007; Cavalheiro *et al*., 2009).

In **Table 6.2**, a comparison of the specific growth rates on glycerol for *C. necator* is shown. Fukui *et al*., (2014) produced PHB from glycerol with recombinant strains of *C. necator* H16 harbouring $glpKF_{Ec}$ genes, where the recombinant strains were able to grow on glycerol much faster than the wild-type strain, but the specific growth rates (0.15 h^{-1} on 0.50 % (w/v) glycerol) were lower compared to other preferred carbon sources such as fructose (0.28 h^{-1} on 0.50 % (w/v) fructose). Mothes *et al*. (2007), produced also PHB from crude glycerol in *Paracoccus denitrificans* and *C. necator* JMP134, both strains accumulated PHB from pure grade glycerol or crude glycerol, although PHB accumulation was 65 % of CDW for *P. denitrificans*, while a higher PHB accumulation of 70 % of CDW was obtained for *C. necator* JMP134. The specific growth rate of *C. necator* JMP134 in glycerol was 0.13 h^{-1}, which is also lower to the specific growth rate observed on other preferred carbon sources such as gluconate (0.29 h^{-1} on 0.59 % (w/v) gluconate, reported in this thesis). Higher specific growth rates on glycerol and crude glycerol were obtained in this thesis work, where *v6C6* showed a specific growth rate of 0.20 h^{-1} on 0.50 % (w/v) glycerol, and 0.22 h^{-1} on 4.00 % (v/v) crude glycerol.

Table 6.2 Specific growth rates on glycerol in related species of *C. necator*.

Strain	Specific growth rate [h^{-1}]	[Glycerol]	Reference
C. necator JMP134	0.13	n.r.	Mothes *et al*., 2007
C. necator H16	0.15	0.50 % (w/v)	Fukui *et al*., 2014
C. necator DSM 545	0.17	0.50 % (w/v)	Cavalheiro *et al*., 2009
v6C6	0.20	0.5 % (w/v)	This study

Abbreviations: n.r.: not reported.

One of the main contributions of this work, is that a higher specific growth rate was obtained in an engineered strain of *C. necator* H16 (*v6C6*), which translates to a shorter time of PHB production. Engineered strains of *C. necator* H16 can reach stationary phase is a shorter period of time (28 h) compared to other reported publications (*e.g*., 50 h in Fukui *et al*., 2014). Higher specific growth rates are preferred as they are advantageous with respect to reducing production costs. On the other hand, for the recombinant strain engineering strategy followed in this study, the contribution of the individual gene $glpK_{H16}$, and $glpKD_{H16}$ genes to the enhancement in glycerol assimilation ability by the *C. necator* H16 strain have not been investigated before.

Although in this work, PHB production was performed at low-scale in flasks for research purposes, it is well known that high-cell density fed-batch cultivation is the strategy mainly used to obtain higher PHB yields (Mothes *et al*., 2007; Cavalheiro *et al*., 2009, and 2012). Fed-batch are usually performed in 2 L reactor, where cultivation conditions such as gas composition and aeration rate have been found to be important to achieve high specific growth rates and high cell densities. Cavalheiro *et al*. (2009) maintained high dissolved oxygen concentration (DOC) levels in order to avoid oxygen limited growth using glycerol as carbon source, they demonstrated that high aeration rates and an inlet air stream containing pure oxygen were key factor to achieve high cell densities of approximately 45-50 g_{DW}/L.

C. necator H16 remains as a good candidate to be used for PHB production as this strain has been extensively studied for almost 60 years (Schlegel *et al*., 1961a, b) and is the best studied bacterium producer of PHB. The complete published

genome sequence of the strain allows scientist to look at the genetic potential of the strain, moreover, *C. necator* H16 is not limited to the production of PHB, but it can produce a wider range of copolyesters derived from PHB such as P(3HB-*co*-3HV). Among other milestones, *C. necator* H16 is capable of accumulating PHB to over 80% (w/w) of its cell dry weight (CDW) (reviewed in Reinecke *et al*., 2009), reinforcing this strain as a candidate for chemical production. In **Table 6.3**, results obtained using different microorganisms and carbon sources are given, where it can be observed that related species of *C. necator* H16 always produced >60 % of P(3HB)/CDW.

Table 6.3 P(3HB) production in different microorganisms and carbon substrates.

Strain	Carbon source	P(3HB)/ CDW %	Reference
E. coli CGSC 4401	Whey	87	Ahn *et al*., 2001
E. coli XL1-Blue (pSYL105)	Glucose	65	Kahar *et al.*, 2005
E. coli XL1-Blue (pJRDTrcphaCABRe)	Glucose	80	Kahar *et al.*, 2005
Methylobacterium rhodesianum MB 126	Glycerol	50	Bormann *et al*., 1999
A. lactus	Sucrose	50	Yamane *et al*., 1996
C. necator DSM 545	Glycerol	62	Cavalheiro *et al*., 2009
C. necator DSM 11348	Glucose and peptone	78	Bormann *et al*., 1998
C. necator NCIMB 11599	Glucose	76	Kim *et al*., 1994
C. necator JMP 134	Glycerol	70	Mothes *et al*., 2007

Adapted from Cavalheiro *et al*., 2009.

In this PhD thesis, the results obtained from the development of an efficient method for *C. necator* H16 transformation, a synthetic biology toolbox tailored for *C. necator* H16, a EMS random mutagenesis method for *C. necator* H16, and adaptive evolution studies in *C. necator* H16, have contributed to the expansion of engineering tools for chemical production in *C. necator* H16, methods that can be used in the bioindustry to convert waste stream such as crude glycerol into useful chemicals.

6.2 Future work

There are further investigations that should be performed to improve and to fully understand the results obtained in this work, which could enrich further either the optimisation of the engineering tools developed in this work or their applications.

Some of the future work that could be performed is mentioned in the following bullet points:

Chapter 2. Recombinant strain engineering: Development of a transformation method and a synthetic biology toolbox for metabolic engineering in *Cupriavidus necator* H16

- The synthetic biology toolbox can be further extended by replacing the P_{BAD} and P_{j5} promoters with other promoters of variable strengths that could be useful for engineering *C. necator* H16 for biomanufacturing applications; also, other genetic elements for controlling gene expression could be included in the synthetic biology toolbox, such as the tested RBSs studied in Alagesan *et al.* (2018). This due to similar to using different promoters for increasing protein expression, the RBS has a significant impact in the protein synthesis. The use of RBS with variable strengths could help to improve the levels of protein synthesis.

Chapter 3. **Metabolic engineering: Applicability of the synthetic biology toolbox by the expression of heterologous extended biotin operons in *Cupriavidus necator* H16**

- The inducible system of the synthetic biology toolbox constructed for *C. necator* H16 was applied for the expression of heterologous biotin operons in *C. necator* H16, were the genes occupying the first position of the operons were not visible in the SDS-PAGE. Thus, one of the first point calls of re-investigation would be to optimise the expression of the biotin operons. One way to explore this could be by 1) performing a dose-dependent induction of P_{BAD} promoter, using different L-arabinose concentrations to tune the expression of the biotin operons in *E. coli* and in *C. necator* H16, or 2) use a different inducible promoter such as the anhydrotetracycline-inducible P_{tet} promoter, which is also widely used to tune the expression of gene clusters in *C. necator* H16 (Johnson *et al*.,

2018), or the tunable L-rhamnose inducible system which has been proven to be suitable for heterologous protein expression in *C. necator* H16, where variations in the induction time can be studied for the expression of toxic proteins, which could be the case of the extended biotin operons (Sydow *et al*., 2017).

- Also, the study of the function of single genes of the extended biotin operons could be performed to check for BioC-like function.

Chapter 4. Directed evolution: Development of directed evolution *via* random mutagenesis tool for *Cupriavidus necator* H16 to understand its biotin biosynthesis pathway

- For the applicability of the EMS random mutagenesis method for *C. necator* H16, *C. necator* H16 cells were randomly mutagenised with the EMS chemical mutagen using a biotin analogue (TVA) as selection pressure to identify mutants that were able to grow in the presence of the biotin analogue, these mutants were expected to overproduce biotin in order to survive in the presence of TVA. Nevertheless, after biotin quantitation of the mutants isolated, the biotin production in *C. necator* H16 WT and mutant strains was somewhat the same. Thus, if in further experiments a mutant with a higher fold (*e.g.* 10-fold higher) in cell growth performance in the presence of TVA is obtained, and still the biotin content between the new isolated mutant strains and the WT is the same, one option to investigate could be to use a different biotin analogue as selection pressure such as azelaic acid.

Chapter 5. Rational design and Evolutionary engineering: Engineering *Cupriavidus necator H16* to convert waste stream into useful chemicals by enhancing glycerol assimilation

- The evolutionary engineering approach provided exciting results which require further investigations, some of the future work that could be

performed in order to understand better the impact of the mutations generated during adaptive evolution could be:

- Whole-genome sequencing of variant strains *v6C6*, *v6F8*, and *v6G7*, as this would help to identify other possible mutations in the genome that could be responsible for the enhancement of the glycerol-utilising phenotype observed in the variant strains, and to characterise variant strains fully.
- Plasmid-based protein expression of $GlpK_{v6C6}$ (GlpK W480S mutant) into *C. necator* H16 to check if the expression of this protein in the WT could improve glycerol assimilation with the same –or better- growth performance (*i.e.*, a shorter lag phase than when $GlpK_{H16}$ is overexpressed) as it does in *v6C6*.
- The $GlpK_{v6C6}$ (GlpK W480S mutant) also provides an interesting candidate to be used for directed evolution, which could generate a better glycerol-utilising mutant.

7 REFERENCES

Aboulnaga E, Zou H, Selmer T, and Xian M. (2018) Development of a plasmid-based, tunable, tolC-derived expression system for application in *Cupriavidus necator* H16. *Journal of Biotechnology*, 274: 15-27.

Ahn WS, Park SJ, Lee SY. (2001) Production of poly(3-hydroxybutyrate) from whey by cell recycle fed-batch culture of *recombinant Escherichia coli. Biotechnol Lett*, 23: 235–40.

Ahrens W, and Schlegel HG. (1972) Carbon-dioxide requiring mutants of *Hydrogenomonas eutropha* strain H16. Growth and CO_2-fixation. *Archiv Fur Mikrobiologie*, 85(2): 142–152.

Alagesan S, Hanko EKR, Malys N, Ehsaan M, Winzer K, and Minton NP. (2018) Functional Genetic Elements for Controlling Gene Expression in *Cupriavidus necator* H16. *App Environ Microbiol*, 84(19): 1-8.

Albuquerque MGE, Eiroa M, Torres C, Nunes BR, Reis MAM. (2007) Strategies for the development of a side stream process for polyhydroxyalkanoate (PHA) production from sugar cane molasses. *Journal of Biotechnology*, 130(4): 411-421.

Allaman I, Belanger M, and Magistretti P. (2015) Methylglyoxal, the dark side of glycolysis. *Front Neurosci*, 9: 23.

Anderson MJ, DeLabarre B, Raghunathan A, Palsson BO, Brunger AT, and Quake SR. (2007) Crystal structure of a hyperactive *Escherichia coli* glycerol kinase mutant Gly230 --> Asp obtained using microfluidic crystallization devices. *Biochemistry,* 46: 19.

Andreeßen B, Lange AB, Robenek H, and Steinbüchel A. (2009) Conversion of glycerol to poly(3-hydroxypropionate) in recombinant *Escherichia coli*. *Appl Environ Microbiol,* 76: 622–626.

Ashby RD, Solaiman DKY, Strahan GD, Zhu C, Tappel RC, and Nomura CT. (2012) Glycerine and levulinic acid: Renewable co-substrates for the fermentative synthesis of short-chain poly(hydroxyalkanoate) biopolymers. *Bioresource Technology*, 118: 272–80.

Asif S, Mohsin H, Tanvir R, and Rehman Y. (2017) Revisiting the Mechanisms Involved in Calcium Chloride Induced Bacterial Transformation. *Front. Microbiol*. [online]. 8(2169) [Viewed on 15 March 2019]. Available from: doi: 10.3389/fmicb.2017.02169.

Atsumi S, Higashide W, and Liao JC. (2009) Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nature Biotechnology*, 27(12): 1177–U142.

Attwood PV, and Wallace JC. (2002) Chemical and catalytic mechanisms of carboxyl transfer reactions in biotin-dependent enzymes. *Acc Chem Res,* 35, 113-120.

Bachmann H, Starrenburg MJ, Molenaar D, van Hylckama Vlieg JE, and Kleerebezem M. (2012) Microbial domestication signatures of *Lactococcus lactis* can be reproduced by experimental evolution. *Genome Res*, 22: 115–124.

Barker DF, Campbell AM. (1980) Use of *bio-lac* fusion strains to study regulation of biotin biosynthesis in *Escherichia coli*. *J Bacteriol*, 143: 789–800.

Barrick J, Yu D, Yoon S, Jeong H, Oh T, Schneider D, Lenski R, and Kim J. (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature,* 461: 1243–1247.

Berkovitch F, Nicolet Y, Wan JT, Jarrett JT, and Drennan CL. (2004) Crystal structure of biotin synthase, an *S*-adenosylmethionine-dependent radical enzyme. *Science*, 303(5654): 76-79.

Bhat T, Khan A, and Parveen S. (2005) Comparative analysis of meiotic abnormalities induced by gamma rays, EMS and MMS in Vicia faba L. *J. Ind. Bot. Soc*. 84: 45-48.

Bi C, Su P, Mueller J, Yeh YC, Chhabra SR, Beller HR, Singer S, and Hillson N. (2013) Development of a broad-host synthetic biology toolbox for *Ralstonia eutropha* and its application to engineering hydrocarbon biofuel production. *Microbial Cell Factories*, 12: 107.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, and Schwede T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* [online], 42: W252-258. [Viewed 15 March 2019]. Available from: doi:10.1093/nar/gku340.

Bloom J, Arnold F. (2009) In the Light of Directed Evolution: Pathways of Adaptive Protein Evolution. In: National Academy of Sciences. Ed. *National Academies Press*, USA. p. 256-269.

Bormann EJ, Leibner M, Roth M, Beer B, and Metzner K. (1998) Production of polyhydroxybutyrate by *Ralstonia eutropha* from protein hydrolysates. *Appl Microbiol Biotechnol*, 50: 604–607.

Bormann EJ, and Roth M. (1999) Production of polyhydroxybutyrate by *Methylobacterium rhodesianum* and *Ralstonia eutropha* in media containing glycerol and casein hydrolysates. *Biotechnol Lett*, 21: 1059–63.

Bower S, Perkins JB, Pero J. (1996b) Biotin biosynthesis in *Bacillus subtilis*. United States Patent 6,768,180.

Bower S, Perkins JB, Yocum RR, Howitt CL, Rahaim P, and Pero J. (1996a) Cloning, sequencing, and characterization of the *Bacillus subtilis* biotin biosynthetic operon. *J. Bacteriol,* 178: 4122–4130.

Bower S, Perkins JB, Yocum RR, and Pero JG. (2001) Biotin biosynthesis in *Bacillus subtilis*. United States Patent 6,303,377 B1.

Bowien B, and Kusian B. (2002) Genetics and control of CO_2 assimilation in the chemoautotroph *Ralstonia eutropha*. *Archives of Microbiology*, 178(2): 85–93.

Brigham CJ, Budde CF, Holder JW, Zeng Q, Mahan AE, Rha C, and Sinskey A. (2010) Elucidation of beta-Oxidation Pathways in *Ralstonia eutropha* H16 by Examination of Global Gene Expression. *Journal of Bacteriology*, 192(20): 5454-64.

Brigham C, Zhila N, Shishatskaya E, Volova T, and Sinskey A. (2012) Chapter 17: Manipulation of *Ralstonia eutropha* Carbon Storage Pathways to Produce Useful Bio-Based products. In: Wang X. Ed *Reprogramming Microbial Metabolic Pathways, Subcellular Biochemistry*, p. 343-366.

Byrom D. (1987) Polymer synthesis by microorganisms - technology and economics. *Trends in Biotechnology*, 5(9): 246-50.

Byrom D. (1992) Production of poly-beta-hydroxybutyrate -poly-beta-hydroxyvalerate copolymers. *Fems Microbiology Letters*, 103(2-4): 247-250.

Cakar ZP, Alkim C, Turanli B, Tokman N, Akman S, Sarıkaya M, Tamerler C, Benbadis L, and François JM. (2009) Isolation of cobalt hyper-resistant mutants of *Saccharomyces cerevisiae* by *in vivo* evolutionary engineering approach. *J Biotechnol*, 143(2): 130-138.

Cakar ZP, Seker UOS, Tamerler C, Sonderegger M, and Sauer U. (2005) Evolutionary engineering of multi-stress resistant *Saccharomyces cerevisiae*. *FEMS Yeast Res,* 5: 569–578.

Cakar ZP, Turanli-Yildiz B, Alkim C, and Yilmaz U. (2011) Evolutionary engineering of *Saccharomyces cerevisiae* for improved industrially important properties. *FEMS Yeast*, 12: 171–182.

Cameron DC, Altaras NE, Hoffman ML, and Shaw AJ. (1998) Metabolic engineering of propanediol pathways. *Biotechnology Progress*, 14(1): 116–25.

Cartman S, Combe J, Foste A, and Kennedy J. (2018) Materials and methods for directing carbon flux and increased production of carbon based chemicals. United States Patent Application 20180023104.

Cavalheiro J, Almeaida M, Grandfils C, Da Fonseca M. (2009) Poly(3-hydroxybutyrate) production by *Cupriavidus necator* using waste glycerol. *Process Biochemistry,* 44(5): 509–515.

Cavalheiro J, Raposo RS, de Almeida MC, Cesário MT, Sevrin C, Grandfils C, and Da Fonseca MMR. (2012) Effect of cultivation parameters on the production of poly(3-hydroxybutyrate-co-4-hydroxybutyrate) and poly(3-hydroxybutyrate-4-hydroxybutyrate-3-hydroxyvalerate) by *Cupriavidus necator* using waste glycerol. *Bioresource Technology*, 111: 391-397.

Choi JC, Shin HD, Lee YH. (2003) Modulation of 3-hydroxyvalerate molar fraction in poly(3-hydroxybutyrate-3-hydroxyvalerate) using *Ralstonia eutropha* transformant co-amplifying phbC and NADPH generation-related *zwf* genes. *Enzyme and Microbial Technology*, 32(1):178-185.

Choi WJ, Hartono MR, Chan WH, Yeo SS. (2011) Ethanol production from biodiesel-derived crude glycerol by newly isolated *Kluyvera cryocrescens*. *App Microbiol Biotechnol,* 89: 1255–1264.

Chu S. (2009) Carbon capture and sequestration. *Science*, 325(5948): 1599.

Cobb R, Chao R, and Zhao H. (2013) Directed Evolution: Past, Present and Future. *AIChE Journal*, 59(5): 1432-1440.

Connor M, Cann AF, and Liao J. (2010) 3-Methyl-1-butanol production in *Escherichia coli*: random mutagenesis and two-phase fermentation. *App Microbiol Biotechnol*, 86: 1155-1164.

Conrad TM, Frazier M, Joyce AR, Cho BK, Knight EM, Lewis NE, Landick R, Palsson B. (2010) RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. *Proc Natl Acad Sci USA*, 107: 20500–20505.

Crameri A, Dawes G, Rodriguez E Jr, Silver S, Stemmer WPC. (1997) Molecular evolution of an arsenate detoxification pathway by DNA shuffling. *Nat Biotechnol*, 15: 436–438.

Cronan J. (2001) The biotinyl domain of *Escherichia coli* acetyl-CoA carboxylase. Evidence that the "thumb" structure is essential and that the domain functions as a dimer. *J Biol Chem*, 276: 37355–37364.

Cronan J. (2014) Biotin and Lipoic Acid: Synthesis, Attachment, and Regulation. *EcoSal Plus* [online], 6(1): 2324. [Viewed on 15 March 2019]. Available from: doi:10.1128/ecosalplus.ESP-0001-2012.

Cronan J, and Lin S. (2011) Synthesis of the α,ω-dicarboxylic acid precursor of biotin by the canonical fatty acid biosynthetic pathway. *Current Opinion in Chemical Biology*, 15(3): 407-413.

Cronan J, and Reed KE. (2000) Biotinylation of proteins in vivo: a useful posttranslational modification for protein analysis. *Methods Enzymol*, 326: 440-458.

Cupples C, and Miller J. (1989) A set of *lacZ* mutations in *Escherichia coli* that allow rapid detection of each of the six base substitutions. *Proc. Natl. Acad. Sci. USA*, 86: 5345-5349.

Cymbalyuk E, Chernomordik L, Broude N, and Chizmadzhev Y. (1988) Electro-stimulated transformation of *E. coli* cells pretreated by EDTA solution. *FEBS LETTERS*, 234(1): 203-207.

Dalby PA. (2011) Strategy and success for the directed evolution of enzymes. *Curr Opin Struct Biol*, 21: 473–480.

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

Delamarre SC, and Batt CA. (2006) Comparative study of promoters for the production of polyhydroxyalkanoates in recombinant strains of *Wautersia eutropha*. *Appl. Microbiol. Biotechnol*., 71: 668–679.

Delli-Bovi T, Spalding M, and Prigge S. (2010) Overexpression of biotin synthase and biotin ligase is required for efficient generation of sulfur-35 labeled biotin in *E. coli. BMC Biotechnology*, 10(73): 1-11.

Dennis D, McCoy M, Stangl A, Valentin HE, Wu Z. (1998) Formation of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) by PHA synthase from *Ralstonia eutropha*. *Journal of Biotechnology*, 64(2-3): 177-186.

DeTitta GT, Edmonds JW, Stallings W, and Donohue J. (1976) Molecular structure of biotin. Results of two independent crystal structure investigations. *J Am Chem Soc,* 98: 1920-1926.

Dobson R, Gray V, and Rumbold K. (2012) Microbial utilization of crude glycerol for the production of value-added products. *J Ind Microbiol Biotechnol*, 39: 217–226.

Dower WJ, Miller JF, and Ragsdale CW. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res*., 16(13): 6127.

Dragosits M, and Mattanovich D. (2013) Adaptive laboratory evolution – principles and applications in biotechnology. *Microbial Cell Factories*, 12: 64.

Drake JW. (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA,* 88: 7160–7164.

Drummond ML, Cundari TR, and Wilson AK. (2012) Protein-based carbon capture: progress and potential. *Greenhouse Gases-Science and Technology*, 2(4): 223–238.

Du GC, Chen J, Yu J, and Lun SY.(2001) Feeding strategy of propionic acid for production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) with *Ralstonia eutropha. Biochemical Engineering Journal*, 8(2): 103–110.

Du J, Shao Z, Zhao H. (2011) Engineering microbial factories for synthesis of value-added products. *J Ind Microbiol Biotechnol*, 38: 873– 890.

Ducat DC, Way JC, and Silver PA. (2011) Engineering cyanobacteria to generate high-value products. *Trends in Biotechnology*, 29(2): 95–103.

Durnin G, Clomburg J, Yeate Z, Alvarez PJJ, Zygourakis K, Campbell P, Gonzalez R. (2009) Understanding and Harnessing the Microaerobic Metabolism of Glycerol in *Escherichia coli*. *Biotechnology and Bioengineering*, 103(1): 148-161.

Eisenberg M. (1975) Mode of Action of α-Dehydrobiotin, a Biotin Analogue. *J. Bacteriol*. 123(1): 248-254.

Ewering C, Heuser F, Benoelken JK, Braemer CO, Steinbüchel A. (2006) Metabolic engineering of strains of *Ralstonia eutropha* and *Pseudomonas putida* for biotechnological production of 2-methylcitric acid. *Metabolic Engineering*, 8(6): 587-602.

Ezeji TC, Qureshi N, and Blaschek HP. (2007) Bioproduction of butanol from biomass: from genes to bioreactors. *Current Opinion in Biotechnology*, 18(3): 220–227.

Finkenwirth F, Kirsch F, and Eitinger T. (2014) A versatile *Escherichia coli* strain for identification of biotin transporters and for biotin quantification. *Bioengineered*, 5(2): 129-132.

Franz A, Rehner R, Kienle A and Grammel H. (2012) Rapid selection of glucose-utilizing variants of the polyhydroxyalkanoate producer *Ralstonia eutropha* H16 by incubation with high substrate levels. *Letters in Applied Microbiology*, 54(1): 45–51.

Foster P. (1991) *In Vivo* Mutagenesis. *Methods Enzymol*, 204: 114-125.

Friedrich B, Hogrefe C, and Schlegel HG. (1981) Naturally occurring genetic transfer of hydrogen-oxidizing ability between strains of *Alcaligenes eutrophus. J. Bacteriol*., 147: 198–205.

Fukui T, Abe H, and Doi Y. (2002) Engineering of *Ralstonia eutropha* for Production of Poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from Fructose and Solid-State Properties of the Copolymer. *Biomacromolecules*, 3: 618-624.

Fukui T, Mukoyama M, Orita I, and Nakamura S. (2014) Enhancement of glycerol utilization ability of *Ralstonia eutropha* H16 for production of polyhydroxyalkanoates. *App Microbiol and Biotechnol*, 98(17): 7559-68.

Gai CS, Lu J, Brigham CJ, Bernardi AC, and Sinskey AJ. (2014) Insights into bacterial CO_2 metabolism revealed by the characterization of four carbonic anhydrases in *Ralstonia eutropha* H16. *AMB Express*, 4(1): 2.

Gloeckler R, Ohsawa I, Speck D, Ledoux C, Bernard S, Zinsius M, Villeval D, Kisou T, Kamogawa K, and Lemoine Y. (1990) Cloning and characterization of the *Bacillus sphaericus* genes controlling the bioconversion of pimelate into dethiobiotin. *Gene*, 87 (1): 63-70.

Goh E-B, Baidoo EEK, Keasling JD, Beller HR. (2012) Engineering of Bacterial Methyl Ketone Synthesis for Biofuels. *Applied and Environmental Microbiology*, 78(1): 70-80.

Goldberg MW, and Sternbach L. (1949) United States Patent 2,489,237.

Gonzalez A, and Bell G. (2013) Evolutionary rescue and adaptation to abrupt environmental change depends upon the history stress. *Philos Trans R Soc Lond B Biol Sci*, 368(1610): 20120079.

González-Pajuelo M, Andrade JC, and Vasconcelos I. (2004) Production of 1,3-propanediol by *Clostridium butyricum* VPI 3266 using a synthetic medium and raw glycerol. *J Ind Microbiol Biotechnol,* 31: 442–446.

González-Villanueva Miriam, Tee KL, Staniland P, Staniland J, Savil I, and Wong TS. (2018) Unlocking glycerol utilization in *Cupriavidus necator* H16 for bioplastic synthesis (submitted).

Grabherr R, Nilsson E, Striedner G, and Bayer K. (2002) Stabilizing plasmid copy number to improve recombinant protein production. *Biotechnol Bioeng*, 77: 142–147.

Greene E, Codomo C, Taylor N, Henikoff JG, Till BJ, Reynolds SH, Enns LC, Burtner C, Johnson JE, Odden AR, Comai L, and Henikoff S. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics* 164(2): 731–740.

Grothe E, Moo-Young M, Chisti Y. (1999) Fermentation optimization for the production of poly(beta-hydroxybutyric acid) microbial thermoplastic. *Enzyme and Microbial Technology*, 25(1-2): 132-141.

Grover LW, Holden HM, and St Maurice M. (2012) The enzymes of biotin dependent CO_2 metabolism: What structures reveal about their reaction mechanisms. *Protein Sci*., 21(11): 1597-1619.

Gruber S, Hagen J, Schwab H and Koefinger P. (2014) Versatile and stable vectors for efficient gene expression in *Ralstonia eutropha* H16. *Journal of Biotechnology*, 186: 74–82.

Grzesiuk E, Janion C. (1993) Some aspects of EMS-induced mutagenesis in *Escherichia coli. Mutation Research*, 297: 313-321.

Guan Y, Zhu Q, Huang D, Zhao S, Jan Lo L, and Peng J. (2015) An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide. *Scientific Reports*. *Nature Publishing Group*, 5(1): 13370.

Hall B. (1998) Adaptive mutagenesis: a process that generates almost exclusively beneficial mutations. *Genetica*, 102/103: 109-125.

Hall BG, Acar H, Nandipati A, and Barlow M. (2013) Growth rates made easy. *Molecular Biology and Evolution*, 31(1): 232-238.

Hanahan D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166: 557–580.

Hanka L, and Martin DG. (1942) Antibiotic α-methyldethiobiotin, α-methylbioti, and their esters. United States Patent Office 3,658,837.

Hatti-Kaul R, Tornvall U, Gustafsson L, and Borjesson P. (2007) Industrial biotechnology for the production of bio-based chemicals - a cradle-to-grave perspective. *Trends in Biotechnology*, 25(3): 119–24.

He Q, McNutt J, and Yang J. (2017) Utilization of the residual glycerol from biodiesel production for renewable energy generation. *Renewable & Sustainable Energy Reviews,* 71: 63-76.

Hong A, Cheng K, Peng F, Zhou S, Sun Y, Lui C, and Lui D. (2009) Strain isolation and optimization of process parameters for bioconversion of glycerol to lactic acid. *J Chem Technol Biotechnol,* 84: 1576–1581.

Hoshino T, Akimufi F, and Masaaki Y. (1999) Process for the production of d-Biotin. United States Patent 5,922,581.

Hua Q, Joyce AR, Palsson B, Fong SS. (2007) Metabolic characterization of *Escherichia coli* strains adapted to growth on lactate. *Appl Environ Microbiol*, 73: 4639–4647.

Ibarra R, Edwards J, Palsson B. (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420: 186–189.

Ifuku O, Haze S, Kishimoto J, Koga N, Yanagi M, and Fukushima S. (1993) Sequencing Analysis of Mutation Points in the Biotin Operon of Biotin-overproducing *Escherichia coli* Mutants. *Bioscience, Biotechnology, and Biochemistry*, 57(5): 760-765.

Ifuku O, Miyaoka H, Koga N, Kishimoto J, Haze S, Wachi Y, and Kajiwara M. (1994) Origin of carbon atoms of biotin. 13C-NMR studies on biotin biosynthesis in *Escherichia coli*. *Eur. J. Biochem.,* 220: 585–591.

Ishizaki A, Tanaka K, and Taga N. (2001) Microbial production of poly-D-3-hydroxybutyrate from CO2. *Applied Microbiology and Biotechnology*, 57(1–2): 6-12.

Izumi Y, Fukuda H, Tani Y, and Ogata K. (1977) Action of 5-(2-thienyl)valeric acid as a biotin antagonist. *Biochemia et Biophysica Acta*, 499: 315-317.

Izumi Y, Fukuda H, Tani Y, and Ogata K. (1978) The Mode of Action of 5-(2-Thienyl)-valeric Acid on Biotin Biosynthesis. Agric. Biol. Chem., 42(3): 579-584.

Jajesniak P, Omar-Ali HEM and Wong TS. (2014) Carbon dioxide capture and utilization using biological systems: Opportunities and challenges. *Journal of Bioprocessing and Biotechniques*, 4(3).

Jitrapakdee S, and Wallace JC. (2003) The biotin enzyme family: conserved structural motifs and domain rearrangements. *Curr Protein Pept Sci*, 4: 217-229.

Johnson AO, Gonzalez Villanueva M, Wong TS. (2015) Engineering *Cupriavidus necator* for chemical production. In: Industrial Processes and Nanotechnology. 10. Studium Press LLC, USA. p. 255-292.

Johnson AO, Gonzalez Villanueva M, Tee KL, and Wong TS. (2018) An engineered constitutive promoter set with broad activity range for *Cupriavidus necator* H16. *ACS Synth* [online]. [Viewed 15 March 2019]. Available from: doi: 10.1021/acssynbio.8b00136.

Johnson DT, and Taconi KA. (2007) The glycerin glut: Options for the value-added conversion of crude glycerol resulting from biodiesel production. *Environ. Progress*, 26 (4): 338-348.

Jun S, Moon C, Kang C, Kong S, Snag B, and Um Y. (2010) Microbial fed-batch production of 1,3-propanediol production using raw glycerol with suspended and immobilized *Klebsiella pneumoniae*. *Appl Biochem Biotechnol,* 161: 491–501.

Jung Y-M, Lee Y-H. (2000) Utilization of oxidative pressure for enhanced production of poly-β-hydroxybutyrate and poly(3-hydroxybutyrate-3-

hydroxyvalerate) in *Ralstonia eutropha*. *Journal of Bioscience and Bioengineering*, 90(3): 266-270.

Kahar P, Agus J, Kikkawa Y, Taguchi K, Doi Y, and Tsuge T (2005). Effective production and kinetic characterization of ultra-high-molecular-weight poly[(R)-3-hydroxy-butyrate] in recombinant E*scherichia coli. Polym Degrad Stabil*, 87(1): 161–169.

Kahar P, Tsuge T, Taguchi K, Doi Y. (2004) High yield production of polyhydroxyalkanoates from soybean oil by *Ralstonia eutropha* and its recombinant strain. *Polymer Degradation and Stability*, 83(1): 79-86.

Kanzaki N, Kawamoto T, Matsui J, Nakahama K, Ifuku O. (2001) Microorganism resistant to threonine analogue and production of biotin. United States Patent 6,284,500.

Kessler B, Witholt B. (2001) Factors involved in the regulatory network of polyhydroxyalkanoate metabolism. *Journal of Biotechnology*, 86(2): 97-104.

Khalil AS, and Collins JJ. (2010) Synthetic biology: Applications come of age. *Nature Reviews Genetics*, 11(5): 367–79.

Kim BS, Lee SC, Lee SY, Chang HN, Chang YK, and Woo SI. (1994) Production of poly(3-hydroxybutyric acid) by fed-batch culture of *Alcaligenes eutrophus* with glucose-concentration control. *Biotechnology and Bioengineering*, 43(9): 892–8.

Kim HY, Park JS, Shin HD, and Lee YH. (1995) Isolation of glucose utilizing mutant of *Alcaligenes eutrophus*, its substrate selectivity, and accumulation of poly-beta-hydroxybutyrate. *Journal of Microbiology*, 33(1): 51–8.

Kim Y., Schumaker K, and Zhu J. (2005) EMS mutagenesis of *Arabidopsis*. In: Methods in Molecular Biology. 323. Humana Press, Inc., USA. p. 101 – 103.

Kiyasu T, Nagahashi Y, and Hoshino T. (2001) Cloning and characterization of biotin biosynthetic genes of *Kurthia* sp. *Gene*, 265, (1-2): 103-113.

Knowles JR. (1989) The mechanism of biotin-dependent enzymes. *Annu Rev Biochem,* 58: 195-221.

Koga N, Kishimoto J, Haze S, Ifuku O. (1996) Analysis of the *bioH* gene of *Escherichia coli* and its effect on biotin productivity. *J Ferment. Bioeng*, 81: 482–487.

Kovach ME, Phillips RW, Elzer PH. (1994) pBBR1MCS: a broad-host range cloning vector. *BioTechniques*; 16: 800–801.

Kovalchuk I, Kovalchuk O, and Hohn B. (2000) Genome-wide variation of the somatic mutation frequency in transgenic plants. *EMBO J.* 19: 4431–4438.

Koutinas AA, Wang R, Webb C. (2007) The biochemurgist—bio- conversion of agricultural raw materials for chemical production. *Biofuels Bioprod Bioref,* 1: 24–38.

Kram K, and Finkei S. (2014) Culture volume and vessel affect long-term survival, mutation frequency, and oxidative stress of *E. coli. Appl Environ Microbiol*, 80(15): 1732-1738.

Krieg D. (1963) Ethyl methanesulfonate-induced reversion of bacteriophage T4*r*II mutants. *Genetics* 48: 561–580.

Kumar A, Singh S. (2013) Directed evolution: tailoring biocatalysts for industrial applications. *Critical reviews in biotechnology*, 33(4): 365–78.

Kurosawa K, Radek A, Plassmeier J, and Sinskey A. (2015) Improved glycerol utilization by a triacylglycerol- producing *Rhodococcus opacus* strain for renewable fuels. *Biotechnology and Biofuels*, 8: 31.

Kusian B, Sultemeyer D, and Bowien B. (2002) Carbonic anhydrase is essential for growth of *Ralstonia eutropha* at ambient CO_2 concentrations. *Journal of Bacteriology*, 184(18): 5018–26.

Lane MD, Rominger KL, Young DL, and Lynen F. (1964) The enzymatic synthesis of holotranscarboxylase from apotranscarboxylase and (+)-biotin. *J Biol Chem,* 239: 2865-287.

Lee IY, Kim GJ, Shin YC, Chang HN, Park YH. (1995) Production of poly(beta-hydroxybutyrate-co-beta-hydroxyvalerate) by 2-stage fed-batch fermentation of *Alcaligenes eutrophus. Journal of Microbiology and Biotechnology*, 5(5): 292-296.

Lee SY. (1996) Plastic bacteria? Progress and prospects for polyhydroxyalkanoate production in bacteria. *Trends in Biotechnology*, 14(11): 431-438.

Lee TS, Krupa RA, Zhang F, Hajimorad M, Holtz WJ, Prasad N, Lee SK, Keasling JD. (2011) BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Eng.*, 20(5): 12.

Lemoine Y, Wach A, and Jeltsch JM. (1996) To be free or not: the fate of pimelate in *Bacillus sphaericus* and in *Escherichia coli. Mol. Microbiol*, 19: 645-647.

Lenz R, and Marchessault R. (2005) Bacterial Polyesters: Biosynthesis, Biodegradable Plastics and Biotechnology. *American Chemical Society*, 6(1): 1-8.

Li C, Lesnik KL, and Liu H. (2013) Microbial Conversion of Waste Glycerol from Biodiesel Production into Value-Added Products. *Energies*, 6: 4739-4768.

Li H, and Liao JC. (2015) A synthetic anhydrotetracycline-controllable gene expression system in *Ralstonia eutropha* H16. *ACS Synthetic Biology*, 4(2): 101–6.

Li H, Opgenorth PH, Wernick DG, Rogers S, Wu T-Y, Higashide W, *et al*. (2012) Integrated Electromicrobial Conversion of CO_2 to Higher Alcohols. *Science*, 335(6076): 1596.

Lin S, and Cronan J. (2011) Closing in on complete pathways of biotin biosynthesis. *Molecular BioSystems*, 7(6): 1811-1821.

Lin S, and Cronan J. (2012) The BioC O-methyltransferase catalyzes methyl esterification of malonyl-acyl carrier protein, an essential step in biotin synthesis. *J Biol Chem*, 287: 37010–37020.

Lin S, Hanson R, and Cronan J. (2010) Biotin synthesis begins by hijacking the fatty acid synthetic pathway. *Nature Chemical Biology*, 6 (9): 682-688.

Linde M, Galbe M, and Zacchi G. (2008) Bioethanol production from non-starch carbohydrate residues in process streams from a dry-mill ethanol plant. *Bioresource Technology*, 99 (14): 6505–11.

Lindlbauer AK, Marx H, and Sauer M. (2017) 3-Hydroxypropionaldehyde production from crude glycerol by *Lactobacillus diolivorans* with enhanced glycerol uptake. *Biotechnol Biofuels*, 10: 295.

Liu ZL, Slininger PJ, Gorsich SW. (2005) Enhanced biotransformation of furfural and hydroxymethylfurfural by newly developed ethanologenic yeast strains. *Appl Biochem Biotechnol*, 5: 451–460.

Loo CY, Lee WH, Tsuge T, Doi Y, Sudesh K. (2005) Biosynthesis and characterization of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from palm oil products in a *Wautersia eutropha* mutant. *Biotechnology Letters*, 27(18): 1405-1410.

Loo CY, Sudesh K. (2007) Polyhydroxyalkanoates: Bio-based microbial plastics and their properties. *Malaysian Polymer Journal*, 2(2): 31-57.

Loveless A. (1959) Mutation of bacteria at high levels of survival by ethyl methane sulphonate. *Nature*, 5: 1780-1782.

Lu J, Brigham CJ, Gai CS, Sinskey AJ. (2012) Studies on the production of branched-chain alcohols in engineered *Ralstonia eutropha. Applied Microbiology and Biotechnology*, 96(1): 283-297.

Lu P, Vogel C, Wang R, Yao X, Marcotte EM. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol*, 25: 117–124.

Lundby A, and Olsen JV. (2011) GeLCMS for in-depth protein characterization and advanced analysis of proteomes. *Methods Mol Biol*, 753: 143-155.

Lüte S, Pohlmann A, Zaychilkov E, Schwartz E, Becher JR, Heumann H, and Friedrich B. (2012) Autotrophic Production of Stable-Isotope-Labeled Arginine in *Ralstonia eutropha* Strain H16. *Appl Environ Microbiol*., 78(22): 7884-7890.

Madden LA, Anderson AJ, Asrar J, Berger P, and Garrett P. (2000) Production and characterization of poly(3-hydroxybutyrate-co-3-hydroxyvalerate-co-4-hydroxybutyrate) synthesized by *Ralstonia eutropha* in fed-batch cultures. *Polymer*, 41(10): 3499–3505.

Manandhar M, and Cronan J. (2017) Pimelic acid, the first precursor of the *Bacillus subtilis* biotin synthesis pathway, exists as the free acid and is assembled by fatty acid synthesis. *Molecular Microbiology*, 104(4): 595-607.

Mann S, Carillon S, Breyne O, and Marquet A. (2002) Total Synthesis of Amiclenomycin, an Inhibitor of Biotin Biosynthesis. *Chem. Eur.*, 8(2): 439-450.

Marangoni C, Furigo A, and Aragao G. (2001) The influence of substrate source on the growth of *Ralstonia eutropha*, aiming at the production of polyhydroxyalkanoate. *Braz. J. Chem. Eng*., 18(2): 1678.

Marchler-Bauer A, and Bryant SH. (2004) CD-Search: protein domain annotations on the fly. *Nucleic Acids Research*, 32.

Min YN, Yan F, Liu FZ, Coto C, and Waldroup PW. (2011) Glycerine—a new energy source for poultry. *Int J Poult Sci,* 9:1–4.

Moon C, Ahn J, Kim S, Sang B, and Um Y. (2010) Effect of biodiesel- derived raw glycerol on 1,3-propanediol production by different microorganisms. *Appl Biochem Biotechnol,* 161: 502–510.

Moser BR. (2011) Biodiesel Production, Properties, and Feedstocks. *In Biofuels*; Tomes, D., Lakshmanan, P., Songstad, D., Eds.; Springer: New York, NY, USA; pp. 285–347.

Mothes G, Schnorpfeil C, and Ackermann JU. (2007) Production of PHB from crude glycerol. *Eng. Life Sci*., 5: 475-479.

Mozumder SI, De Wever H, Volcke EIP, and García-González L. (2014) A robust fed-batch feeding strategy independent of the carbon source for optimal polyhydroxybutyrate production. Process Biochemistry 49(3): 365-373.

Mu Y, Teng H, Zhang DJ, Wang W, and Xiu ZL. (2006) Microbial production of 1,3-propanediol by *Klebsiella pneumoniae* using crude glycerol from biodiesel preparations. *Biotechnol Lett,* 28: 1755–1759.

Müller J, MacEachran D, Burd H, Sathitsuksanch N, Bi C, Yeh Y, Lee TS, Hillson NJ, Chhabra SR, Singer SW, and Beller HR. (2013) Engineering of *Ralstonia eutropha* H16 for Autotrophic and Heterotrophic Production of Methyl Ketones. *Appl Environ Microbiol*, 79(14): 4433-4439.

Nannemann D, Birmingham W, Scism R, and Bachmann B. (2001) Assessing directed evolution methods for the generation of biosynthetic enzymes with potential in drug biosynthesis. *Future Med Chem*, 3(7): 809-819.

Nielsen J. (1998) Metabolic engineering: techniques for analysis of targets for genetic manipulations. *Biotechnol Bioeng*, 58: 125–132.

Oh B, Seo J, Choi MH, and Kim CH. (2008) Optimization of culture conditions for 1,3-propanediol production from crude glycerol by *Klebsiella pneumoniae* using response surface methodology. *Biotechnol Bioprocess Eng,* 13: 666–670.

Oh B, Seo J, Heo S, Hong W, Luo L, Joe M, Park D, and Kim C. (2011) Efficient production of ethanol from crude glycerol by a *Klebsiella pneumoniae* mutant strain. *Bioresour Technol,* 102: 3918–3922.

Ohsugi M, Inoue Y. (1981) Formation of pimelic acid and azelaic acids, biotin intermediates, derived from Oleic acid by *Micrococcus* sp. *Agric. Biol. Chem*, 45(10): 2355-2356.

Okamoto H, Miyagawa A, Shiota T, Tamura Y, and Endo T. (2014) Intramolecular disulfide bond of Tim22 protein maintains integrity of the TIM22 complex in the mitochondrial inner membrane. *The Journal of biological chemistry. American Society for Biochemistry and Molecular Biology*, 289(8): 4827–38.

Oleoline. (2018) Glycerine Market Report June 2018 [online]. *The Independent Oleo reporter* [Viewed 15 March 2019]. Available from: http://www.hbint.com/datas/media/590204fd077a6e381ef1a252/sample-quarterly-glycerine.pdf

Otsuka A, and Abelson J. (1978) The regulatory region of the biotin operon in *Escheriehia coli*. *Nature*, 276 (5689): 689-694.

Overhage J, Steinbüchel A, and Priefert H. (2002) Biotransformation of Eugenol to Ferulic Acid by a Recombinant Strain of *Ralstonia eutropha* H16. *Appl Environ Microbiol*, 68(9): 4315-4321.

Park JM, Kim TY, Lee SY. (2011) Genome-scale reconstruction and *in silico* analysis of the *Ralstonia eutropha* H16 for polyhydroxyalkanoate synthesis,

lithoautotrophic growth, and 2-methyl citric acid production. *BMC Systems Biology*, 5(101).

Park HC, Lim KJ, Park JS, Lee YH, and Huh TL. (1995) High frequency tranformation of *Alcaligenes eutrophus* producing poly-β-hydroxybutyric acid by electroporation. *Biotechnol. Tech*., 9: 31- 34.

Paulo da Silva G, Mack M, and Contiero J. (2009) Glycerol: A promising and abundant carbon source for industrial microbiology. *Biotechnology Advances*, 27: 30-39.

Perkins JB, Bower S, Howitt CL, Yocum RR, and Pero J. (1996) Identification and characterization of transcripts from the biotin biosynthetic operon of *Bacillus subtilis*. *J. Bacteriol,* 178: 6361–6365.

Perkins DN, Pappin DJ, Creasy DM, and Cottrell JS. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS: An International Journal*, 20(18): 3551-3567.

Pohlmann A, Fricke WF, Reinecke F, Kusian B, Liesegang H, Cramm R Eitinger T, Ewering C, Pötter M, Schwartz E, Strittmatter A, Voß I, Gottschalk G, Steinbüchel A, Friedrich B, and Bowien B. (2006) Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16. *Nature Biotechnology*, 24(10): 1257–62.

Polyak SW, Abell AD, Wilce MC, Zhang, L, and Booker GW. (2012) Structure, function and selective inhibition of bacterial acetyl-CoA carboxylase. *Appl. Microbiol. Biotechnol*., 93(3): 983−992.

Porcar M. (2010) Beyond directed evolution: Darwinian selection as a tool for synthetic biology. *Sis Synth Biol*, 4: 1-6.

Posada JA, Cardona CA. (2010) Design and analysis of fuel ethanol production from raw glycerol. *Energy*, 35: 5286–5293.

Pries A, Steinbüchel A, and Schlegel HG. (1990) Lactose-utilizing and galactose-utilizing strains of poly(hydroxyalkanoic acid)-accumulating *Alcaligenes eutrophus* and *Pseudomonas saccharophila* obtained by recombinant-DNA technology. *Applied Microbiology and Biotechnology*, 33(4): 410–7.

Raberg M, Peplinski K, Heiss S, Ehrenreich A, Voigt B, Doering C, Bömeke M, Hecker M, and Steinbüchel A. (2011) Proteomic and transcriptomic elucidation of the mutant Ralstonia eutropha G(+)1 with regard to glucose utilization. *Applied and Environmental Microbiology*, 77(6): 2058–70.

Rausch KD, Belyea RL. (2006) The future of coproducts from corn processing. *Appl Biochem Biotechnol*, 128: 47–86.

Rehm BHA. (2010) Bacterial polymers: biosynthesis, modifications and applications. *Nature Reviews Microbiology*, 8(8): 578-92.

Reinecke F and Steinbüchel A. (2009) *Ralstonia eutropha* strain H16 as model organism for PHA metabolism and for biotechnological production of technically interesting biopolymers. *Journal of Molecular Microbiology and Biotechnology*, 16(1–2): 91–108.

Riedel SL, Bader J, Brigham CJ, Budde CF, Yusof ZAM, Rha C, and Sinskey A. (2012) Production of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) by *Ralstonia eutropha* in high cell density palm oil fermentations. *Biotechnology and Bioengineering*; 109(1): 74-83.

Rodionov D, Dubchak I, Arkin A, Alm E, and Gelfand M. (2004) Reconstruction of regulatory and metabolic pathways in metal-reducing δ-proteobacteria.*Genome Biology*, 5: R90.

Rodionov D, Mironov A, and Gelfand M. (2002) Conservation of the Biotin Regulon and the BirA Regulatory Signal in Eubacteria and Archaea. *Genome Research*, 12: 1507-1516.

Rymowicz W, Rywijska A, and Goldkowski W. (2008) Simultaneous production of citric acid and erythritol from crude glycerol by *Yarrowia lipolytica*. *Chem Pap* 62: 239–246.

Sakurai N, Imai Y, Masuda M, Komatsubara S, and Tosa T. (1993) Construction of a Biotin-Overproducing Strain of *Serratia marcescens. Appl Environ Microbiol*, 59: 2857-2863.

Sakurai N, Imai Y, and Komatsubara S. (1995) Instability of the mutated biotin operon plasmid in a biotin-producing mutant of *Serratia marcescens*. *Journal of Biotechnology*, 43(1): 11-19.

Sakurai N, Imai Y, Masuda M, Komatsubara S, Tosa T. (1994) Improvement of a d-biotin-hyperproducing recombinant strain of *Serratia marcescens. J Biotechnol*, 36: 63–73.

Sambrook J, and Russel D. (2011) Molecular Cloning: A Laboratory Manual, third ed. Cold Spring Harbor Laboratory Press, Cold Spring, NY.

Sanyal I, Lee S, and Flint DH. (1994) Biosynthesis of pimeloyl-CoA, a biotin precursor in *Escherichia coli*, follows a modified fatty acid synthesis pathway. 13C-Labeling studies. *J. Am. Chem. Soc,* 116: 2637–2638.

Sato S, Fujiki T, and Matsumoto K. (2013) Construction of a stable plasmid vector for industrial production of poly(3-hydroxybutyrate-*co*-3-hydroxyhexanoate) by a recombinant *Cupriavidus necator* H16 strain. *J. Biosc. Bioeng*, 116(6): 677-681.

Schlegel HG, Gottschalk G, and Von Bartha R. (1961a) Formation and utilization of poly-[beta]-hydroxybutyric acid by Knallgas bacteria (Hydrogenomonas). *Nature, 191*(4787): 463–465.

Schlegel HG, and Gottscha G. (1965) Verwertung von glucose durch eine mutante von hydrogenomonas H16. *Biochemische Zeitschrift*, 341(3): 249–259.

Schlegel HG, Kaltwasser H, and Gottschalk G. (1961b) Ein Submersverfahren zur Kultur wasserstoffoxydierender Bakterien: Wachstumsphysiologische Untersuchungen. *Archiv für Mikrobiologie,* 38(3): 209–222.

Schlegel HG, Lafferty R, Krauss I. (1970) Isolation of mutants not accumulating poly-beta-hydroxybutyric acid. *Archiv Fur Mikrobiologie*; 71(3): 283-294.

Schrag DP. (2007) Preparing to capture carbon. *Science*, 315(5813): 812–3.

Schwartz E, Henne A, Cramm R, Eitinger T, Friedrich B, and Gottschalk G. (2003) Complete nucleotide sequence of pHG1: A *Ralstonia eutropha* H16 megaplasmid encoding key enzymes of H2-based lithoautotrophy and anaerobiosis. *Journal of Molecular Biology*, 332(2): 369–383.

Sega G. (1984) A review of the genetic effects of ethyl methanesulfonate. *Mutation Research*, 134: 113-142.

Sen M, Yılmaz U, Baysal A, Akman S, Cakar ZP. (2011) In vivo evolutionary engineering of a boron-resistant bacterium: *Bacillus boroniphilus. Antonie Van Leeuwenhoek*, 99: 825–835.

Shahnawaz K, and Samiullah K. (2014) Ethyl methanesulphonate (EMS), A Potent Chemical Mutagen: A Review. *International Journal of Scientific Research*, 3(10): 488-490.

Shang LG, Jiang M, and Chang HN. (2003) Poly(3-hydroxybutyrate)synthesis in fed-batch culture of *Ralstonia eutropha* with phosphate limitation under different glucose concentrations. *Biotechnology Letters*, 25(17): 1415–1419.

Shi C, and Alrich C. (2012) Design and Synthesis of Potential Mechanism-Based Inhibitors of the Aminotransferase BioA Involved in Biotin Biosynthesis. *J. Org. Chem*., 77: 6051-6058.

Shimizu R, Chou K, Orita I, Suzuki Y, Nakamura S, and Fukui T. (2013) Detection of phase-dependent transcriptomic changes and Rubisco-mediated CO_2 fixation into poly (3-hydroxybutyrate) under heterotrophic condition in *Ralstonia eutropha* H16 based on RNA-seq and gene deletion analyses. *BMC Microbiology,* 13: 169.

Sichwart S, Hetzler S, Broeker D, and Steinbüchel A. (2011) Extension of the substrate utilization range of *Ralstonia eutropha* strain H16 by metabolic engineering to include mannose and glucose. *Applied and Environmental Microbiology*, 77(4): 1325–34.

Sydow A, Pannek A, Krieg T, Huth I, Guillouet SE, and Holtmann D. (2017) Expanding the genetic tool box for *Cupriavidus necator* by a stabilized L-rhamnose inducible plasmid system. *J Biotechnol*, 263: 1-10.

Sniegowski P, Gerrish P, Lenski R. (1997) Evolution of high mutation rates in experimental populations of *E. coli. Nature*, 387: 703–705.

Solaiman D, Swingle BM, and Ashby R. (2010) A new shuttle vector for gene expression in biopolymer-producing *Ralstonia eutropha. J Microbiol M*, 82: 120-123.

Sonderegger M, and Sauer U. (2003) Evolutionary Engineering of *Saccharomyces cerevisiae* for Anaerobic Growth on Xylose. Appl. Environ. Micriobiol, 69(4): 1990-1998.

Soucaille P. (2002) Metabolic pathway engineering for the production of 1,3-propanediol from glucose. *Abstracts of Papers American Chemical Society*, 224(1– 2): 129-BIOT.

Srinivasan S, Barnard GC, Gerngross TU. (2003) Production of recombinant proteins using multiple-copy gene integration in high-cell-density fermentations of *Ralstonia eutropha. Biotechnol. Bioeng.*, 84: 114–120.

Steinbüchel A, Lutke-Eversloh T. (2003) Metabolic engineering and pathway construction for biotechnological production of relevant polyhydroxyalkanoates in microorganisms. *Biochemical Engineering Journal*, 16(2): 81-96.

Stok JE, and De Voss JJ. (2000) Expression, purification, and characterization of BioI: a carbon–carbon bond cleaving cytochrome P450 involved in biotin biosynthesis in *Bacillus subtilis*. *Arch Biochem Biophys,* 384: 351-360.

Streit WR, and Entcheva P. (2003) Biotin in microbes, the genes involved in its biosynthesis, its biochemical role and perspectives for biotechnological production. *Appl Microbiol Biotechnol*, 61: 21-31.

Sullivan JT, Brown SD, Yocum RR, and Ronson CW. (2001) The *bio* operon on the acquired symbiosis island of *Mesorhizobium* sp. strain R7A includes a novel gene involved in pimeloyl-CoA synthesis. *Microbiology,* 147: 1315–1322.

Taga N, Tanaka K, and Ishizaki A. (1997) Effects of rheological change by addition of carboxymethylcellulose in culture media of an air-lift fermentor on poly-D-3-hydroxybutyric acid productivity in autotrophic culture of hydrogen- oxidizing Bacterium, *Alcaligenes eutrophus*. *Biotechnology and Bioengineering*, 53(5): 529–33.

Taketo, A. (1974) Sensitivity of *Escherichia coli* to viral nucleic acid. *J. Biochem*. 75: 895–904.

Tanadchangsaeng N, and Yu J. (2012) Microbial Synthesis of Polyhydroxybutyrate From Glycerol: Gluconeogenesis, Molecular Weight and Material Properties of Biopolyester, *Biotechnology and Bioengineering*, 109(11): 2808-18.

Tcherkez GGB, Farquhar GD, and Andrews TJ. (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences of the United States of America*, 103(19): 7246–51.

Tee KL, and Wong TS. (2013) Polishing the craft of genetic diversity creation in directed evolution. *Biotechnology Advances*, 31: 1707-1721.

Tee KL, and Wong TS. (2014) Directed evolution: A powerful algorithm for advancing synthetic biology. In: Singh, V., *Ed. Advanced Synthetic Biology*. 4. USA: Studium Press LLC, 465–97.

Tizei P, Csibra E, Torres L, and Pinheiro V. (2016) Selection platforms for directed evolution in synthetic biology. *Biochem Soc Trans*, 44(4): 1165–1175.

Tobin MB, Gustafsson C, Huisman GW. (2000) Directed evolution: the 'rational' basis for 'irrational' design. *Curr Opin Struct Biol*, 10: 421–427.

Todd P, Brouwer J, and Glickman B. (1981) Influence of DNA-repair deficiencies on MMS and EMS induced mutagenesis in *Escherichia coli* K-12. *Mutation Research*, 82: 239-250.

Truniger V, and Boss W. (1993) Glycerol uptake in *Escherichia coli* is sensitive to membrane lipid composition. *Res. Micriobiol*, 144: 565-574.

Tyo KE, Zhou H, and Stephanopoulos GN. (2006) High-throughput screen for poly-3-hydroxybutyrate in *Escherichia coli* and *Synechocystis* sp. strain PCC6803. *Appl Environ Microbiol*, 72(5): 3412-3417.

U.S. Energy Information Administration: Washington, DC, USA. (2013) *Monthly Biodiesel Production Report*.

Van Arsdell SW, Perkins JB, Yocum RR, Luan L, Howitt CL, Chatterjee NP, and Pero JG. (2005) Removing a bottleneck in the *Bacillus subtilis* biotin pathway: BioA utilizes lysine rather than *S*- adenosylmethionine as the amino donor in the KAPA-to-DAPA reaction. *Biotechnology and Bioengineering*, 91(1): 75-83.

Volodina E, Raberg M, and Steinbüchel A. (2016) Engineering the heterotrophic carbon sources utilization range of *Ralstonia eutropha* H16 for applications in biotechnology. *Critical Reviews in Biotechnology*, 36(6): 978-991.

Voss I, and Steinbüchel A. (2006) Application of a KDPG-aldolase gene-dependent addiction system for enhanced production of cyanophycin in *Ralstonia eutropha* strain H16. *Metabolic Engineering*, 8: 66-78.

Washburn MP, Wolters D, and Yates III JR. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature biotechnology*, 19(3): 242.

Wensink P, Finnegan D, Donelson J, and Hogness D. (1974) A system for mapping DNA sequences in the chromosomes of *Drosophila melanogaster*. *Cell* 3: 315–325.

Wilde E. (1962) Untersuchungen über Wachstum und Speicherstoff synthese von Hydrogenomonas. *Archiv für Mikrobiologie,* 43(2): 109–137.

Wilms B, Hauck A, Reuss M, Syldatk C, Mattes R, Siemann M, Altenbuchner J. (2001) High cell density fermentation for production of L-*N*-carbamoylase using an expression system based on the *Escherichia coli* rhaBAD promoter. *Biotechnol Bioeng*, 73: 95–103.

Witthoff S, Schmitz K, Niedenführ S, Nöh K, Noack S, Bott M, Marienhagen J. (2015). Metabolic engineering of *Corynebacterium glutamicum* for methanol metabolism. *Appl Environ Microbiol*, 81: 2215–2225.

Wu CF, Valdes JJ, Rao G, Bentley WE. (2001). Enhancement of organophosphorus hydrolase yield in *Escherichia coli* using multiple gene fusions. *Biotechnol Bioeng*, 75: 100–103.

Wu G, Yan Q, Jones A, Tang Y, Fong SS, and Koffas, MAG. (2016) Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends in Biotechnology,* 34(8): 652-664.

Yadav D, Tanveer A, Malviya N, and Yadav S. (2018) Overview and Principles of Bioengineering: The Drivers of Omics Technologies. In: Omics Technologies and Bio-Engineering. 1. Academic Press, p. 3-23.

Yamada H, Osakai M, and Tani Y. (1983) Biotin Overproduction by Biotin Analog-resistant Mutants of *Bacillus sphaericus. Agric. Biol. Chem.*, 47(5): 1011-1016.

Yamane T, Fukunaga M, and Lee YW (1996). Increased P(3HB) productivity by high-cell-density-fed-batch culture of *Alcaligenes latus*, a growth-associated P(3HB) producer. *Biotechnol Bioeng*, 50: 197–202.

Yang F, Hanna MA, and Sun R. (2012) Value-added uses for crude glycerol— a byproduct of biodiesel production. *Biotechnol Biofuels,* 5: 13.

Yeh JI, Charrier V, Paulo J, Hou L, Darbon E, Claiborne A, Deutscher J. (2004) Structures of *enterococcal* glycerol kinase in the absence and presence of

glycerol: correlation of conformation to substrate binding and a mechanism of activation by phosphorylation. *Biochemistry*, 43(2): 362-373

Yu J. (2014) Bio-based products from solar energy and carbon dioxide. *Trends in Biotechnology*, 32(1): 5-10.

Zambrano MM, Siegele DA, Almirón M, Tormo A, and Kolter R. (1993) Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. *Science*, 259: 1757-1760.

Zha W, Rubin-Pitel S, Shao Z, and Huimin Z. (2009) Improving cellular malonyl-CoA level in *Escherichia coli* via metabolic engineering. *Metabolic Engineering*, 11(3): 192-198.

Zhang WW, Yang MM, Li Hx, and Wang D. (2011) Construction of recombinant *Bacillus subtilis* strains for efficient pimelic acid synthesis. *Electronic Journal of Biotechnology.* ISSN: 0717-3458.

Zhao H. (2007) Directed evolution of novel protein functions. *Biotechnol Bioeng*, 98: 313–317.

Zhu C, Nomura CT, Perrotta JA, Stipanovic AJ, and Nakas JP. (2010) Production and Characterization of Poly-3-hydroxybutyrate From Biodiesel-Glycerol by *Burkholderia cepacia* ATCC 17759. *Biotechnol. Prog*., 26(2): 424-430.

APPENDICES

Appendix 1: List of reagents

List of Reagents	Catalogue #; Manufacturer details
Antibiotics	
Chloramphenicol	Cat#: CLA01; FORMEDIUM, Norfolk, UK
Kanamycin	Cat#: A1493,0025; AppliChem, Cheshire, UK
Gentamicin	Cat#: A1492,0005; AppliChem, Cheshire, UK
Media components	
Agar granulated	Cat#: AGR03; FORMEDIUM, Norfolk, UK
Ammonium chloride (NH_4Cl)	Cat#: A9434-500; Sigma-Aldrich, Dorset, UK
Ammonium iron(III) citrate ($Fe(III)NH_4$-Citrate)	Cat#: 09713-50; Fluka, Bremen, Germany
Beef extract	Cat#: B4888-50G; Sigma-Aldrich, Dorset, UK
Calcium chloride ($CaCl_2 \bullet 2H_2O$)	Cat#: 437055N; VWR, Leicestershire, UK
Coomassie blue	Cat#: A1092.0025; AppliChem, Cheshire, UK
Crude glycerol (sweet water)	Obtained from a high-pressure fat splitting process as industrial waste kindly provided by Croda Europe Ltd (batch 8/1/18), Hull, UK
Glycerol, BioReagent	Cat#: G2025-500; Sigma-Aldrich, Dorset, UK
Magnesium sulphate ($MgSO_4 \bullet 7H_2O$)	Cat#: 437044k; VWR, Leicestershire, UK
Peptone	Cat#: PEP03; FORMEDIUM, Norfolk, UK
Potassium phosphate monobasic (KH_2PO_4)	Cat#: P979; Sigma-Aldrich, Dorset, UK
Sodium chloride (NaCl)	Cat#: NAC02; FORMEDIUM, Norfolk, UK
Sodium gluconate (NaGlu)	Cat#: A3712.0250; AppliChem, Cheshire, UK
Sodium phosphate dibasic ($Na_2HPO_4 \bullet 7H_2O$)	Cat#: 71505; Sigma-Aldrich, Dorset, UK
Trace elements solution SL6	
$ZnSO_4 \bullet 7H_2O$	Cat#: Z/1550/53; Fisher Chemical, Loughborough, UK
$MnCl_2 \bullet 4H_2O$	Cat#: 36526; Alfa aesar, Lancashire, UK
H_3BO_3	Cat#: 327130010; Acros, Leicestershire, UK
$CoCl_2 \bullet 6H_2O$	Cat#: C8661; Sigma-Aldrich, Dorset, UK
$CuCl_2 \bullet 2H_2O$	Cat#: C3279-100; Sigma-Aldrich, Dorset, UK
$NiCl_2 \bullet 6H_2O$	Cat#: N6136-100; Sigma-Aldrich, Dorset, UK
$Na_2MoO_2 \bullet 2H_2O$	Cat#: 331058-5; Sigma-Aldrich, Dorset, UK
Tryptone	Cat#: TRP02; FORMEDIUM, Norfolk, UK
Yeast extract	Cat#: YEA03, FORMEDIUM, Norfolk, UK
Miscellaneous	
Acetic acid	Cat#: 20104.334; VWR, Leicestershire, UK
Acrylamide Bis-Acrylaminde	Cat#: 20-2100-10; Severn Biotech Ltd., Worcestershire, UK
Agarose	Cat#: 50004; Lonza, Slough, UK
Ammonium persulphate (APS)	Cat#: A2941.0500; AppliChem, Cheshire, UK
Boric acid	Cat#: 327130010; VWR, Leicestershire, UK
Bromophenol blue	Cat#: 32641; Alfa Aesar, Lancashire, UK
Coomassie brilliant blue staining dye	Cat#: A1092.0025; AppliChem, Cheshire, UK
DNA ladder 1 kb Quick-Load ® [50 µG/ML]	Cat#: N0468S; New England BioLabs, Herts, UK
EDTA DISODIUM	Cat#: EDTA500; FORMEDIUM, Norfolk, UK
Ethanol, BioUltra	Cat#: 51976-500ML-F; Sigma-Aldrich, Dorset, UK
Ethidium bromide	Cat#: HC001596; MERCK, Darmstadt, Germany
Gel Loading Dye Blue (6x)	Cat#: B707 S; New England Biolabs, Herts, UK
Glycerol	Cat#: 444485B; VWR, Leicestershire, UK

L-(+)-Arabinose	Cat#: SC-221794B; ChemCruz, Heidelberg, Germany
Lysozyme	Cat #: 89833; Thermo Scientific, Loughborough, UK
Methanol	Cat#: 10598240; Fisher Chemical, Loughborough, UK
Nile Red	Cat#: N3013-100; Sigma-Aldrich, Dorset, UK
Potassium phosphate dibasic (K_2HPO_4)	Cat#: A2945,1000; AppliChem, Chesire, UK
PageRuler Broad Range Unstained Protein Ladder	Cat#: 26630; Thermo Fisher Scientific, Loughborough, UK
Sodium dodecyl sulphate (SDS)	Cat#: 444464T; VWR, Leicestershire, UK
Sodium hydroxide (NaOH)	Cat#: A6829-0500; AppliChem, Cheshire, UK
Sodium thiosulphate ($Na_2S_2O_3$)	Cat#: A17629; Alfa Aesar, Lancashire, UK
Sucrose	Cat#: 84100-250; Sigma-Aldrich, Dorset, UK
Synthetic genes (DV and DD)	GenScript; Piscataway, USA
Tetramethylethylenediamine (TEMED)	Cat#: A1148,0100; AppliChem, Cheshire, UK
Tris Molecular Biology grade	Cat#: BP152-1; Fisher Chemical, Loughborough, UK
T4 DNA ligase [400,000 U/ml]	Cat#: M0202S; New England BioLabs, Herts, UK
T4 DNA ligase 10x Buffer with 10 mM ATP	Cat#: B0202S; New England BioLabs, Herts, UK
5-(2-thienyl)-pentanoic acid (TVA)	Cat#: L13122; Alfa Aesar, Lancashire, UK
β-mercaptoethanol	Cat#: A1108,0100; Applichem, Damstadt, Germany
Restriction enzymes	
AvrII [5,000 U/ ml]	Cat#: R0174S; New England BioLabs, Herts, UK
DpnI [20,000 U/ml]	Cat#: R0176S; New England BioLabs, Herts, UK
NdeI [20,000 U/ml]	Cat#: R0111S; New England BioLabs, Herts, UK
PstI-HF [20,000 U/mL]	Cat#: R3140S; New England BioLabs, Herts, UK
SacI-HF [20,000 U/mL]	Cat#: R3156S; New England BioLabs, Herts, UK
XhoI [20,000 U/ml]	Cat#: R0146S; New England BioLabs, Herts, UK
10x CutSmart Buffer	Cat#: B7204S; New England BioLabs, Herts, UK
PCR reagents	
dNTPs (10 mM)	Cat#: N0447S; New England BioLabs, Herts, UK
Primers	Eurofins Genomics; Ebersberg, Germany
Pfu Turbo DNA Polymerase [2.5 U/µl]	Cat#: 600250-52; Agilent, Cheshire, UK
10x Cloned Pfu Reaction Buffer	Cat#: 600153-82; Agilent, Cheshire, UK
Pfu Ultra HF DNA Polymerase [2.5 U/µl]	Cat#: 600380-51; Agilent, Cheshire, UK
10x Pfu Ultra HF Reaction Buffer	Cat#: 600380-52; Agilent, Cheshire, UK
Q5 HF DNA Polymerase [2.0 U//µl]	Cat#: M0491S; New England BioLabs, Herts, UK
5x Q5 Reaction Buffer	Cat#: B9027S; New England BioLabs, Herts, UK
5x Q5 High GC Enhancer	Cat#: M0491S; New England BioLabs, Herts, UK

Appendix 2: List of kits

Experiment	Catalogue #; Manufacturer details
Biotin quantitation	Cat#: F30751; FluoReporter Biotin Quantitation Assay Kit, Molecular Probes, Paisley, UK
HiFi DNA Assembly Cloning	Cat#: E5520S; NEBuilder® HiFi DNA Assembly Cloning Kit, New England Biolabs Ltd, Hitchin, UK
Isolation of plasmid DNA	Cat#: D6942-01; E.Z.N.A.® Plasmid Mini Kit I, Omega Bio-Tek, Norcross, USA
Gel and PCR purification	Cat#: 740609.250; NucleoSpin® Gel and PCR Clean-up Macherey-Nagel, Düren, Germany
Genomic DNA extraction	Cat#: D3350-01; E.Z.N.A.® Bacterial DNA Kit, Omega Bio-Tek, Norcross, USA

Appendix 3: List of equipment

Experiment	Details
Bacterial cultivation	For 37 ºC → ES-20 shaker-incubator; Grant Instruments, Shepreth, UK For 30 ºC → MaxQTM 4450 benchtop orbital shaker; Thermo Fisher Scientific, Loughborough, UK
Bacterial incubation	For 30º C and 37 ºC → INCU-Line Incubator; VWR, Leicestershire, UK
Bacterial Optical Density (OD_{600}) – Large cuvettes	BioPhotometer Plus UV/Vis photometer; Eppendorf, Stevenage, UK
Bacterial Optical Density (OD_{595}, OD_{600}) – Microtitre plates	**Chapter 4** → Multiskan^{TM} FC microplate photometer; Thermo Fisher Scientific, Loughborough, UK **Chapter 5** → SpectraMax M2e microplate reader, Molecular Devices, Winnersh, UK
Microscopy	RX30F Brunel Microscope Ltd, Chippenham, UK
Bacterial transformation	For heat shock → Eppendorf Thermomixer C; Eppendorf, Stevenage, UK
Centrifugation	Microcentrifuge accuSpin Micro 17; Fisher Scientific, Loughborough, UK
DNA concentration	VersaWave Spectrophotometer; Expedeon, Cambridge, UK
Electrophoresis	Mini-Sub Cell GT Systems; Bio-Rad, Hertfordshire, UK
Electroporation	Electroporator →Eppendorf Eporator; Eppendorf, Stevenage, UK Electroporation cuvettes → Eppendorf 1 mm and 2 mm; Eppendorf, Stevenage, UK
Fluorescence	**Chapter 4** → Fluoroskan Ascent, Thermo Scientific; Loughborough, UK **Chapter 5** → SpectraMax M2e microplate reader; Molecular Devices, Wokingham, UK
Gel documentation system	Genosmart2, VWR, Leicestershire, UK
PCR	Eppendorf Mastercycler; Eppendorf, Stevenage, UK
pH	pH meter Jenway 3510; Jenway, Staffs, UK

Appendix 4: List of software

Experiment	Details
Analysis of data and creation of graphs	GraphPad Prism 7 Software, La Jolla, CA, USA
DNA and protein data analysis	SnapGene, GSL Biotech. Chicago, USA
Protein graphics	PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC
Protein modelling	SWISS-MODEL (Biasini *et al*., 2014)

Appendix 5: Media preparation

Media	Preparation
2x YT	Per 1 L 16 g Tryptone 10 g Yeast extract 5 g NaCl Sterilise by autoclaving
TYE Agar plate	Per 1 L 10 g Tryptone 5 g Yeast extract 8 g NaCl 15 g Agar Sterilise by autoclaving
NB	Per 1 L 5 g Peptone 1 g Beef extract 2 g Yeast extract 5 g NaCl 15 g agar (if required) Sterilise by autoclaving
M9 salts	Prepare M9 salts 5x stock per 1 L 30.0 g $Na_2HPO_4 \bullet 7H_2O$ 15.0 g KH_2PO_4 2.5 g NaCl 5.0 g NH_4Cl Sterilise by autoclaving Prepare stocks of the following components: 0.01 M $CaCl_2 \bullet 2H_2O$ 1.00 M $MgSO_4 \bullet 7H_2O$ 20 % (w/v) Glucose 1 mg/mL Thiamine (*sterilize with 0.2 μm filter*) 40 mg/mL Casamino acids 4 x 10^{-6} M Biotin (*sterilize with 0.2 μm filter*) Sterilise by autoclaving the previous components separately unless specified otherwise Prepare M9 supplemented with: 1x M9 salts 1 x 10^{-4} M $CaCl_2 \bullet 2H_2O$

	1×10^{-3} M $MgSO_4 \bullet 7H_2O$ 0.2 % (w/v) Glucose 1×10^{-3} mg/mL Thiamine 0.04 mg/mL Casamino acids 2×10^{-9} M Biotin (Biotin-free media was prepared as well)
MSM	Per 1 L 6.74 g $Na_2HPO_4 \bullet 7H_2O$ 1.50 g KH_2PO_4 1.00 g NH_4Cl (*Use 0.50 g for N_L condition*) 0.20 g $MgSO_4 \bullet 7H_2O$ (*Take 400 µL from a 0.5 g/mL stock*) 20.00 mg $CaCl_2 \bullet 2H_2O$ (*Take 80 µL from a 0.25 g/mL stock*) 1.20 mg Fe(III)NH_4-Citrate (*Take 4.8 µL from a 0.25 g/mL stock*) 10 g Sodium gluconate (*A different carbon source can be used in MSM medium*) 0.10 mL SL6 (*Trace elements solution SL6 must be prepared separately and sterilised by 0.2 µm filter. Add to MSM after it has been autoclaved*) Sterilise by autoclaving Prepare SL6 stock per 1 L: 10 mg $ZnSO_4 \bullet 7H_2O$ (*Take 40 µL from a 0.25 g/mL stock*) 3 mg $MnCl_2 \bullet 4H_2O$ (*Take 12 µL from a 0.25 g/mL stock*) 30 mg H_3BO_3 (*Measure 30 mg. Not very soluble in water*) 20 mg $CoCl_2 \bullet 6H_2O$ (*Take 80 µL from a 0.25 g/mL stock*) 1 mg $CuCl_2 \bullet 2H_2O$ (*Take 4 µL from a 0.25 g/mL stock*) 2 mg $NiCl_2 \bullet 6H_2O$ (*Take 8 µL from a 0.25 g/mL stock*) 3 mg $Na_2MoO_2 \bullet 2H_2O$ (*Take 12 µL from a 0.25 g/mL stock*) Sterilised by 0.2 µm filter

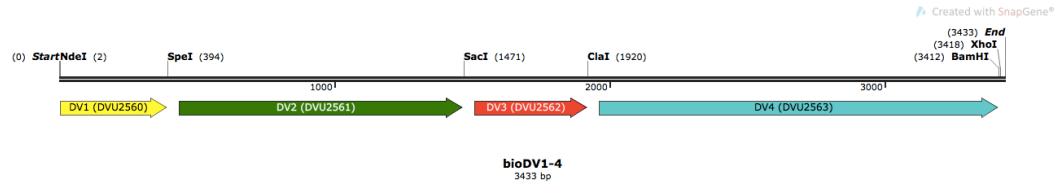
Appendix 6: List of miscellaneous

List of miscellaneous	**Catalogue #; Manufacturer details**
Filter Unit 0.2 µm	Cat #: 10462200; GE Healthcare Life Sciences Whatman™, Buckinghamshire, UK
Filter paper Circles 595	Cat #: 10311612; Whatman GmbH, Dassel, Germany
Gene sequencing	Eurofins Genomics sequencing was used to check quality of plasmids, and for DNA sequencing of *glpK* of *C. necator* H16 WT and *glpK* of *C. necator* H16 variants; Eurofins Genomics UK, Wolverhampton, UK.
Microscope slides 0.8 -1.0 mm	Cat #: B57011/2; Thermo Scientific, Loughborough, UK
Microscope cover glasses	Cat #: MNJ-350-020H; Thermo Scientific, Loughborough, UK

Appendix 7: Synthetic *bioDV1-4* and *bioDD1-5* extended biotin operons

Optimised codon sequence of the designed *bioDV1-4* and *bioDD1-5* extended biotin operons synthesised by GenScript. Genes are marked with colours according to diagrams, and restriction sites are in italics and underlined.

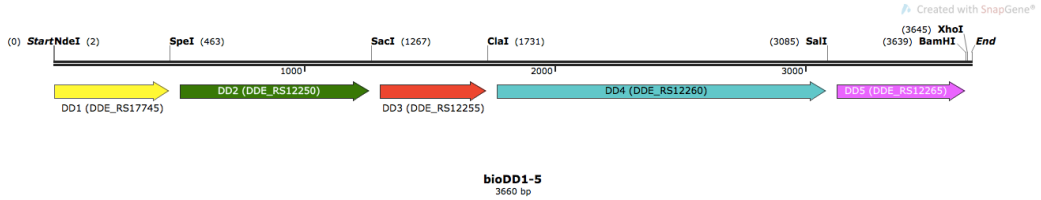
Synthetic *bioDV1-4* extended biotin operon:



*CATATG*AAGACCGTGCACGAGGGTACCTTCCTGTTTCCGGCGACCCACCCGCTGTTTGC
GGACCACTTTGAAGGTGCGCCGGTGGTTCCGGGCAGCCTGCTGCTGCGTGCGTTTGTG
GAGGAAGCGGGTACCGTTTGGCCGGATTTTGCGACCCAGGGTGCGACCGGCTTCCGTT
TTCGTCGTTTCGTTACCCCGGGCCGTCACCCGTTTCGTATGGAGCGTACCGACCTGAC
CGTGCGTTGCACCCTGATGGATGCGCAAGGTGGCATCCTGGTTCGTGGTGACCTGACC
ACCGCGGACGCGACCGGTGATGCGGGTACCGACAACGATCGTACCGGTGCGGATGGT
GAAGCGGCGCTGCACACCAGCCCGCGTGGCACCACCGCGTAATGA*ACTAGT*GAATTCA
AAAGATCTTTTAAGAAGGAGATATACAAATGACCCTGCACCAGAAGGATGGTGACGA
TTACGCGGGTGGCACCCGTCGTCCGACCGCGGTGGCGCACGCGCAACATGACATTGC
GCCGGGCGTGGTTCTGATTCTGGGTGGCAGCAGCACCCTGGGTCTGGCGGCGGGTCGT
GCGCTGCTGGCGGCGGGTCACCCGGTTGCGCTGACCACCCGTAACGCGGAGGGTACC
GGTCGTATCCTGGCGGCGCTGGCGCCGGGTGAAGCGGGCATTCCGGACGGTGGCACC
GATTTCAGCGGCAACCGTCCGGACCTGTTTAGCAACCGTCCGGATCTGAGCAACGGTG
GCGCGGACCGTTGCCGTGTGTGCCGTCACACCGGTCACGCGCCGTATGCGGTGCTGCC
GGTTGAGAGCGCGGAAAGCCTGCCGGACCGTTGCGCGGATGCGCTGGGTACCCCGCC
GCGTCACATGCTGGATTTCATGCATGGCCGTTTTGAAACCCTGGTGGCGGCGGCGCAT
GTTCATGACATCGACCGTTGGGCGATCGACGATATTGCGCTGCGTGCGCGTTGCCTGC
GTGCGGTTAGCCGTGCGATGCTGGCGGAACGTGCGGGTCGTTGCGTGTTCGTTAGCAG
CACCGCGGCGGCGATGGCGGCGCACGGTCAAGGTTACTATGCTGCGGCGAAAGCGGC
GGGTGAAGCGCTGTACCGTAGCGCGGGTCTGGAACTGGCGAGCCGTGGCGTGACCAC
CTGTAGCCTGCGTCTGGGTTGGGTTGATGCGGGTCGTGGCGCGGATTTTATCCGTGGT
AACGATGCGCCGTATCGTGATGGTACCGCGCACGCGGTTCCGACCCCGCACACCGCG
ACCCAACTGGCGGATGGTGTTCCGACCGGCCGTCTGGTGACCGTTGACGAGGTGGTTG
GTACCCTGATGTTCCTGCTGAGCGGTGCGGCGACCAGCCTGAACGCGACCCACATTAC
CCTGGACGGTGGCTTTACCGCGGGTAAACCGCGTGCGGGTACCCCGCGTACCGGTAG
CGGCACCCGTCGTACCGGTTAATGA*GAGCTC*GAATTCAAAAGATCTTTTAAGAAGGAG
ATATACAAATGCCGGAAACCCCGGCGTTCGATACCGTGCGTGCGATCGTTGCGGACA
TTTTTGAAGTTGCGCCGGAGACCCTGACCCCGGCGACCCGTCTGTTCGAAGATCTGCC
GTGCGAGAGCATCGACCTGCTGGAGATTGGTGCGGGCCTGAACCGTCGTTGCCGTATC
CCGGTTGACGATGACGCGGCGTTTCTGCGTAGCCTGCGTATTCACCTGATGGAGGCGG
AACGTGAGCGTCGTGATGCGACCCAACACCTGGCGAGCGTGTACCCGCACCTGACCG
CGCGTCGTGTGGCGAGCATCGTTGAGGCGCTGCGTGCGCCGGCGAGCCCGCCGGTTCT
GACCCTGGCGGACATTACCGCGTATGTGGATCACGTTCGTAACGACGTGCAACGTACC
GGTGAGGTTTAATGA*ATCGAT*GAATTCAAAAGATCTTTTAAGAAGGAGATATACAAAT
GAACGATAGCGCGCACCACCAGCGTACCCAAGACGCGACCGGTAACCGTATGCGTGT
GGTTATCAGCGGTATGGATTGCGTGACCCCGCTGGGTGATGATGCGAGCCTGGCGGC
GCACCTGCACCACGGTACCCCGCCGTTCATGCCGGCGCCGGTGCTGCCGGGTGCGGC
GTGGTGCCCGGTTAACGCGTTTGATCCGGCGCAGGCGACCGACCACTGGCGTCACCGT
CGTTACCTGAGCCGTGGTGGCCTGTTTGCGCTGGCGAGCGCGCTGGGTGCGGCGCGTG
ATGCGGGTTGGGGCGCGAGCGCGCCGGATGGTGCGGCGCTGGTGACCGCGCACGGCC

CGAACCTGGACATTACCAGCGATTTTCCGCACTGCCACTGCGACGTTAGCGGTACCGG
TGATACCCAAGGTGATGCGCCGAGCCCGTGCGCGACCAGCACCCATGGTGGCCCGAC
CGCGGCGCACGGTGTGAACGGCATGAACGGTGTTGATGGTCCGGATGAAGCGCGTCA
TGGTCTGAGCCGTGACACCGCGGGTCTGGATCATCCGGGTCTGGATGCGCTGTGGCTG
CTGCGTTGGCTGCCGAACACCGCGGCGAGCGCGATTGCGCAGCGTCTGGGTATTCAC
GGTGAAGGCCTGGTGGTTGGTACCGCGTGCGCGGCGGGTCTGCAAGCGCTGGGCGAA
GCGTATCGTCGTGTGCGTCATGGTCTGGCGCCGACCGTTCTGGCGGCGGGTGGCGATA
GCCGTCTGAGCGCGGGTGGTATGCTGGGTTATGCGCGTGCGGATGCGCTGTGGCACA
GCAACGCGCGTCCGGCGAGCCACGCGGAGGCGCTGCGTGCGATGCGTCCGTTTGATG
CGGCGCATGGTGGCTTTGTGCCGGGTGAAGGTGGCGCGGCGTTCGTTCTGGAGACCG
AAGCGGCGGCGTGCGCGCGTGGTGCGACCATCCACGGCGAGATTCTGGGTTTTGGCG
CGACCCTGGATGGCGCGAGCCTGACCGCGCCGGATGCGACCGCGCACCATGCGGAAT
GCGCGGTGCGTAAAGCGCTGGATGATGCGGGTTTCACCCCGGGTGACATTGCGTGGG
TTGCGGCGCACGGTACCAGCACCCCGCGTGGCGATGCGGCGGAGGCGCTGCTGCTGC
AACGTGTGTTTACCGATGCGGGTCACCGTCCGGCGGTTACCGCGCTGAAAAGCTGGA
CCGGTCACCTGGCGAGCGCGTGCGGTCTGGCGGAATGCGCGCTGATGCTGCGTGCGG
CGCGTTGCGGTGTGCTGCCGAGCATTCGTAACCTGGATACCCCGTGCAGCCCGGCGGC
GGAGGGTCTGGACCTGGTTCGTGAACCGCGTCCGTTCCCGCAGGGTCCGGGTCTGATC
CAAAGCTTCGGTTTTGGTGGCCAGAACGCGGCGCTGATTGTGATGCCGGCGGTTACCC
GTGCGGGCCAATAATGAG*GATCCCTCGAG*GATCGCATGC

Synthetic *bioDD1-5* extended biotin operon:



*CATATG*AGCGAGAAGAACATGCCGGAAAAAACCGTGAGCGAACGTGATGGTGAAAGC
AGCGATGCGCCGCGTGATACCGCGATGAGCGGCGCGCAGAAGCACAGCGCGGAGAA
GATGGGTCGTAAAAGCAGCGCGGCGGTTCCGGCGGCGCGTAGCGCGAGCGTTTTCTG
CTGCTTTCCGGCGGATGATCCGGTGTATGCGGAGCACTTTCCGGGTGCGCCGTGCGTT
CCGGGCAGCCTGCTGATGCAGGCGTTTATCCGTGCGGCGGAACAAATGACCGGTCCG
GCGCCGGATGCGGGCCAGTGGACCTTCACCGGTGTGCGTTTCCGTAAATTTGTTCCGC
CGGGTACCCACGTGTGCACCGTTAGCGCGGTGACCGATGGTAGCGGCAGCGCGTATC
GTTGCACCCTGCTGGTTGATGGTGTGGCGGCGGTTACCGGTACCATTCAATGCGCGTA
ATGA*ACTAGTG*AATTCAAAAGATCTTTTAAGAAGGAGATATACAAATGCGTCTGAGCC
ATACCCGTGGCTGCGTGCTGGTTGCGGGTGGCGGTAGCGCGCTGGGTACCGCGAGCG
CGACCCTGCTGGCGCGTGCGGGCGTGCCGCTGTGCCTGACCGCGGGTAGCGCGGAAG
CGGCGGCGCGTATTCACAACGCGTTCAGCGCGGCGCACCTGAGCTGCCCGCCGGTGC
TGATCGTTCGTATGGAGGACACCACCCGTGCGGATGCGCTGCCGGAACAAGCGGCGG
CGGCGGCGGGCATGCCGGCGGCGTACCTGCTGGACCTGATGCACACCGATCATGAGA
GCCTGCTGGCGAGCGCGAGCCCGGACACCATCAGCGCGTATATTGAAGCGAACATTA
CCTGGCGTGCGCGTCTGGTGCGTGCGGTTAGCCGTCACATGCTGCACCGTCGTTGCGG
CCGTATGGTGTTCGTTAGCAGCACCGCGGCGGCGCTGCCGGCGGCGGGTCAGGGTCT
GTACGCGGCGGCGAAGCTGGCGGCGGAGGCGCTGTATCGTAGCACCGGTACCGAACT
GGCGGCGCGTGGTATTACCACCTGCAGCCTGCGTCTGGGTTATGTTGCGGCGGGTCGT
GGTGCGCGTTATCTGAAACAACACCCGGACCTGCTGCAGCGTATCCCGCTGCAACGTG
CGGTGGATGTTCATGAAGCGGCGGGCGCGCTGGCGTTTCTGCTGAGCGATGGTGCGG
CGGCGATTAACGCGACCGCGATCACCATGGATGGCGGTCTGACCGCGTGCAAGTAAT
GA*GAGCTC*GAATTCAAAAGATCTTTTAAGAAGGAGATATACAAATGACCCAACAAAC
CGCGCCGTGCGCGGTGACCGAGCGTGTTATCGGTATCATTGCGGAAGTGCTGGACATC
CCGCCGCACACCATTACCCCGGAGACCCGTGTTCTGCGTGACCTGGAGACCGAGAGC
ATCGACCTGCTGGAGTTCGGTGTGGGCCTGAACGCGGCGTTTGGCATTCCGGTGGCGG
ACGATACCGTTTTCCTGACCAGCCTGCGTGTTCACCTGGAGGAAGCGCGTCATGCGGA
TGCGGACCCGTACAAGATGATGGCGGGTCGTTATCCGCACCTGAGCTGCGATCGTGTG
GCGTGCATCCTGCGTACCCTGGACGATGGCCCGGTGCTGAGCGTTCAGGACATTGCGG

ATTACGTTCTGCACGCGCTGCGTGCGGGTCAACACGAGAAAGCGCCGAGCCACGCGG
GCTAATGA*ATCGAT*GAATTCAAAAGATCTTTTAAGAAGGAGATATACAAATGCAGGCG
GAGCGTCGTGTGGTTATTACCGGTGCGGGTGCGGTTACCCCGCTGGGTCACACCCCGC
AGGATATTCGTCAAAGCGTGCTGACCGGTCGTACCGTTTTCGAACCGACCAGCGTGCC
GGCGGGTTATGCGCAGGCGCCGGCGCAATTTGATGCGGACGATGCGCTGCGTGGTTG
GCGTCACCGTCGTTATCTGAACCGTGGCGCGGCGATGGCGGTGCATGCGGCGCGTCA
GGCGGCGGGTGCGGCGGGTTTCGCGGCGGGCCTGCCGGGTTACTGCGGTATCTTTGCG
GGCAGCGGTCCGCACCTGGACATGGATAGCGACTATGCGCAGCCGCAACACCTGCCG
CAACGTCACGATACCCTGTGCGCGCTGTGGATGCTGCGTTACCTGCCGAACACCGCGG
CGAGCGCGATCAGCCAACTGCTGGGTGCGCGTGGCGAGAGCCTGGTTACCGGTACCG
CGTGCGCGGCGAGCCTGCAAGCGGTGGGTGAAGCGTTCCGTCGTATTCGTCATGGTGT
GCTGCCGGTTGCGCTGGCGGGTGGCGGTGACAGCCGTCTGAGCTACGGCGGTCTGCT
GGCGTATGCGAAAGCGGGTGCGCTGGCGCCGTGCAGCGAGGATACCCATGCGGCGGC
GGCGGCGGCGTGCCGTCCGTTCGACACCGGTCGTCAGGGTTTTGCGGCGGGTGAAGG
CGCGGCGTTCTTTGTTCTGGAAAGCCTGGATCACGCGCAAAAACGTGGTGCGGTGCCG
CTGGCGGAAGTTTGCGGTTATGGTTGCAGCATGGATGGTCATGCGATGACCGCGCCGC
ACCCGCAGGGTATGCATGCGGAGCAAGCGGTGCGTACCGCGCTGCATGAAGCGGGTG
CGGGTCCGGCGGATGTGACCGCGGTTTGCGCGCATGGTACCGGTACCCGTCGTAACG
ACACCGCGGAGAGCGTTATGCTGACCCGTCTGTTTGCGCCGACCGGTCATCGTCCGCC
GGTGCTGGCGCTGAAGAGCTGGACCGGTCACTGCGCGAGCGCGTGCGGTGCGGTTGA
ACTGGCGGTTCTGCTGGCGTGCGCGCGTCATGGCTTCCTGCCGCCGGTTCGTAACCTG
CAGACCCCGCTGTGCGATAGCCTGAGCTTTGTGCGTCGTATTCACCCGTGCACCGCGG
CGGAGCTGTATGCGGCGGGTCGTCCGGAAAAACTGAGCGCGCTGACCCTGCTGCAGA
GCTTCGGCTTTGGCGGTCAAAACGCGGCGCTGGCGGTTCGTCCGTTCCGTGCGAGCAC
CGCGGCGGCGATCCAAGACGATAACATTCCGGCGTAATGA*GTCGAC*GAATTCAAAAG
ATCTTTTAAGAAGGAGATATACAAATGAAGCACAAAAACACCATCACCCTGCGTTGC
AACGCGGCGCCGGCGCCGGAGACCCTGCACACCCCGGCGACCGCGCTGTTTGTGGTT
CCGGCGGCTGTGAGCGCGGTTTGCGGCAGCAGCATTCGTGGTACCGTGCGTGGTGGC
GGTCACAGCGGCGAGCCGTGGTTCATGACCGAAGCGCTGGCGCAGTTTGCGGCGCTG
CACACCCGTTGGCTGTGCGGTTTCAGCTGCCACGCGTTTCTGCTGAGCGTGCGTCGTA
TGCCGCACATCCGTCTGAGCGCGGCGGACCTGCCGGCGGGTGCGACCGCGCCGGTTC
ACGCGCTGCTGACCAGCCGTACCGGTAGCGATAGCCGTCACGCGTTTGAGTACACCGT
GCAGATGCAACTGGGCGCGCGTACCGTTCAGGGCATTCTGCTGACCGGTACCCTGCCG
TACGACGGTCGTTTCAGCGCGGATCGTCTGCAAAAGCGTTATCGTGAACTGTTTGCGT
GGCTGACCCGTCACTAATGA*GGATCCCTCGAG*GATCAAGCTT