Metabolic flux analysis in *Euglena gracilis*



Sahutchai Inwongwan

Department of Plant Sciences

New College

University of Oxford

Thesis submitted for the degree of *Doctor of Philosophy*

Michaelmas term 2020

Abstract

In this study, ^{13}C metabolic flux analysis and flux balance analysis are used to investigate the metabolic phenotypes of *Euglena gracilis* grown on glucose, acetate, ethanol and pyruvate. A set of metabolic flux maps of *E. gracilis* are presented. These flux maps emphasise the limited contribution of plastidic metabolic processes and the oxidative pentose phosphate pathway in heterotrophic metabolism of *E. gracilis*. Comparison of the flux maps of wild-type and aplastidic *E. gracilis* on glucose reveals that disruption of chloroplast increases demand for the cofactors required for cell maintenance, resulting in higher fluxes through the central pathways of carbohydrate oxidation and a decrease in carbon conversion efficiency in aplastidic cells. Moreover, flux maps of *E. gracilis* on different respiratory substrates emphasise the importance of the glyoxylate shunt and malate dehydrogenase in supporting anaplerosis when grown on acetate and ethanol, but not on pyruvate in which the precursor for gluconeogenesis is mainly provided by pyruvate carboxylase. The flux maps reported in this study reveal the impact of chloroplast disruption on the metabolic phenotype of *Euglena* and provide new insights into the metabolism of substrates differing in carbon composition and chemical redox potential.

Acknowledgements

I would love to express my deepest gratitude to Dr Nick Kruger, Prof. George Ratcliffe and Dr Ellis O'Neil for their mentorship, guidance, encouragement, support and understanding throughout these unforgettable years of DPhil.

I would like to thank Prof. Hugh Dickinson for his support on TEM, and Dr Sanu Shameer and Mann for their advice on FBA.

I acknowledge the financial support from the Development and Promotion of Science and Technology Talents Project (DPST), Thailand.

Thank you to Shaobo and Steffi – I surely could not have made it through a year in Oxford without you. Thanks to Anya, Martin, P' Santamind, P' Beer and others for making Oxford feel a bit more like home, and to Kalamung, Dao, Ja and Reena for always being there whenever I need the emotional supports.

Lastly, I would like to thank my family who always supports me on every step I take and every decision I make. I am nothing without you.

Abbreviations

ADH	Alcohol dehydrogenase
ALDH	Acetaldehyde dehydrogenase
AcCoA	Acetyl coenzyme A
aq	Aqueous
C	Carbon
CCE	Carbon conversion efficiency
CoA	Coenzyme A
cTP	Chloroplast transit peptides
DW	Dry weight
EC	Enzyme commission
F6P	Fructose 6-phosphate
FBA	Flux balance analysis
G6P	Glucose 6-phosphate
GC-MS	Gas chromatography mass spectrometry
Glyc3P	Glyceraldehyde 3-phosphate Glyc3P
IMP	Inosine monophosphate
KEGG	Kyoto encyclopaedia of genes and genomes
KO	KEGG orthology
MDH	Malate dehydrogenase
MFA	Metabolic flux analysis

MID	Mass isotopic distribution
mTP	Mitochondrial transit peptides
nonoxPPP	Non-oxidative pentose phosphate pathway
OAA	Oxaloacetate
oxPPP	Oxidative pentose phosphate pathway
PC	Pyruvate carboxylase
PDH	Pyruvate dehydrogenase
PEP	Phosphoenolpyruvate
PEPC	Phosphoenolpyruvate carboxylase
PEPCK	Phosphoenolpyruvate carboxykinase
PPP	Pentose phosphate pathway
RMSD	Root mean square deviation
SM	Streptomycin
SP	Signal peptide
SSR	Sum of squared residuals
TBDMS	tert-Butyldimethylsilyl
TCA	Tricarboxylic acid
TEM	Transmission electron microscopy
TMP	Thymidine monophosphate
TP	Transit peptide
UMP	Uridine monophosphate

Table of contents

Chapter 1 | Introduction 11

1.1 Outline of the thesis 11

1.2 Background 13

1.2.1 *Euglena* 13

1.2.2 Evolution of *Euglena* 15

1.2.3 Industrial significance of *Euglena* 17

1.2.4 Growth and cultivation of *Euglena* 18

1.2.5 The quantification of metabolic fluxes 21

1.2.6 FBA 23

1.2.7 ^{13}C MFA 24

1.3 Objectives 30

Chapter 2 | Materials and methods 31

2.1 Materials 31

2.2 *E. gracilis* cultivation 31

2.3 Generation of aplastidic *E. gracilis* 32

2.4 Sample preparation for transmission electron microscopy 33

2.5 Substrates utilisation analysis 34

2.5.1 ^{14}C Glucose uptake assay 35

2.5.2 Pyruvate quantification 35

2.5.3 Acetate quantification 36

2.5.4 Ethanol quantification 37

2.6 Biomass composition analysis 38

2.6.1 Protein 39

2.6.2 Total carbohydrate 39

2.6.3 Insoluble carbohydrate 40

2.6.4 Total Lipid 41

2.7 Carbon conversion efficiency analysis 42

2.8 Steady state ^{13}C MFA 43
2.8.1 ^{13}C labelling experiment 43
2.8.2 GC-MS analysis of ^{13}C labelled sample 43
2.8.3 MID extraction 45
2.8.4 ^{13}C Metabolic flux modelling 45
2.9 Metabolic network model construction and FBA 47
2.9.1 Metabolic network model construction 47
2.9.2 Model analysis 48
Chapter 3 | Metabolic model of *E. gracilis* 49
3.1 Introduction 49
3.1.1 Overview of the metabolic pathways of *E. gracilis* from transcriptome analyses 49
3.1.2 Functions of the *Euglena* plastid in heterotrophic conditions 50
3.1.3 Metabolic network reconstruction of *Euglena* 53
3.1.4 Objectives 54
3.2 Subcellular localisation of enzymes of the central metabolic pathways of *E. gracilis* 55
3.2.1 Prediction of subcellular localisation of *E. gracilis* proteins 55
3.2.2 Organisation of the central metabolic pathways in *Euglena* 57
3.3 Construction of the metabolic network model of *E. gracilis* 60
3.3.1 Drafting the metabolic network of *E. gracilis* 60
3.3.2 Model curation 63
3.4 Formulating biomass synthesis reactions for metabolic flux analysis 64
3.4.1 Protein 66
3.4.2 Carbohydrate 67
3.4.3 Lipid 68
3.4.4 Nucleic acid 72
3.4.5 The biomass synthesis reactions of *E. gracilis* grown on glucose 73
3.5 Testing the stoichiometric model of the metabolic network in *E. gracilis* 74
3.6 Discussion 76

3.6.1 The metabolic pathway distribution of *E. gracilis* is substantially different from other photosynthetic organism, which results in its ability to survive in the absence of a plastid. 76
3.6.2 The reconstruction of the metabolic network model of *E. gracilis* provides the basis for flux analysis under heterotrophic conditions. 77
3.7 Conclusion 78
Chapter 4 | Metabolic phenotypes of wild-type and aplastidic *E. gracilis* grown on glucose 79
4.1 Introduction 79
4.1.1 Glucose metabolism in *Euglena* 79
4.1.2 Chloroplasts of *Euglena* 80
4.1.3 Objectives 82
4.2 Physiological characterisation of wild-type and aplastidic *E. gracilis* grown on glucose 83
4.2.1 Anatomy of *E. gracilis* strains 83
4.2.2 Growth of wild-type *E. gracilis* grown heterotrophically on glucose 86
4.2.3 Comparison of growth characteristics of wild-type and aplastidic *E. gracilis* 88
4.2.4 Biomass composition of *E. gracilis* grown heterotrophically on glucose 90
4.3 ^{13}C MFA of *E. gracilis* grown on glucose 93
4.3.1 Flux map of wild-type *E. gracilis* grown on glucose 93
4.3.2 Flux map of aplastidic *E. gracilis* grown on glucose 96
4.3.3 Comparing flux maps of wild-type and aplastidic *E. gracilis* grown on glucose 98
4.3.4 CCE of wild-type and aplastidic *E. gracilis* grown on glucose 98
4.3.5 Cofactor provision in wild-type and aplastidic *E. gracilis* grown on glucose 100
4.4 FBA of wild-type and aplastidic *E. gracilis* grown on glucose 104
4.4.1 FBA of *E. gracilis* strains grown on glucose with biomass constraint 104
4.4.2 Estimating cell maintenance of *E. gracilis* strains grown on glucose 108
4.4.3 Accounting for maintenance in FBA of wild-type and aplastidic *E. gracilis* 110
4.4.4 Metabolic efficiency of *E. gracilis* strains grown on glucose based on CCE 112

4.5 Discussion 114
4.5.1 The limited contribution of plastidic metabolic processes, the oxPPP and the glyoxylate shunt in heterotrophic glucose metabolism in *E. gracilis*. 114
4.5.2 Chloroplast disruption increases cell maintenance costs and decreases the metabolic efficiency of *E. gracilis* during growth on glucose. 116
4.6 Conclusion 118
Chapter 5 | Metabolic phenotypes of *E. gracilis* grown on C2 and C3 substrates 120
5.1 Introduction 120
5.1.1 Acetate and ethanol metabolism in *E. gracilis* 121
5.1.2 Pyruvate metabolism 123
5.1.3 Objectives 123
5.2 Physiological characterisation of *E. gracilis* grown on acetate, ethanol and pyruvate 124
5.2.1 Growth of heterotrophically *E. gracilis* cultures on acetate 124
5.2.2 Growth of heterotrophically *E. gracilis* cultures on ethanol 125
5.2.3 Growth of heterotrophically *E. gracilis* cultures on pyruvate 127
5.2.4 Comparison of growth characteristics of *E. gracilis* grown on acetate, ethanol and pyruvate 128
5.2.5 Biomass composition of *E. gracilis* grown heterotrophically on acetate, ethanol and pyruvate 132
5.3 ^{13}C MFA of *E. gracilis* grown on acetate, ethanol and pyruvate 135
5.3.1 Flux map of *E. gracilis* grown on acetate 135
5.3.2 Flux map of *E. gracilis* grown on ethanol 138
5.3.3 Flux map of *E. gracilis* grown on pyruvate 141
5.3.4 Comparison of ^{13}C MFA flux maps of *E. gracilis* grown on acetate, ethanol and pyruvate 143
5.3.5 CCE of *E. gracilis* grown on acetate, ethanol and pyruvate 144
5.3.6 Cofactor production by *E. gracilis* grown on acetate, ethanol and pyruvate 146
5.4 FBA of *E. gracilis* grown on acetate, ethanol and pyruvate 149

5.4.1 FBA of *E. gracilis* grown on acetate, ethanol and pyruvate with biomass constraint 140
5.4.2 Estimating cell maintenance of *E. gracilis* grown on acetate, ethanol and pyruvate 154
5.4.3 Accounting for maintenance in FBA of *E. gracilis* grown on acetate, ethanol and pyruvate 156
5.4.4 Metabolic efficiency of *E. gracilis* grown on acetate, ethanol and pyruvate 157
5.5 Discussion 159
5.5.1 ^{13}C MFA flux maps reveal the metabolic flux phenotypes of *E. gracilis* and emphasise the limited contribution of plastid metabolic processes and the oxPPP in heterotrophic metabolism. 159
5.5.2 Despite the same carbon composition and similar metabolic network entry points, the metabolic phenotypes of *E. gracilis* grown on acetate and ethanol are substantially different. 161
5.5.3 Flux through MDH in *E. gracilis* was elevated when grown on acetate and ethanol, responding to the precursors and NADPH demand for biomass production. 164
5.6 Conclusion 165
Chapter 6 | General discussion 167
6.1 Summary 167
6.2 The metabolic flexibility of the central metabolic network of *E. gracilis* 168
6.3 The implementation of MFA and FBA for analysing the metabolic network of *E. gracilis* 171
6.4 Future direction 173
References 177
Appendices 204

Chapter 1 | Introduction

1.1 Outline of the thesis

Euglena has long been a focus of research as a biological model and for its diverse industrial applications. Despite this, understanding of the metabolic network of *Euglen*a and its capacity is limited. Analysis of the *Euglena* genome is incomplete and restricted by its size and complexity (Ebenezer et al., 2019), reflecting its origin in a series of secondary endosymbiosis events and horizontal gene transfer during the course of evolution. Transcriptome sequencing has been used as an alternative to genome structure analysis (O'Neill et al., 2015) to study the genetic capacity of *Euglena.* The transcriptomes indicate that many genes are regulated at the post-transcriptional level. Thus, analysis based on only transcriptomic information may not be sufficient to gain insights into the metabolic network of *Euglena* (Yoshida et al., 2016). To obtain a greater understanding of the metabolic capacity of *Euglena*, in this study, ^{13}C isotopic labelling-based metabolic flux analysis (^{13}C MFA or MFA) and constraint-based metabolic flux modelling (flux balance analysis – FBA), relatively direct techniques for determining metabolic phenotypes (Wiechert, 2001), have been used to investigate the metabolic phenotypes of *Euglena gracilis* grown on different carbon substrates.

Euglena is able to utilise a variety of organic carbon compounds as respiratory substrates (Buetow, 1989). These substrates vary in their carbon composition and chemical redox potential, and therefore requiring (or permitting) different distributions of flux through

the reactions of the central metabolic network to maintain cellular processes and support growth. To examine this, I have chosen to investigate the metabolism of glucose, acetate, ethanol and pyruvate. These compounds were chosen for two reasons: first, each can be used as the sole carbon substrate to support heterotrophic growth of *Euglena*; and secondly, each enters the metabolic network by different routes. Defining the metabolic phenotypes of *Euglena* on these diverse respiratory substrates will establish the extent to which the network responds to variation in metabolic inputs.

In addition, *E. gracilis* is known to survive without its chloroplasts. This unusual feature of *E. gracilis* physiology offers the potential of examining the fundamental essentiality of plastids and the subcellular compartmentation of pathways of carbohydrate oxidation in photoautotrophic eukaryotes. Comparing metabolic phenotypes of wild-type and chloroplast-less strains will lead to the understanding of the roles of chloroplasts in heterotrophic carbon metabolism in *Euglena*.

The framework shown in Figure 1.1 summarises the structure of the approach used in this project. In brief, for each carbon substrate, three major experiments were conducted: (i) the analysis of growth and biomass composition of the *E. gracilis* culture; (ii) isotopic labelling experiment; and (iii) metabolic network modelling. The results from these experiments are combined to undertake metabolic flux analysis in *Euglena* through both ^{13}C MFA and FBA.

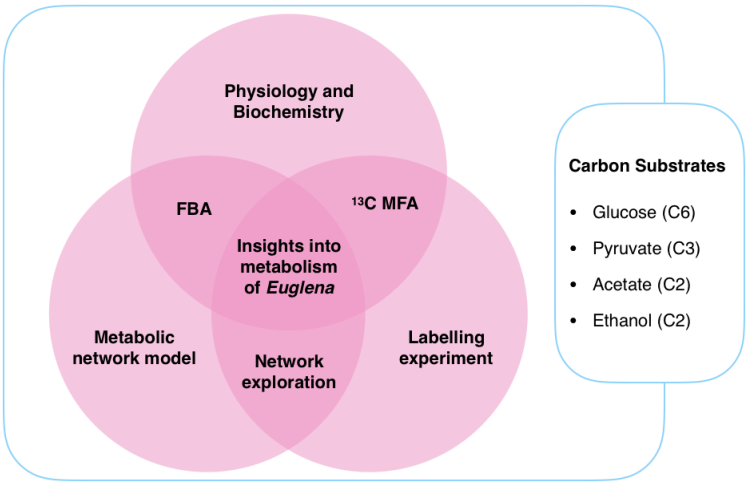


Figure 1.1. The framework of the research. To undertake metabolic flux analysis in *Euglena* through both ^{13}C MFA and FBA, three major experiments were conducted: (i) the analysis of growth and biomass composition of the *E. gracilis* culture; (ii) isotopic labelling experiment; and (iii) metabolic network modelling.

1.2 Background

1.2.1 *Euglena*

Euglena is a unicellular, protozoan, flagellate alga, which has long been one of the most intriguing groups of organisms in terms of evolution due to its combination of both "plant-" and "animal-" like features (Buetow, 1989). *Euglena* can obtain energy hetero-, mixo- or autotrophically, and its cultivation is relatively easy, fast and well established. This alga can be found in a broad range of ecological niches including fresh water, brackish water, snow, acidic conditions, alkaline conditions, aerobic and anaerobic environments (Kitaoka et al., 1989). The typical morphology (Figure 1.2) of *Euglena* is characterised by its elongated spherical, ridged, and spirally twisted body shape which greatly varies in size. Two flagella with different lengths are present; the longer one has a sweeping movement and the shorter

flagellum does not emerge from the reservoir, a cavity opening via a gullet. *Euglena* has either grass-green colour or can be colourless and may have pyrenoids. Their chloroplasts are found to be greatly diverse in shape and size. *Euglena* store carbohydrate principally in granules of a β-1,3-glucan, called paramylon. Near to the gullet lies a stigma, the light tracking organelle, composed of a hematochrome granule, and *Euglena* has a proteinaceous inner wall with striae called the pellicle (Wolken, 1961).

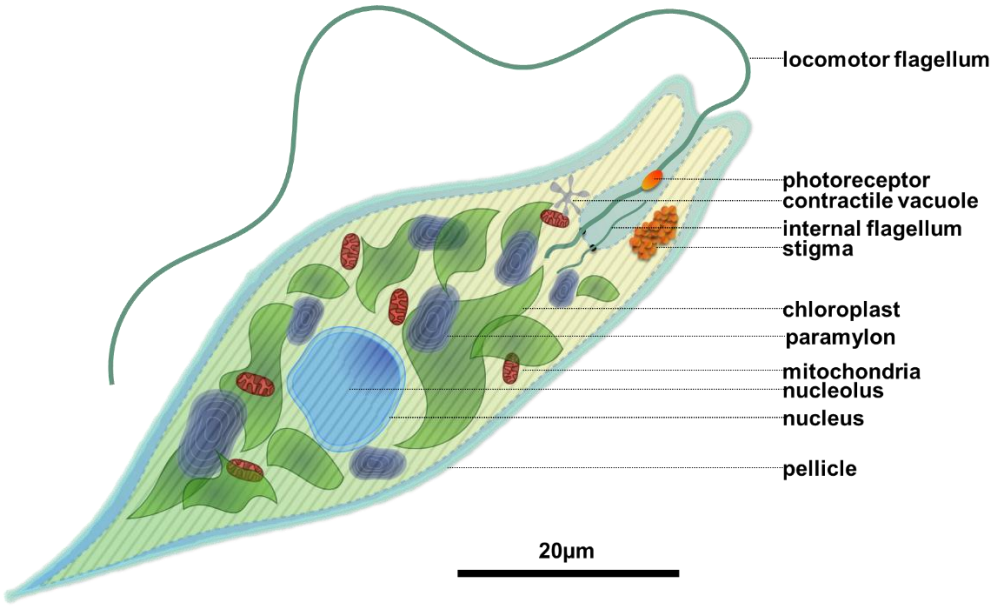


Figure 1.2. Morphology of *Euglena*

Due to the chimeric features of *Euglena* (Wolken, 1961), the classification and evolutionary history of *Euglena* were questioned for decades. Euglenoids are now classified into the special group Excavata, superphylum Discoba, subphylum Euglenozoa (Adl et al.,

2012), which is phylogenetically very distant from green algae, plants or animals. *Euglena gracilis* is the most studied species in the *Euglena* genus and it is considered to be a good model organism for several reasons. *E. gracilis* cultivation is relatively easy, fast and well established and it has remarkable metabolic plasticity, which allows it to grow in various conditions (Einicker-Lamas et al., 2002). Furthermore, many aspects of *E. gracilis* biology have been intensively studied, including industrial potential, physiology and biochemistry (Buetow, 1989). Most importantly, *E. gracilis* has two sets of *de novo* transcriptomic data published (O'Neill et al., 2015, Yoshida et al., 2016) which will be useful for metabolic network studies. Hence, *E. gracilis* was chosen as the model organism for this thesis.

1.2.2 Evolution of *Euglena*

The taxonomic complexity of *Euglena* has arisen from a series of endosymbiosis events during its evolution (Linton et al., 2010). Studies of the distribution of homologues of *2770 ESTs* of *E. gracilis* have shown that the photoautotrophic eukaryotes are the ancestral precursors of *E. gracilis* (Ahmadinejad et al., 2007). *Euglena* first split out from the Euglenozoa a billion years ago from the ancestral eukaryotic protozoa that had mitochondria, which contain some alphaproteobacteria derived genes (Parfrey et al., 2011). After the horizontal transfer of genes from a since-lost red algal endosymbiont to the nuclear genome (Maruyama et al., 2011a), the endosymbiosis of a eukaryotic green alga (Martin et al., 1992) happened. This event brought many genes involved in the chloroplast function and maintenance to *Euglena* cell (Figure 1.3). The transcriptome of *Euglena*, suggests many genes were acquired from distantly related species and the genetic control mechanisms in *Euglena* involve genes which are as complex as those in multicellular eukaryotes (O'Neill et al., 2015). *Euglena*

contains secondary chloroplasts with three layers of envelope membrane that were acquired relatively recently during evolution (~600 MYA (Jackson et al., 2018)). The genome of *E. gracilis* is estimated to be around 500 Mb in size, with large amounts of highly repetitive sequences (Ebenezer et al., 2019), which leads to difficulty in genome sequencing and analysis. The chimeric genome of *Euglena* derived from endosymbiosis of multiple endosymbionts, which could lead to its resilient metabolic flexibility (Ebenezer et al., 2017).

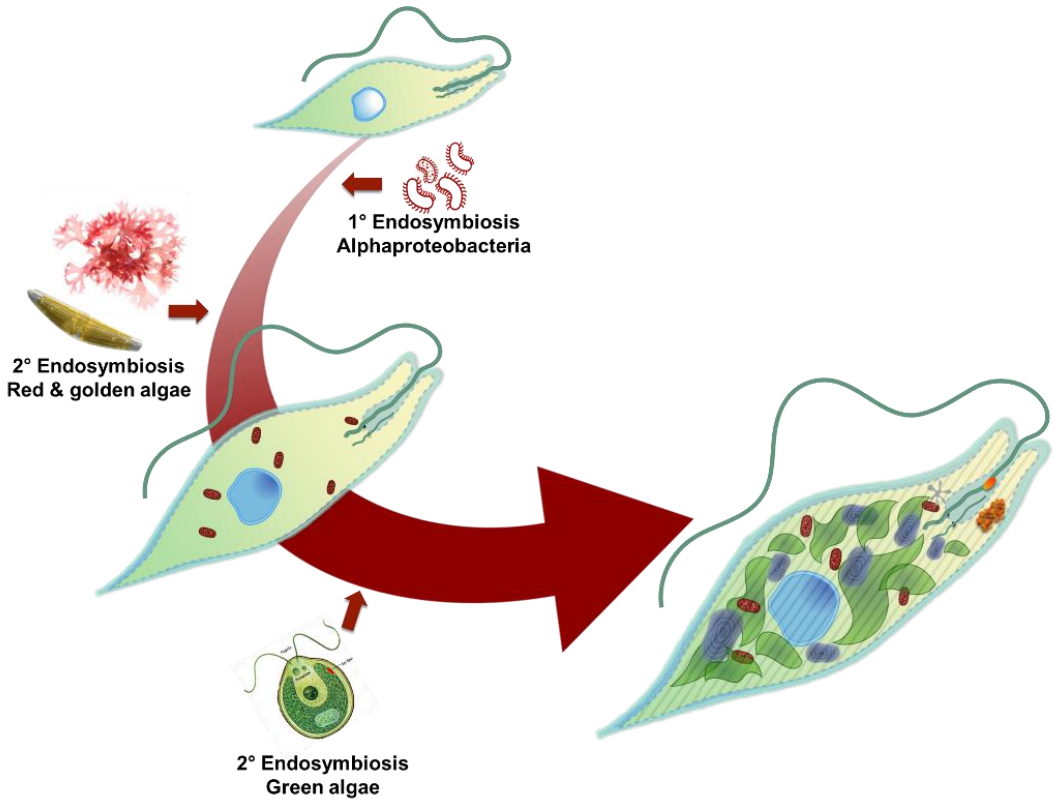


Figure 1.3. The evolution of *Euglena*. The primary endosymbiosis of alphaproteobacteria resulted in the origin of mitochondria. The multiple events of secondary endosymbiosis of red and golden algae provided *Euglena* the extensive nuclear genetic materials. The secondary chloroplasts of *Euglena* were acquired relatively recently from the secondary endosymbiosis of a green alga.

1.2.3 Industrial significance of *Euglena*

Euglena is a promising organism for industrial application due to its ability to produce various bioactive and valuable compounds. *Euglena* has also been considered as a potential dietary supplement due to its capacity to produce various nutritious compounds. For example, *Euglena* is a useful source of proteins and polyunsaturated fatty acids (Barsanti et al., 2000, Korn, 1964). *Euglena* also produces vitamin A, vitamin C and vitamin E (Kusmic et al., 1998, Ogbonna et al., 1998, Takeyama et al., 1997). *E. gracilis* is a rich source of β-1,3-glucan, which can make up to 85% of cell dry weight (DW) (Inui et al., 1982), accumulated as the storage carbohydrate polymer paramylon. Paramylon is a bioactive compound (Koizumi et al., 1993), with sulphated derivatives of paramylon from *Euglena* having anti-HIV activity. In addition, inclusion of paramylon in human and animal diets has a cholesterol-lowering effect (Barsanti et al., 2000, Wang et al., 1997) and β-1,3-glucans have been shown to stimulate the immune system by the cytokine-related immunopotentiating activities (Kondo et al., 1992).

Besides its uses in the food and medicine industries, *Euglena* has been studied for wastewater treatment systems, energy sources and environmental bioindicators. *Euglena* sp. isolated from sewage treatment plants had higher nutrient removal capability and growth rate compared to other algae, with biomass containing up to 24.6% (w/w) lipid and a mass density of 1.24 g L^{-1} (Mahapatra et al., 2013). These results indicated that *Euglena* could be considered as a viable source for biofuel production from wastewaters. *E. gracilis* has also been used as a bioindicator in a comet assay for characterising genotoxic effects, based on DNA fragmentation and quantification of antioxidants in *Euglena* (Li et al., 2014).

1.2.4 Growth and cultivation of *Euglena*

Euglena can grow in an extremely diverse set of conditions, such as autotrophic, heterotrophic or mixotrophic conditions (Takeyama et al., 1997), and aerobic or anaerobic, and acidic environments (Yamane et al., 2001). While growing photo-autotrophically, *Euglena* uses CO_2 as a carbon source for sugar synthesis through photosynthetic reactions. On the other hand, in heterotrophic conditions, *Euglena* has the ability to utilise a range of organic carbon substrates as the sole source of carbon and energy; these include sugars, short-chain alcohols, short-chain fatty acids and amino acids (Table 1.1). Interestingly, the physiological properties of *Euglena* differ significantly, depending on the growth conditions of the culture, specifically the organic carbon sources for mixo-/heterotrophic cultivation.

Table 1.1 Organic carbon substrates usable by *Euglena*

Substrate	Type of compound	References
alanine	amino acid	(Buetow, 1989)
aspartate	amino acid	(Buetow, 1989)
glutamate	amino acid	(Barsanti et al., 2000, Buetow, 1989, Kempner and Miller, 1965b, Kusmic et al., 1998, Navarro et al., 1997, Oda et al., 1982)
sucrose	disaccharide	(Buetow, 1989)
maltose	disaccharide	(Ivušić and Šantek, 2015)
lactose	disaccharide	(Ivušić and Šantek, 2015)
trehalose	disaccharide	(Maréchal and Belocopitow, 1972)
palmitate	long-chain fatty acid	(Blum and Wittels, 1968)
glucose	monosaccharide	(Barsanti et al., 2000, Blum and Wittels, 1968, Buetow, 1989, Feldman and Bruce,

		1972, Ogbonna et al., 1998, Yamane et al., 2001)
fructose	monosaccharide	(Blum and Wittels, 1968, Buetow, 1989)
galactose	monosaccharide	(Barsanti et al., 2000, Oda et al., 1982)
lactate	organic acid	(Buetow, 1989, Calvayrac et al., 1981, Feldman and Bruce, 1972, Navarro et al., 1997, Ogbonna et al., 1998)
acetate	organic acid	(Barsanti et al., 2000, Blum and Wittels, 1968, Buetow, 1989, Feldman and Bruce, 1972, Kusmic et al., 1998, Ogbonna et al., 1998)
ethanol	short-chain alcohol	(Barsanti et al., 2000, Blum and Wittels, 1968, Buetow, 1989, Kusmic et al., 1998, Ogbonna et al., 1998)
butanol	short-chain alcohol	(Buetow, 1989)
propanol	short-chain alcohol	(Buetow, 1989)
propionate	short-chain fatty acid	(Buetow, 1989)
butyrate	short-chain fatty acid	(Buetow, 1989)
pyruvate	Tricarboxylic acid (TCA) cycle intermediate	(Barsanti et al., 2000, Blum and Wittels, 1968, Buetow, 1989, Feldman and Bruce, 1972, Kusmic et al., 1998, Ogbonna et al., 1998)
oxaloacetate	TCA cycle intermediate	(Blum and Wittels, 1968)
succinate	TCA cycle intermediate	(Barsanti et al., 2000, Blum and Wittels, 1968, Buetow, 1989, Feldman and Bruce, 1972, Kusmic et al., 1998)
fumarate	TCA cycle intermediate	(Buetow, 1989)
malate	TCA cycle intermediate	(Barsanti et al., 2000, Navarro et al., 1997, Oda et al., 1982, Yamane et al., 2001)

The effects of different carbon substrates on the physiology of *Euglena* have been studied for decades. Carbon sources have marked effects on various physiological characteristics including, growth, respiration, energy metabolism, circadian rhythm, stress responses and biosynthesis (Buetow, 1989). The growth rate of *Euglena* varies depending on the source of carbon. For example, it was reported growth rates of *E. gracilis* differ when grown on six different carbon sources, including acetate, glucose, glutamate, succinate, pyruvate and ethanol (Kusmic et al., 1998). Furthermore, amino acid utilisation in the bleached *E. gracilis* strain, in which the chloroplast has been eliminated by chemical or heat treatment, was investigated. This study showed that, among 21 amino acids, glutamate gave the highest growth rate, while methionine, cysteine, threonine, leucine and phenylalanine each inhibited growth of *Euglena* using glutamate as the carbon and nitrogen source (Oda et al., 1982). Besides carbohydrate and amino acid compounds, the fact that *Euglena* has the ability to use long chain hydrocarbons, like palmitic acid, as a sole organic carbon source showed its capacity to adapt to using various substrates (Blum and Wittels, 1968).

Carbon source also has a remarkable effect on *Euglena* respiration. Two-carbon compounds like ethanol or acetate were found to result in significantly higher rates of oxygen consumption, while glucose and intermediates of the TCA cycle gave lower oxygen consumption. For other compounds, including propionate, butyrate, lactate, and pyruvate, the oxygen consumption was at an intermediate level (Buetow, 1989). Adding acetate to the autotrophic culture was shown to have an effect on the period of the circadian rhythm of phototaxis of *E. gracilis*, although this might be an effect caused by the transition from an autotrophic state to a mixotrophic state, which would have required the switching of

metabolic pathways (Feldman and Bruce, 1972). Succinate, lactate or pyruvate also have an effect on temporary phototaxis inhibition. On the other hand, glucose affected phototaxis or rhythmicity much less (Feldman and Bruce, 1972). In addition, carbon sources were shown to have an effect on the heavy metal response in *Euglena*, with sensitivity to heavy metals dependent on the carbon sources in the culture media (Navarro et al., 1997).

It can be seen that *Euglena* can adapt to different organic carbon sources and physiological changes in *Euglena* due carbon source have long been documented. However, there has been only limited investigation into the effect of carbon substrate on the operation of metabolic pathways that must adjust to support these physiological changes in *E. gracilis*.

1.2.5 The quantification of metabolic fluxes

Genome, transcriptome and proteome analyses are the most important approaches in systems biology. In the race to obtain these data, alternative techniques for determination of metabolic phenotypes have been increasingly explored and have become major tools in metabolic engineering (Wiechert, 2001). Metabolism is the set of biochemical process to maintain the cellular activity. The term “flux” or “metabolic flux” has been widely used in biochemistry to represent the flow of substrates through reactions catalysed by enzymes in the metabolic network. The distribution of fluxes in the metabolic network is usually described as a flux map (Kruger and Ratcliffe, 2015).

Metabolic flux represents the final integrative outcome of many different levels of cellular regulation that contribute to the flow of metabolites through the metabolic network. Thus metabolic fluxes are a function of gene expression, translation, posttranslational protein

modifications, and protein-metabolite interactions (Nielsen, 2003). Thus, it is the fluxes between metabolites that define the actual functional metabolic phenotype (Ratcliffe and Shachar-Hill, 2005). Measuring metabolic fluxes reveals the metabolic phenotype of an organism under the specific conditions. The results help in understanding how the cells metabolically respond or operate, and can also be used to identify targets for metabolic engineering to improve productivity in organisms of interest (Guo et al., 2015).

Even though fluxes are useful to depict the characteristics of the physiological state of the cell, measuring the actual flow of metabolites or real-time activity of the intracellular enzymes is difficult. There are three major strategies for metabolic flux determination, each strategy is based on different approaches and giving different outcomes: kinetic modelling, FBA and MFA or ^{13}C MFA (Kruger and Ratcliffe, 2015).

Kinetic modelling is an approach for modelling the chemical reactions in the metabolic network based on the kinetic parameters of the enzymes in the network. This method can predict the responses of well-characterised fragments of the network, as well as the provision of precursors or demand for products, with high predictive value. However, this approach is limited by the availability of known kinetic parameters and incomplete description of the regulatory processes (Kruger and Ratcliffe, 2015). This technique is particularly difficult to apply to *E. gracilis* due to the lack of kinetic characterisation of its enzymes and the poor understanding in its regulatory processes. Thus, this thesis specifically focuses on exploiting the ^{13}C MFA and FBA techniques to investigate the metabolic phenotypes of *E. gracilis.*

1.2.6 FBA

FBA has been widely used for the prediction of metabolic fluxes based on the genetic capacity of the system, involving mathematical calculation including stoichiometry and thermodynamics (Orth et al., 2010, Fang et al., 2020). FBA has the ability to estimate the fluxes of metabolites through a large-scale metabolic model, usually genome-based reconstructions, which allows FBA to predict the rates of growth or production of compounds of interest. The network is reconstructed as a stoichiometric model that contains all of the metabolic reactions based on the genome annotation of an organism (Thiele and Palsson, 2010).

FBA predicts flux based on the assumption that the metabolic network has evolved towards an evolutionary optimum. The function that is set to be optimised is called the objective function (Kauffman et al., 2003), which defines the scope of fluxes and to optimise and predict the flow through the network. Maximising the synthesis of biomass is the most commonly used objective for FBA, with the proportions of metabolites needed for biomass components defined by the biomass composition (Feist and Palsson, 2010). With biomass constraints and the objective to maximise biomass production, FBA can generate a range of optimal solutions. In order to narrow down to the most realistic solution, running a second optimisation step, such the parsimonious FBA (pFBA) (Lewis et al., 2010), can be applied. This approach is based on the principle of simplicity to minimise the sum of reaction fluxes while maintaining the optimal primary objective (Foguet et al., 2019).

Extensive availability of transcriptomic, proteomic and metabolomic data increases the reliability of FBA (Thiele and Palsson, 2010). However, FBA can only predict net flux, rather

than both forward and backward directions (or the exchange flux) in a reversible reaction (Orth et al., 2010). For the past decade, FBA has been widely used for studying plant metabolism (Shameer et al., 2020, Shameer et al., 2019, Cheung et al., 2015). Several flux-balance models of photosynthetic organisms have been published, such as *Arabidopsis* (Cheung et al., 2015), tomato (Shameer et al., 2020) and *Chlamydomonas* (Cogne et al., 2011).

FBA of a diel genome-scale model of *Arabidopsis* leaf was exploited to explore the flexibility of the network. The analysis could identify alternative flux modes in the Calvin-Benson cycle that revealed the alternative carbon stores in leaves (Cheung et al., 2015). FBA is also able to accurately predict the metabolic phenotype of *Arabidopsis* cells under elevated temperature and hyper-osmotic conditions. This study demonstrated that cell maintenance costs was substantially increased only under the elevated temperature, and not under hyper-osmotic conditions (Cheung et al., 2013). A whole plant model of tomato metabolism has been developed based on the genome-scale model and FBA (Shameer et al., 2020). The model was able to accurately predict the levels of flux through the central metabolic pathways throughout fruit development. In *Chlamydomonas reinhardtii*, FBA suggests the presence of a light-driven respiration functioning in light limited conditions (Cogne et al., 2011). These studies demonstrate the utility of FBA in predicting the metabolic phenotypes of photosynthetic cells.

1.2.7 ^{13}C MFA

MFA or ^{13}C-based MFA is a quantitative method for deducing metabolic fluxes, by tracking the fate and rearrangement of labelled atoms throughout the metabolic network. ^{13}C MFA does not only provides net fluxes, but is also potentially able to define the

unidirectional forward and reverse fluxes in reversible reactions. The technique is based on an isotopic labelling approach typically involving the supply of ^{13}C-substrates to the cell. When the labelled materials distribute throughout the network and cells reach both metabolic and isotopic steady states, the labelling pattern is recorded using nuclear magnetic resonance spectroscopy (NMR) or, more commonly, mass spectrometry (MS) (Wiechert, 2001, Kruger and Ratcliffe, 2015). This data is combined with the metabolic model to determine *in vivo* fluxes. The empirical flux map generated from the mass isotopic distribution (MID) patterns of labelled cells at isotopic steady state provides quantitative metabolic fluxes for the specific state of cells and in specific growth conditions. However, this flux map has very limited ability to predict the effect of a network change. In addition, the MFA experimental procedure requires a considerable experimental effort in order to establish reliable protocols for the analysis of the chosen cellular system (Kruger et al., 2012).

To conduct the steady state ^{13}C labelling experiment, it is necessary to understand the concepts of isotopic and metabolic steady state. Currently, most ^{13}C MFA experiments are performed in isotopic and metabolic steady state (steady state MFA), in which all metabolic fluxes and isotopic abundance in the cellular metabolites are assumed to be constant at the point when the labelled metabolites are extracted from the tissue (Antoniewicz et al., 2006). Metabolic steady state and isotopic steady state can be achieved in batch culture during the exponential growth phase, where the growth rate is steady (Zamboni et al., 2009). For a better visual understanding, an overview of the ^{13}C MFA strategy in *E. gracilis* is presented in Figure 1.4.

First, in the labelling experiment, *E. gracilis* is cultivated heterotrophically on a ^{13}C labelled carbon substrate. The choice of ^{13}C-labelled substrate for ^{13}C MFA depends the objectives of the experiment (Crown et al., 2016). The labelled cells biomass is harvested at late exponential growth phase and the isotope distribution is measured by gas chromatography-mass spectrometry (GC-MS). The fragments of tert-Butyldimethylsilyl (TBDMS) derivatised amino acids in total cell protein are commonly used for ^{13}C MFA (Antoniewicz et al., 2007). However, the determination of the MID or the amount of label taken up (isotopic abundance) is complicated by the naturally occurring stable isotopes including: ^{13}C, 1.1%; ^{2}H, 0.0115%; ^{17}O, 0.038%; ^{18}O, 0.2%; ^{15}N, 0.366%; and ^{34}S, 4.2% (Poskar et al., 2012). Thus, when analysing the MID of an isotopically labelled metabolite or fragment, systematic correction for natural isotopic effects must be taken in account. This can be done automatically using several well-established tools (Wahl et al., 2004, Poskar et al., 2012). After mass isotope correction, the MID is subsequently used for flux analysis, in conjunction with the stoichiometric model of the central metabolic pathways. The parameters (fluxes) within the model are systematically varied to obtain the best fit between the simulated and measured MIDs (Guo et al., 2015), using a range of available software, e.g. 13CFLUX (Wiechert, 2001), FiatFlux (Zamboni et al., 2005), influx_s (Sokol et al., 2011) or INCA (Young, 2014).

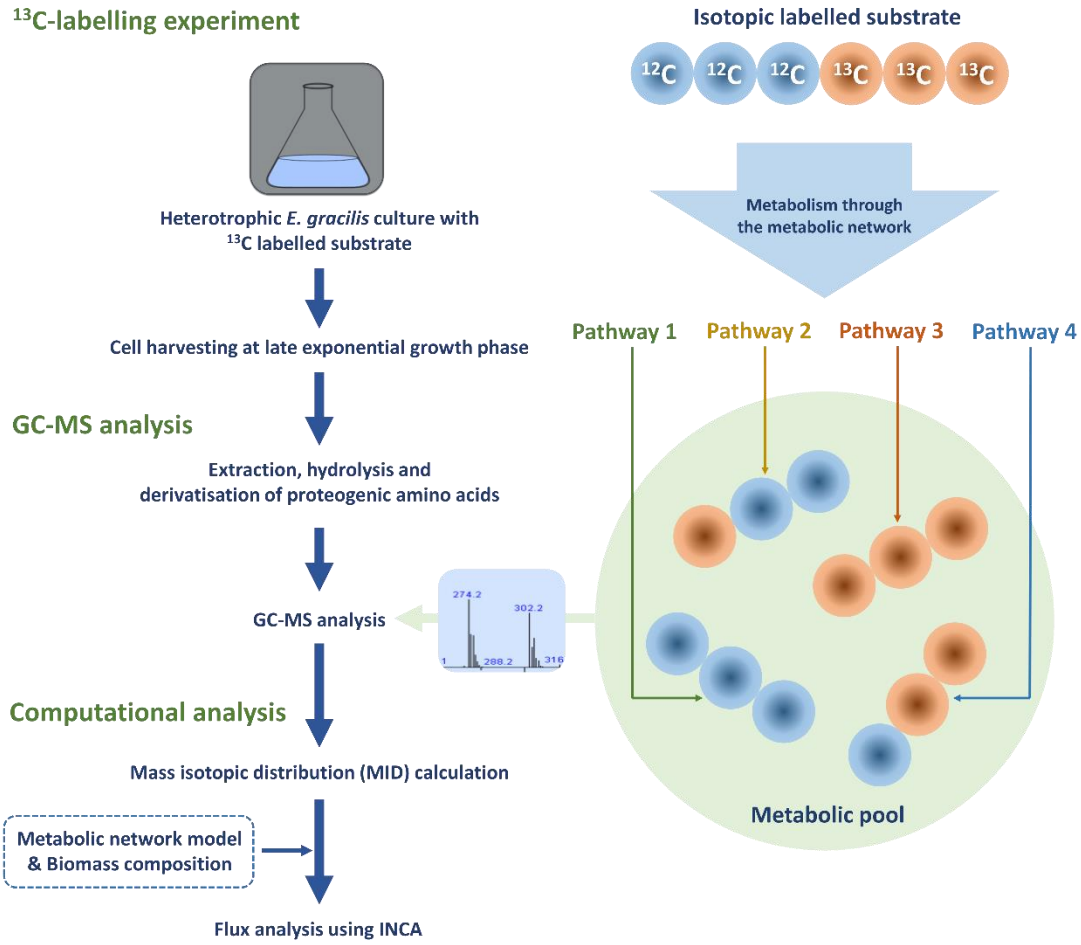


Figure 1.4. General workflow of ^{13}C MFA in *E. gracilis*. Labelling experiment is conducted in a batch heterotrophic culture of *E. gracilis*, and is harvested at pseudo metabolic/isotopic stationary state (late exponential growth phase). The rearrangement of ^{13}C atoms in metabolites of different pathways is determined by GC-MS. The analysed MIDs are used in conjunction with biomass composition data to perform ^{13}C MFA based on a model defining the metabolic network of *Euglena*.

The central network of carbon metabolism of *Escherichia coli* was one of the first systems to be studied using MFA, and the earliest studies employed [$^{13}C_6$]glucose as the labelled substrate. The ^{13}C tracer was successfully used to quantify the metabolic fluxes of *E.*

coli under aerobic and anaerobic conditions. The study demonstrated the use of isotopomer data to measure the intracellular reaction rates of the metabolic network of *E. coli* (Schmidt et al., 1999). Critically, it has been demonstrated that there is no significant effect on the growth of *E. coli* when grown on [$^{13}C_6$]glucose compared to the unlabelled glucose (Sandberg et al., 2016). This implies that fluxes measured in the ^{13}C MFA studies reflect those occurring in the equivalent unlabelled cultures. ^{13}C MFA has been successfully used to explore the response of the metabolic network to various conditions, such as glucose and xylose metabolism in *E. coli* in both aerobic and anaerobic conditions. Flux maps from different conditions showed that β-oxidation was important for anaerobic growth using xylose as the carbon source (Gonzalez et al., 2017).

In plants, ^{13}C MFA is an effective tool to measure the fluxes through known pathways, to address unknown pathways or to define fluxes of parallel pathways from different compartments (Iyer et al., 2008). However, performing ^{13}C MFA in plant cells is more complicated than in bacteria due to subcellular compartmentation in plant cells increasing the complexity of the network. It is necessary to revise the underlying metabolic model to accommodate duplication of pathways and the occurrence of specific pathways in the cytosol, mitochondria and plastids (Kruger and Ratcliffe, 2015).

It has been shown that using the isotopically labelled sugars, including [1-^{13}C]glucose, [$^{13}C_6$]glucose or the mixture of [$^{13}C_6$]glucose and unlabelled glucose, as the carbon sources do not perturb the metabolism of *A. thaliana* cell cultures (Kruger et al., 2007). ^{13}C MFA was used to show that changing the temperature changed the flux directions of cytosolic phosphoenolpyruvate (PEP) carboxylase (PEPC) and plastidic malic enzyme in soybean

cotyledons, causing changes in protein and lipid production. This indicated that these two enzymes could be targeted using metabolic engineering to improve protein and lipid production in soybean (Iyer et al., 2008).

To address the flux of carbohydrate oxidation in the cytosol and plastid of heterotrophic *A. thaliana*, ^{13}C MFA was performed, focusing on the pentose phosphate pathway (PPP). Three PPP models with different localisations of oxidative (oxPPP) and nonoxidative (nonoxPPP) phases were tested. The study found that flux data could be fitted with all three models, but biochemical evidences indicate the best match to the model with full PPP duplication. This highlighted the importance of using all available biochemical evidence and suggested that accurately predefining reaction locations is essential to obtain good quality flux maps (Masakapalli et al., 2010). Subsequently, ^{13}C MFA was used to investigate the metabolic impacts of different nitrogen supplies in heterotrophic *A. thaliana*. Analysis of the fluxes of *A. thaliana* grown in a medium with only nitrate and one containing a mixture of ammonium and nitrate as the nitrogen supply demonstrated that heterotrophic *A. thaliana* grown in the media with ammonium had a higher cell maintenance cost. Furthermore, by measuring the reductant demand for biosynthesis and the provision of reductant, it was found that flux through the oxidative pentose pathway was influenced by the reductant demand, showing the complexity of the metabolic response of the *Arabidopsis* cell. Moreover, the study showed that non-plastidic processes had a significant role in reductant provision for the plastid (Masakapalli et al., 2013).

The metabolic flux studies on well-established model organisms like *E. coli* and *A. thaliana* have demonstrated the value of using ^{13}C MFA to investigate metabolic networks.

Performing ^{13}C MFA on non-model organism, like *Euglena*, allows us to investigate missing biochemical pathways (McAtee et al., 2015) and to gain more insight into the metabolic capacity of the organism of interest.

1.3 Objectives

Despite a long history into the research of *Euglena* and the industrial application, the metabolic network of *Euglena* has not been systematically analysed. The aim of this thesis is to explore the metabolic network of *E. gracilis* using the metabolic quantification approaches, ^{13}C MFA and FBA. Within this overall aim, several research questions will be addressed. First, how is the metabolic network of *Euglena* configured and how do its component reactions operate? Secondly, what is the role of the chloroplast in heterotrophic carbon metabolism in *Euglena*? Thirdly, how does the metabolic network of *Euglena* allow the use of diverse organic substrates? In order to answer these questions, this research will:

(i) Extend the application of ^{13}C MFA and FBA to *Euglena* (Chapter 3)

(ii) Explore the impact of the absence of chloroplasts on heterotrophic metabolism in *Euglena* (Chapter 4)

(iii) Determine metabolic phenotype of *Euglena* during growth on a range of respiratory substrates differing in carbon composition and chemical redox potential (Chapter 5)

Chapter 2 | Materials and methods

2.1 Materials

Unless otherwise stated, the basic chemicals were purchased from Fisher Scientific (Loughborough, UK) or Sigma-Aldrich Company Ltd (St. Louis, USA). ^{13}C substrates for labelling experiments were purchased from various providers. [1-^{13}C]glucose and [2-^{13}C]glucose were purchased from Omicron Biochemicals Inc. (Indiana, USA). [$^{13}C_2$]acetate, [1-^{13}C]acetate, [2-^{13}C]acetate, [$^{13}C_3$]pyruvate, [1-^{13}C]pyruvate, [3-^{13}C]pyruvate and [$^{13}C_2$]ethanol were purchased from Sigma-Aldrich Company Ltd (St. Louis, USA). [$^{13}C_6$]glucose, [1-^{13}C]ethanol and [2-^{13}C]ethanol were purchased from CK Isotopes Ltd (Desford, UK). The radioisotopes [U-^{14}C]glucose (specific activity 9.25 MBq, 9.25-13.3 GBq mmol^{-1}) and [U-^{14}C]acetate (specific activity 1.85 MBq, 1.66 – 2.22 Gbq mmol^{-1}) were purchased from PerkinElmer (Massachusetts, USA).

2.2 *E. gracilis* cultivation

Axenic *E. gracilis* Klebs (strain 1224/5Z) culture was provided by Dr Ellis O'Neill, the original culture was obtained from the Culture Collection of Algae and Protozoa. The culture was maintained in a plastic cell culture flask (canted neck, surface area 25 cm^2, capacity 50 ml; tissue culture flask T-25 with filter cap, SARSTEDT AG & Co. KG) wrapped with aluminium foil containing 30 ml of AF6 medium (Kato, 1982) supplemented with 5.0 g dm^{-3} glucose at 21 °C on an orbital shaker at 120 rpm. The culture was sub-cultured every three weeks.

To prepare starting *E. gracilis* culture for different carbon substrates, 300 µl of maintaining culture at stationary phase was inoculated to plastic culture flasks containing 30 ml of AF6 medium supplemented with 5.00 g dm^{-3} glucose, 5.00 g dm^{-3} sodium pyruvate, 1.64 g dm^{-3} sodium acetate or 7.88 g dm^{-3} ethanol. The culture flask was wrapped with aluminium foil to exclude light and kept in a growth facility at a constant temperature of 21 °C with shaking at 120 rpm. Cell density was monitored daily with a hemocytometer until the culture reached stationary phase.

To establish the optimal concentration of each substrate, the starting culture at late exponential growth phase was inoculated into 30 ml AF6 medium supplemented with different concentrations of the carbon source: 5.00, 10.00, 15.00, 20.00, and 30.00 g dm^{-3} glucose, 0.10, 0.50, 1.00, 2.50 and 5.00 g dm^{-3} sodium pyruvate, 0.1, 0.50, 1.00, 1.64, 2.50 and 5.00 g dm^{-3} sodium acetate and 0.08, 0.14, 0.79, 7.88 and 15.78 g dm^{-3} ethanol. In addition, starting culture was inoculated into pure AF6 medium as a control. Starter culture volume of 100-300 µl, depending on cell density, was used to provide an initial cell density of approximately 10,000 cells ml^{-1}. *E. gracilis* culture was grown in four replicates for each concentration under the standard growth conditions as defined above. Cell density of each culture was monitored daily using a haemocytometer for two weeks or until cell growth reached stationary phase. Cells inoculation and sampling were conducted in a dark tissue culture cabinet to minimise the exposure to light.

2.3 Generation of aplastidic *E. gracilis*

To generate aplastidic *E. gracilis* strain, 300 µl of *E. gracilis* culture grown on glucose at late exponential growth phase was inoculated into 30 ml of AF6 medium containing 15 g

dm^{-3} glucose and 500 µg ml^{-1} streptomycin (SM). The culture was grown in the dark under the standard growth conditions and sub-cultured three times in media containing SM, before maintaining in AF6 (15 g dm^{-3} glucose) without SM.

2.4 Sample preparation for transmission electron microscopy (TEM)

This analysis was conducted in collaboration with Prof. Hugh Dickinson, Department of Plant Sciences, University of Oxford. 10 ml of late exponential (~ 5 days old) wild-type *E. gracilis* culture grown under a 16 h light/8 h dark photoperiod of about 40 µmol m^{-1} s^{-1} white light, wild-type *E. gracilis* grown in the dark, or aplastidic *E. gracilis* grown in the dark on glucose-supplemented medium was harvested by centrifugation at 4,000 g for 5 min. After decanting the supernatant, the cell pellet was washed twice, each time with 1 ml of AF6 medium. One ml of fixative reagent (aqueous (aq) solution of 0.05 M phosphate buffer, 2.5 % (w/v) glutardialdehyde and 2-3% (v/v) formaldehyde) was added to the sample, mixed well and incubated at 25 °C for 4 h. Thereafter, fixative was removed by centrifugation at 4,000 g for 5 min and the fixed cell pellet was washed twice with 1 ml 0.05 M phosphate buffer. In a fume hood, 1 ml of 2% (w/v, aq) OsO_4 was added to the sample before incubating for 1-2 h. Then, the OsO_4 solution was carefully discarded from the sample, and the pellet was washed twice by soaking in 1 ml water for 10 min before removing the water.

To embed the sample for sectioning, 1 ml of molten warm 2 % (w/v, aq) agar was added to the rinsed cell pellet and quickly mixed using a cocktail stick. Immediately afterwards, the cell-agar mixture was poured onto a glass slide and left to set for 15 min. To

ensure dehydration, the hardened agar sample was cut into small pieces and placed in glass vial for dehydration by successive washes with 25, 50, 75, 95 and finally 100 % (v/v, aq) acetone solution. The sample was fully submerged for 30 min before transfer to the next acetone concentration, and finally washed twice with 100 % acetone. Subsequently, the sample was washed with propylene oxide solution in acetone in a series of 25, 50, 75, 95 and 100 % (v/v). Thereafter, 25, 50 and 75 % (v/v) freshly prepared resin (TAAB 812 Resin Premix Kit – Hard Plus T027/1) in propylene oxide was gradually introduced into the sample by placing the sample in the resin mixture and incubating on a rotator for 4 h for each resin concentration before transferring to the higher concentration. Thereafter, the sample was placed into 2.5 ml of 100 % resin and incubated on a rotator overnight before sample-resin mixture was poured into a silicone mould. The sample was then baked at 60 °C in an oven in a flow hood for 24 h before sectioning using microtome and examining with FEI Tecnai T12 TEM.

2.5 Substrate utilisation analysis

To analyse substrate utilisation by *Euglena*, a 200 µl aliquot of wild-type *E. gracilis* culture grown heterotrophically on AF6 medium supplemented with 1.00 g dm^{-3} sodium pyruvate, 1.64 g dm^{-3} sodium acetate or 0.79 g dm^{-3} ethanol (optimal concentrations) was taken and placed in a 1.5 ml tube every other day starting from day 0 to day 14 of cultivation. *E. gracilis* cells were immediately removed by centrifuging at 5,000 g for 5 min and the media was kept at -80 °C for substrate quantification. Due to the relatively small changes in the amount of glucose in the medium, the glucose uptake rate of *E. gracilis* strains (wild-type and

aplastidic) grown on AF6 medium with 15 g dm^{-3} glucose was analysed by a ^{14}C radiolabelling technique. The measurement of each substrate was done with four biological replicates.

The substrate consumption rates (q) for acetate, pyruvate and ethanol cultures were determined by the average method as the equation below (Görgens et al., 2005), where s represents substrate concertation, x represents cell density and t represents time.

$$q = \frac{\Delta s/\Delta x}{\Delta t}$$

2.5.1 ^{14}C Glucose uptake assay

490 µl of *E. gracilis* culture grown on AF6 medium with 15 g dm^{-3} glucose at early exponential phase (day 3) was transferred into a 2 ml tube wrapped with aluminium foil. Then, 10 µl of 7.4 kBq μl^{-1} [U-^{14}C]glucose (total 74 kBq) was added to each tube to initiate the assay. A 100 µl aliquot was taken at 0, 1, 2 and 3 h, immediately filtered through a GF/A filter disk (Whatman®) and washed with 5 ml of AF6 medium. The filter with the collected *E. gracilis* cells was placed in a scintillation vial containing 350 µl Optiphase Hisafe 3 scintillation fluid (PerkinElmer), incubated in the dark overnight to allow the decay of chemilumiescence and then the amount of radioactivity was measured using a Wallac 1450 Trilux Microbeta scintillation counter (PerkinElmer). The rate of glucose uptake was calculated from the plot between radioactivity and incubation time.

2.5.2 Pyruvate quantification

All derivatisation processes were conducted in a fume hood. Frozen pyruvate medium was left to thaw at room temperature for 30 min. A 25 µl sample was placed in 1.5 ml tube

with the addition of 10 µl of 65.7 mM ribitol in pyridine as an internal standard and dried in a vacuum centrifuge for 2 h 25 µl of pyridine was then added and incubated on a thermomixer at 37 °C and 900 rpm for 30 min. After a short centrifugation to collect pyridine at base of the tube, 35 µl of N-methyl-N-(trimethylsilyl) trifluoroacetamide was added to derivatise the metabolites and the sample was incubated on a thermomixer at 37 °C and 900 rpm for 30 min. Subsequently, the sample was collected at the base of the tube by centrifugation for 1 min and transferred to a GC-MS vial (Borah, 2017).

The derivatised sample was analysed using an Agilent Intuvo 9000 GC system with Agilent 122-5512UI-INT: US17310301 column (15 m x 250 µm x 0.25 µm). 1 µl of the sample was injected with splitless mode with the column gas flow of 1.2 ml of He per min. The oven was programmed to start at 70 °C for 5 min. The temperature was then increased to 330 °C at a rate of 7.5 °C min^{-1}. The total run time was 39.7 min with 7.5 min of solvent delay included. The GC-MS was controlled via Intuvo/Enhanced MassHunter software. Agilent Mass Hunter Quantitative Analysis (MS) software was used for data processing. The pyruvate peak was identified using m/z 74 and 73 ions as the key fragments, and m/z 307 and m/z 319 ions were used to identify ribitol as the internal standard. The sample was quantified based on a standard curve of 0.00, 0.10, 0.25, 0.50, 1.00 and 2.00 g dm^{-3} of sodium pyruvate in AF6 medium.

2.5.3 Acetate quantification

The method of acetate quantification (Tumanov et al., 2016) was modified for measuring the acetate concentration in *Euglena* growth media. All derivatisation processes were conducted in a fume hood. The frozen medium sample was left to thaw at room

temperature for 30 min, before a 25 µl sample was aliquoted to 1.5 ml tube followed by the addition of 20 µl of 1 mM sodium [$^2H_3$]acetate, 25 µl 1-propanol and 25 µl pyridine. After vortexing for 30 s, the sample was placed on ice for 5 min before adding 50 µl of 1 M sodium hydroxide, followed by 15 µl methyl chloroformate. Immediately afterwards, the sample was vortexed for 30 s before the lid was opened to release pressure and the vortexing was continued for another 30 s. The sample was then centrifuged at 10,000 g for 5 min and 100 µl of the upper phase was placed in GC-MS glass vial for GC-MS analysis.

The derivatised acetate sample was analysed with an Agilent Intuvo 9,000 GC system with Agilent 122-5512UI-INT: US17310301 column (15 m x 250 µm x 0.25 µm). The oven was programmed to start at 45 °C for 48 s and the temperature was then increased to 60 °C at a rate of 25 °C min^{-1}. Thereafter, the temperature was decreased back to 45 °C before an increase to 190 °C at 50 °C min^{-1}. The total run time was 4 min. The injection volume was 1 µl using splitless mode with the column gas flow of 1.0 ml of He per min. The GC-MS was controlled by Intuvo/Enhanced MassHunter software.

Agilent Mass Hunter Quantitative Analysis (MS) software was used for data processing. The acetate peak was identified using m/z 61 ions as the key fragment, and m/z 63 ions was used to identify $^2H_3$-acetate as the internal standard. Acetate in the medium sample was quantified based on the standard curve of 0.00, 0.10, 0.25, 0.50, 1.00 and 2.00 g dm^{-3} of sodium acetate in AF6 medium.

2.5.4 Ethanol quantification

A spectrophotometric approach to quantify ethanol (Kristoffersen et al., 2005) based on the activity of alcohol dehydrogenase (ADH) (Walker, 1992, Fibla and Gonzàlez-Duarte, 1993) was optimised for ethanol quantification. The enzymatic assay mixture was prepared by adding 1.25 ml of 150 mM sodium pyrophosphate pH 8.5, 25 µl of 10 mM NAD^+ (aq) and 25 µl media sample to a plastic cuvette. After mixing thoroughly, 5 µL of 3,000-unit ml^{-1} ADH was added to the cuvette, mixed well and incubated at 30 °C for 20 min. The absorbance of the sample was measured at 340 nm using an Evolution 260 Bio UV-Visible spectrophotometer (Thermo Scientific). The concentration of ethanol in the medium was calculated by comparing with the standard curve of 0.00, 0.39, 0.79, 3.94 and 7.88 g dm^{-3} ethanol which were processed using the same procedures.

2.6 Biomass composition analysis

E. gracilis biomass was obtained from 4 dm^3 culture grown on AF6 media supplemented with either 15 g dm^{-3} glucose, 1.64 g dm^{-3} sodium acetate, 1.0 g dm^{-3} sodium pyruvate or 0.79 g dm^{-3} ethanol as the respiratory substrate. The culture was grown in the dark at a constant temperature of 21 °C and 120 rpm shaking. After reaching late exponential growth phase, *E. gracilis* cells were harvested by centrifugation at 5,000 g for 30 min and stored at -80 °C for 2 days before freeze-drying until the weight of the sample was constant. Approximately 2-3.5 g DW of heterotrophic *E. gracilis* biomass was obtained from each 4 dm^3 of culture. The dried biomass was kept in a desiccator until analysis. The analysis was done with four replicates.

2.6.1 Protein

The protein content of *E. gracilis* biomass was quantified using a Biorad DC protein assay kit (based on Lowry assay (Lowry et al., 1951)) according to the manufacturer's instruction, with bovine serine albumin (BSA) as protein standard. The BSA standards were freshly prepared at concentrations of 0.01, 0.05, 0.10, 0.20, 0.50 and 1.00 mg ml^{-1}. 10 mg sample of ground dried *Euglena* was resuspended in 10 ml of water to make 1 mg ml^{-1} *E. gracilis* biomass. 100 µl of resuspended cells, BSA standards and water (blank) were placed into separate test tubes before 500 µl of reagent A was added to each tube and mixed well. The tubes were then incubated for 10 min with constant shaking at 900 rpm at 60 °C before 4 ml of reagent B was added to each tube and vortexed immediately, followed by incubating at 25 °C for 15 min. Absorbance was measured at 750 nm using an Evolution 260 Bio UV-Visible spectrophotometer (Thermo Scientific). Subsequently, protein concentration was calculated by comparing absorbance of the sample to that of the protein standard.

2.6.2 Total carbohydrate

Heterotrophic *E. gracilis* biomass has an obvious pale orange colour, which is potentially from the carotenoids that are known to be produced in this organism. Due to the fact that carotene absorbs at the wavelength used in the phenol-sulfuric acid colorimetric assay (490 nm) (Laurens et al., 2012), 10 g of ground dried *Euglena* biomass was washed by vortexing with 5 ml acetone for 2 min to remove the carotenoids. The washed biomass was collected by centrifugation at 12,000 g for 5 min before acetone was discarded. The cell pellet was then resuspended in 10 ml of water to make 1 mg ml^{-1} biomass. Glucose standards were prepared in the concentrations of 0.01, 0.05, 0.10, 0.20, 0.50 and 1.00 mg ml^{-1}. In a fume

hood, 1 ml of the sample, standards and a water (blank) were aliquoted into separate test tubes, followed by the addition of 1 ml 5% (w/v, aq) phenol. Subsequently, 3 ml of 72% sulfuric acid was added to each tube before immediately vortexing for 5 s. The tubes were then incubated in a water bath at 90 °C for 5 min, and then placed in an ice bath for 2 min to stop the reaction, before measuring the absorbance at 490 nm using an Evolution 260 Bio UV-Visible spectrophotometer (Thermo Scientific). The quantity of total carbohydrate was calculated by comparing the absorbances of the sample to the glucose standards.

2.6.3 Insoluble carbohydrate

Insoluble carbohydrate quantification procedures were modified from previous reports (Laurens et al., 2012, Ogawa et al., 2015). 100 mg of ground dried *E. gracilis* biomass was placed in a weighed tube with 600 µl of chloroform/methanol (2:1, v/v) and vigorously vortexed for 2 min. Thereafter, 150 µl of 0.7% (w/v) NaCl (aq) was added to the sample. After vortexing for 2 min, the sample was centrifuged at 12,000 g for 5 min and the solvent was removed to collect the insoluble pellet. The extraction was repeated two more times for each sample. To remove the insoluble protein, 1 ml of 1% (w/v) sodium dodecyl sulphate was added to each tube and vortexed for 2 min before incubating on a thermomixer at 95 °C for 30 min. Thereafter, the sample was centrifuged at 12,000 g for 10 min before the supernatant was removed. Protein extraction was repeated two more times before the pellet was washed twice by vortexing with 1 ml 0.1% (w/v) sodium dodecyl sulphate for 5 min, followed by three washes with 1 ml water. The washed pellet was then dried in at 60 °C until the weight was constant. The weight of the tube was subtracted from the final weight to calculate the amount of insoluble carbohydrate in 100 mg biomass.

2.6.4 Total Lipid

Total lipid in *E. gracilis* biomass was quantified by the sulfo-phospho-vanillin (assay (Laurens et al., 2012, Mishra et al., 2014) using commercial sunflower seed oil as the standard. The phospho-vanillin reagent was freshly prepared by dissolving 150 mg vanillin in 2.5 ml ethanol, followed by adding 22.5 ml of water and 100 ml of phosphoric acid and mixing well. 10 mg of ground dried *E. gracilis* biomass was added to 10 ml of water and vortexed vigorously for 2 min to prepare 1 mg ml^{-1} biomass sample. Lipid standards were prepared in glass vials by dissolving of 0.01, 0.05, 0.10, 0.20, 0.50 and 1.00 mg sunflower oil in 100 ml chloroform. 100 µl of each concentration was transferred into separate glass vial and dried at 60 °C for 20 min. 100 µl biomass sample was aliquoted to a test tube and 100 µl of water was added into each dried standard tube and a clean tube (blank). 2 ml of 90% sulfuric acid was added to each tube before immediate vortexing for 10 s. The tubes were incubated in the 95 °C water bath for 10 min and then placed in an ice bath to cool down for 2 min. Subsequently, 5 ml of phospho-vanillin reagent was added to each tube, followed by incubation at 37 °C with 200 rpm shaking for 15 min before measuring absorbance at 530 nm using the Evolution 260 Bio UV-Visible spectrophotometer (Thermo Scientific). The lipid content of the biomass was calculated by comparing absorbance of the sample to the standard curve. The use of commercial sunflower as the standard was validated by comparing the result from the assay with a gravimetric assay using chloroform/methanol extraction of total lipids from 10 g of *E. gracilis* biomass.

2.7 Carbon conversion efficiency analysis

The carbon conversion efficiency (CCE) of *E. gracilis* was measured using a radiometric assay (Woods, 2014). 30 ml of wild-type *E. gracilis* culture grown on AF6 medium supplemented with either 15 g dm^{-3} glucose or 1.6 g dm^{-3} sodium acetate, and an aplastidic *E. gracilis* culture grown on the glucose medium, were prepared with an initial cell density of 10,000 cell ml^{-1}. The cultures were grown under the conditions described in 2.2. At exponential growth phase (day 4), 1 ml of *E. gracilis* culture was transferred to a sealed 100 ml flask. 9 ml of AF6 medium with supplemented substrate was added to the culture. 10 µl of 3.7 kBq μl^{-1} [U-^{14}C]glucose (total 37 kBq) was injected into the culture with glucose, and 10 µl of 3.7 kBq μl^{-1} [U-^{14}C]acetate (total 37 kBq) was injected to the acetate culture. After incubation for 24 h, the culture was filtered through a GF/A glass microfiber filter (Whatman®) to harvest *E. gracilis* cells, and the filter was washed with 20 ml of pure AF6 medium. The filter was stored at -80 °C for subsequent determination of the amount of radioactivity incorporated into biomass. CO_2 production measurement was done in a duplicated set of flasks with the plastic 1.5 ml tube containing 500 µl 10% (w/v) KOH suspended in the flask to trap CO_2 (Harrison and Kruger, 2008). To stop metabolism of the ^{14}C substrate and release dissolved CO_2, 2 ml of 12 M formic acid was injected into the culture before the flasks were further incubated for 8 h.

The filter containing *E. gracilis* cells and 450 µl KOH used for CO_2 trapping were placed in separate scintillation vials containing 2 ml Optiphase Hisafe 3 scintillation fluid (PerkinElmer) and left to incubate overnight to allow decay of chemiluminescence before quantification of the amount of radioactivity present with a Wallac 1450 Trilux Microbeta

scintillation counter (PerkinElmer). CCE was calculated using the equation below where C represents ^{14}C incorporation into CO_2 and B is ^{14}C incorporation in cell biomass.

$$CEE = 100 \times \frac{1}{(1 + \frac{C}{B})}$$

2.8 Steady state ^{13}C MFA

2.8.1 ^{13}C labelling experiment

E. gracilis culture was grown heterotrophically on AF6 medium supplemented with 15.0 g dm^{-3} glucose, 1.0 g dm^{-3} sodium pyruvate, 1.6 g dm^{-3} sodium acetate, or 0.79 g dm^{-3} ethanol as described in Section 2.2. Three different ^{13}C labelling strategies were used for each of the four substrates: [1-^{13}C]glucose, [2-^{13}C]glucose and 20% [$^{13}C_6$]glucose; [1-^{13}C]acetate, [2-^{13}C]acetate and 20% [$^{13}C_2$]acetate; [1-^{13}C]ethanol, [2-^{13}C]ethanol and 25% [$^{13}C_2$]ethanol; [1-^{13}C]pyruvate, [3-^{13}C]pyruvate and 20% [$^{13}C_3$]pyruvate.

2.8.2 GC-MS analysis of ^{13}C labelled sample

The ^{13}C labelled *Euglena gracilis* culture was harvested at late exponential phase of growth by centrifuging 30 ml culture at 6,000 g for 30 min and the supernatant was removed. Cell pellet was then washed with 1 ml of AF6 and centrifuged at 6,000 g for 15 min to remove the medium before being stored at -80 °C. 150 µl of 100% methanol was added to the frozen cell pellet and vortexed for 1 min to resuspend. The sample was then transferred to a 2 ml tube with screw cap and incubated on a thermomixer at 950 rpm for 15 min at 70 °C, followed by centrifugation at 10,000 g for 10 min. The supernatant was then removed and the

insoluble pellet was washed twice by adding 150 µl of 100% methanol and vortexing for 2 min before spinning down at 10,000 g for 10 min. After the supernatant was removed and the pellet was dried for 10 min, 500 µl of 6 M HCl was added to the sample to hydrolyse the methanol-insoluble material at 100 °C for 22 h. After cooling down and centrifuging the hydrolysed sample for 5 min at 10,000 g to settle cell debris, 25 µl hydrolysate was transferred to new tube and dried using a vacuum centrifuge for 2-4 h until no liquid remained.

All the derivatisation processes were conducted in a fume hood. 25 µl of pyridine was added to the dried sample and incubated at 37 °C on the thermomixer with 900 rpm shaking for 30 min. Subsequently, after a short centrifugation to pull down the evaporated pyridine, 35 µl of N-tert-butyldimethylsilyl-N-methyl trifluoroacetamide with 1% tert-butyldimethylchlorosilane was added and the sample was further incubated at 900 rpm for 30 min at 60 °C. The derivatised sample was transferred to a GC-MS vial and sealed for GC-MS analysis (Orata, 2012, Antoniewicz et al., 2007).

GC-MS analysis was performed using an Agilent Intuvo 9,000 GC system with an Agilent 122-5512UI-INT: US17310301 column (15 m x 250 µm x 0.25 µm). The derivatised sample was analysed using the injection volume of 1 µl with the column gas flow of 1.2 ml of He per min. The oven was programmed to start at 120 °C for 5 min. The temperature was then increased to 270 °C at a rate of 4 °C min^{-1}, before an increase to 310 °C for 3 min at a rate of 20 °C min^{-1}. The total run time was 47.5 min with 7.5 min of solvent delay included. The GC-MS was controlled using Intuvo/Enhanced MassHunter software.

2.8.3 MID extraction

To process the data, Agilent Mass Hunter Qualitative Analysis (MS) software was used. The chromatogram was extracted and integrated using the ChemStation integrator with Classical baseline correction mode. GC peaks of amino acids were identified based on the mass spectra of unlabelled samples using the National Institute of Standards and Technology (NIST) mass spectral database with a cut-off at 80% matching. To eliminate the effect of naturally occurring heavy isotopes from non-carbon atoms on the MID calculation, the m/z values and ion abundances of each identified amino acid peak were copied to a Notepad (.txt) file for mass correction using MSCorr software (Wahl et al., 2004). After mass correction, MS data of the following validated fragments of the detectable TBDMS derivatised amino acids were used for ^{13}C MFA (Antoniewicz et al., 2007, Masakapalli, 2011) (Appendix A).

2.8.4 ^{13}C Metabolic flux modelling

The central metabolic pathways of *Euglena*, including glycolysis, the PPP, the TCA cycle, anaplerotic reactions and amino acid biosynthesis were included in a central metabolic network model within INCA (Young, 2014). The model consists of elementary metabolite unit (EMU) balance equations based on the reactions in the central pathways, the subcellular location of each reaction (Inwongwan et al., 2019) and carbon atom transition information. In addition to metabolic pathway reactions, biomass synthesis equations derived from the biomass compositions of *E. gracilis* on each specific carbon substrate (demonstrated in Chapter 3) were included in the model to constrain the proportion of biomass components synthesised.

To perform ^{13}C MFA, the substrate input equations were adjusted according to the substrate used for each ^{13}C labelling experiment. The flux value for the substrate input equation was fixed as 1. The isotope information of the ^{13}C substrates and MID data of the chosen amino acid fragments obtaied from labelling experiments were incorporated into the model. Flux analysis was performed using INCA to minimise the weighted sum of squared residuals (SSR) and to determine the statistical acceptance comparing between simulated and experimental MIDs with chi-squared test at the confidence level of 95%. Parameter continuation, an inbuilt statistics algorithm in INCA, was used to test the reliability of the best-fit solution determined by the confidence intervals of each flux value. During model optimisation, an F-test we used to compare the fit of two closely related models (1, 2) with different degrees of freedom (DOF) (Antoniewicz et al., 2006, Boyle et al., 2017). The metabolic models for ^{13}C MFA can be accessed by the Oxford University Research Archive (https://ora.ox.ac.uk/).

$$F = \frac{\left(\frac{R1 - R2}{n1 - n2}\right)}{\left(\frac{R2}{p - n2}\right)}$$

The F value was calculated using the equation above, where R represents SSR, n represents the number of fitted variables and p is number of measurement data points (MIDs). The calculated F value was compared to a critical F value at 95% confidence level to test the null hypothesis. An F value greater than the critical value indicates that the more complex model provides a better fit to the experimental data.

2.9 Metabolic network model construction and FBA

2.9.1 Metabolic network model construction

The annotated data from two reports (O'Neill et al., 2015, Yoshida et al., 2016) were collated and used as material for drafting the metabolic network. After gathering and formatting the data to a set of EC (Enzyme Commission) numbers and KO (KEGG orthology), metabolic reactions were obtained from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg/pathway.html) by linking the EC numbers to reactions. The obtained reactions were arranged by their pathway functions in a Microsoft Excel file as an initial draft. The draft was manually curated by assigning the subcellular location, and the spontaneous and transport reactions were added according to a review of the literature (Inwongwan et al., 2019). The direction and stoichiometry of each reaction were then adjusted according to KEGG, the SIB Bioinformatics Resource Portal (ExPASy) and/or the Comprehensive Enzyme Information system (BRENDA) (Thiele and Palsson, 2010). Reactions in the model were classified into pathways such as glycolysis/gluconeogenesis, the PPP, the TCA cycle, and amino acid metabolism. The model was manually checked for presence and absence of enzymes involved in central metabolic pathways. The identified gaps (absent enzymes) were manually filled, and the reported cofactor specificities of the enzymes (Inwongwan et al., 2019) were taken into account. The electron transport chain was included and located in mitochondria, and protons produced from the process were specified to be used to generate ATP through ATP synthase with the ATP/H^+ stichometry of 1/4 (Shameer et al., 2019).

2.9.2 Model analysis

FBA was conducted using COBRApy (Ebrahim et al., 2013). The model was constrained by biomass synthesis reactions (demonstrated in Chapter 3). The objectives to optimise the biomass production (Cheung et al., 2013) and to minimise the sum of reaction fluxes through the network (Foguet et al., 2019) were used. The lower bounds and upper bounds of all fluxes were initially set to -1,000,000 and 1,000,000 respectively, but the flux of carbon substrate input was fixed to 100. In addition, the amounts of cofactors available for maintenance cost in the form of ATP and NADPH estimated from the MFA and FBA results were later introduced to the model. The FBA predicted fluxes were visualised and compared to those from the same reactions in MFA. CCE and biomass production from FBA were compared to those from MFA. Root mean square deviation (RMSD) (Pontius et al., 2008) was calculated based on the reactions of the free fluxes from ^{13}C MFA results to assess the match between FBA and MFA fluxes. Furthermore, the network model was used to assess the metabolic efficiency of *E. gracilis* grown on different carbon substrates in the analyses presented in Chapter 4 and 5. The metabolic model for FBA developed in this thesis can be accessed by the Oxford University Research Archive (https://ora.ox.ac.uk/).

Chapter 3 | Metabolic model of *E. gracilis*

3.1 Introduction

3.1.1 Overview of the metabolic pathways of *E. gracilis* from transcriptome analyses

Genomic analysis elucidates the DNA sequence of an organism of interest, which leads to the investigation of the organism in numerous aspects (Anthony J.F. Griffiths et al., 2015). Genome sequence is usually used in metabolic network analysis as the starting information for metabolic network reconstruction (Thiele and Palsson, 2010). The genome-scale network reconstruction is able to quantify the genotype–phenotype relationships (Fang et al., 2020).

The genome of *Euglena* was analysed and estimated to be approximately 500 Mb in size (Ebenezer et al., 2019), about 20% of the genome sequence has been successfully analysed and only became available after the analysis of this study had been undertaken. In the absence of a complete genome sequence for *Euglena*, transcriptome sequencing has been used as the preliminary alternative, aiming to gain an understanding of gene expression and regulation under different conditions. One of the first reports on transcriptomic analysis of heterotrophic and autophototrophic *E. gracilis* emphasised the versatile metabolic capacity of *Euglena* (O'Neill et al., 2015). In this analysis, all the core metabolic pathways were identified, including glycolysis, gluconeogenesis, the TCA cycle, the PPP, Calvin cycle, carotenoid biosynthesis, and thylakoid glycolipid biosynthesis. In addition, the pathways for the biosynthesis of lipids, amino acids and isoprenoids were also detected in the

transcriptome. The transcriptome showed the capacity for polyketide and non-ribosomal peptide biosynthesis, along with capacities for exploiting vitamin C, vitamin E, and glutathione to respond to stresses. Furthermore, the transcriptome provided evidence that *E. gracilis* produces multifunctional polydomain proteins related to those from both plant and bacteria, and which may have been obtained by horizontal gene transfer during evolution (O'Neill et al., 2015).

A comparative transcriptomic study of *E. gracilis* under aerobic and anaerobic conditions (Yoshida et al., 2016) investigated the regulatory system of wax ester metabolism under anaerobic conditions. The matching expressed genes from both conditions consisted of components of fundamental cellular processes including genetic information processing, translation, and energy metabolism. The differentially expressed genes from these conditions were not involved in wax-ester metabolism. This result indicates that the metabolic pathways for wax-ester biosynthesis were regulated at the post-transcriptional level. Furthermore, the transcriptomic data supported the view that the genome of *Euglena* is a hybrid of genomes from photosynthetic and heterotrophic organisms, which provided *Euglena* with the extensive metabolic capacities to adapt to diverse growth conditions. In combination, these two transcriptome studies demonstrate that the physiology of *E. gracilis* highly involves the regulation at the post-transcriptional level. Thus, using only the transcriptomic approach to explore the metabolic capacity of *Euglena* might be insufficient to understand how the metabolic network of *Euglena* responds to various conditions.

3.1.2 Functions of the *Euglena* plastid in heterotrophic conditions

The evolution of *Euglena* is complex due to the multiple secondary endosymbiosis events, which is the origin of its secondary chloroplast (Turmel et al., 2009). The acquisition of the plastids of *Euglena* associated with many gene losses and gene transfers from both primary and secondary endosymbionts to the host genome (Martin and Herrmann, 1998). As in the primary plastids of plants and green algae, the majority of the *Euglena* secondary plastid proteome is encoded in the nuclear genome (Armbrust et al., 2004, Douglas et al., 2001, Gleason and Hogenkamp, 1970). However, the secondary chloroplasts of *Euglena* are surrounded by three layers of membrane (Bachvaroff et al., 2005, Yoon et al., 2002). This implies that hundreds of chloroplast-targeted proteins synthesised in the cytosol need to be transported through three or four membranes to reach their destination in the stroma or the thylakoid lumen (Bolte et al., 2009).

Three classes of chloroplast pre-protein have been identified based on targeted signal analysis of the *E. gracilis* chloroplast proteome. Class I proteins are composed of a signal peptide (SP), stop-transfer signal and transit peptide (TP), whilst class II proteins contain only SP and TP (Záhonová et al., 2018, Durnford and Gray, 2006). No signal sequence is detected in class III chloroplast proteins and the transport mechanism of this class remains unknown (Novák Vanclová et al., 2019). The transport mechanism of class I and II pre-proteins into the chloroplast of *Euglena* (Figure 3.1) starts with co-translational transport into the endoplasmic reticulum (ER) lumen, where the signal peptide is cleaved. The pre-proteins are then transported to the Golgi body and from there to the outermost plastid membrane via vesicles. Golgi SNAP receptor (GOSR) complex and RAB5 GTPase are proposed to mediate the fusion

of the vesicle to the outermost membrane (Záhonová et al., 2018). The next steps of transporting across the inner two membranes of the plastids in euglenophytes remains unclear (Bolte et al., 2009, Záhonová et al., 2018, Novák Vanclová et al., 2019). The TOC/TIC-like unit was believed to be used due to the presence of a plant-like TP in the preproteins (Sláviková et al., 2005). However, TOC subunits have not been detected in the transcriptome of *E. gracilis*, even though homologues of several TIC subunits were detected (Ebenezer et al., 2019). Two proteins (seqids 13308 and 11030) of a novel subgroup of rhomboid-related pseudoproteases have been detected in the *Euglena* proteome. These proteins function in the symbiont-derived ERAD-like machinery (SELMA) transport system, which was able to recognise TP in chromalveolates (Maier et al., 2015, Novák Vanclová et al., 2019).

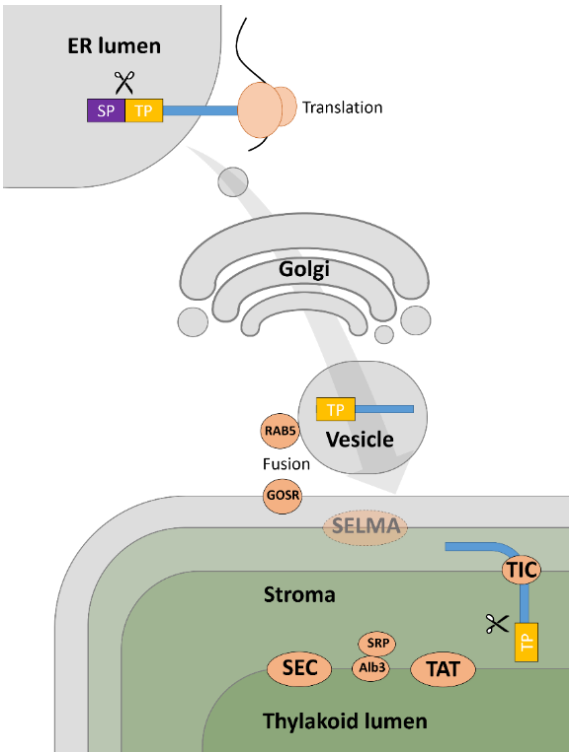


Figure 3.1. Protein transport into the secondary chloroplast of *Euglena* (Inwongwan et al., 2019). Nuclear encoded chloroplast pre-proteins (blue strip) are synthesised into the lumen

of the endoplasmic reticulum (ER) where the signal peptide (SP) is cleaved. Pre-proteins with transit peptides (TP) are subsequently transferred to the outermost chloroplast membrane through the Golgi body via vesicles. GOSR and RAB5 GTPase are proposed to mediate the fusion of the vesicle to the outermost membrane. After transport of proteins into the stroma, where the TP is removed, the mature protein can enter the thylakoid lumen via the SEC, TAT, or Alb3/SRP pathways (Záhonová et al., 2018).

3.1.3 Metabolic network reconstruction of *Euglena*

A metabolic network model, representing all the biochemical reactions in a cell, is a prerequisite for several approaches used in systems biology and in order to explore the metabolic phenotypes of various groups of organism, including bacteria, fungi, algae, animals and plants (Thiele and Palsson, 2010). In general, metabolic network reconstruction consists of three essential steps. First, the network is drafted using the genome data. This step involves the annotation of the genome sequences and identifying the metabolic reactions. Secondly, the draft model is curated according to available databases and experimental data. Thirdly, the curated network is converted into a computable format, and the network model is evaluated as to whether the network is functional for FBA (Thiele and Palsson, 2010).

There are no reports of a reconstructed metabolic network of *E. gracilis*, partly due to the lack of a complete genome. Transcript sequences and topology gap filling were used to attempt to reconstruct the metabolic network of *E. mutabilis* (Prigent et al., 2017). The transcriptomic data used to reconstruct this network was obtained and partially analysed from one growth condition (Halter et al., 2015). The network model was reported to be incomplete as it could not simulate the growth of *E. mutabilis* in the dark (Prigent et al., 2017),

despite the capacity of this species to do so. In contrast to that of *E. mutabilis*, the analyses of transcriptomes of *E. gracilis* provided a relatively complete overview of the metabolic network (O'Neill et al., 2015, Yoshida et al., 2016). In this chapter, I attempted to use these published transcriptomes to reconstruct the metabolic network of *E. gracilis*.

3.1.4 Objectives

The objectives of this chapter are to reconstruct a definitive metabolic model of *E. gracilis* for investigating the metabolic phenotype of the cells under heterotrophic growth conditions and to deduce the subcellular distribution of the metabolic pathways of *E. gracilis*. Due to the unavailability of complete genome data, I assembled the network using the available transcriptomic data of *E. gracilis*. The model was curated manually by adding reactions necessary for biomass component synthesis, before being tested for FBA constrained by biomass synthesis reactions generated from the biomass composition. To provide these empirical constraints, I formulated biomass synthesis reactions which determine the stoichiometry of the precursors required to produce the biomass composition of *E. gracilis*. Furthermore, I undertook a systematic review of the subcellular localisation of *Euglena* enzymes and the predictions of targeting of proteins deduced from the transcriptomes to define the organisation of the pathways in heterotrophic *E. gracilis.*

3.2 Subcellular localisation of enzymes of the central metabolic pathways of *E. gracilis*

Identifying subcellular locations of the metabolic pathways is important for reconstructing the metabolic network, as the locations indicate the function of the pathways and the organisation of the cell (Scott et al., 2005). Most enzymes of the central metabolic pathways of *Euglena* have been biochemically characterised and the subcellular locations of some of these enzymes have been reported (Buetow, 1989). Nevertheless, a comprehensive overview of the organisation of the central metabolic pathways of *Euglena* has not been attempted. In this study, I collated the reported subcellular locations of the enzymes of the central metabolic pathways of *Euglena*, and explored the practicality of exploiting protein targeting prediction tools to identify the location of *E. gracilis* proteins. Whilst the main features of this analysis are presented in this section, the full review can be found in the previously published report (Inwongwan et al., 2019).

3.2.1 Prediction of subcellular localisation of *E. gracilis* proteins

Eukaryotic protein subcellular location prediction tools, including WoLF PSORT (Horton et al., 2007) and TargetP 1.1 (Emanuelsson et al., 2007), were tested for their accuracy in predicting the locations of *Euglena* proteins. Protein sequences of selected marker enzymes of *E. gracilis* deduced from the *E. gracilis* transcriptome (O'Neill et al., 2015) were analysed with these tools. These markers included proteins known to be located in the chloroplast, mitochondria, cytosol or directed through the secretory pathway. Both plant-based and nonplant-based prediction modes of TargetP 1.1 and WoLF PSORT provided plausible results in subcellular location prediction for mitochondrial enzymes, proteins without targeting signals and those targeted for secretion. On the other hand, TargetP

predicted most of the chloroplast marker proteins to be either mitochondrial or secreted. WoLF PSORT correctly predicted many chloroplast-targeted enzymes, but has limitations in predicting chloroplast integral membrane proteins. This analysis demonstrates that, to establish the subcellular localisation of the proteins in *Euglena*, relying on only the protein targeting prediction results is not enough and the results need to be carefully evaluated in conjunction with biochemical and proteomic analysis.

Protein localisation prediction tools, TargetP and WoLF PSORT, showed only limited accuracy in predicting the location of chloroplast targeted proteins in *Euglena.* TargetP predicts the protein localisation based on the presence of N-terminal presequences, including chloroplast transit peptides (cTP) for predicting chloroplast protein import (Emanuelsson et al., 2007, Emanuelsson et al., 2000), and WoLF PSORT prediction is based on sorting signals, amino acid composition and functional motifs (Horton et al., 2007).

However, the structure of cTP or chloroplast sorting signal is not well characterised in *Euglena*, which might lead to the inaccurate prediction of chloroplast proteins. The limitation of predicting chloroplast-targeted proteins was previously reported in *A. thaliana* where the prediction performance of chloroplast-targeted proteins was less accurate than that for mitochondria, with occasional false predictions between mitochondrial targeting peptide (mTP) and cTP (Emanuelsson et al., 2007). This poor discrimination performance between mTP and cTP is also observed in our study with *E. gracilis* proteins. This lack of discrimination is potentially due to the use of the same targeting sequence for some chloroplast- and mitochondria-targeted proteins (Emanuelsson et al., 2007).

Furthermore, apart from the limited understanding of cTP in *Euglena*, chloroplast protein targeting of *Euglena* differs from that in plants due to the different origin and structures of the secondary chloroplast of *Euglena.* The secondary endosymbiosis led to the chloroplast being surrounded by three envelope membranes (Bachvaroff et al., 2005, Yoon et al., 2002, Gibbs, 1978), and protein transport into secondary chloroplasts of *E. gracilis* is not fully understood. Thus, the prediction of TargetP and WoLF PSORT might not cover the protein transport into the chloroplast of *Euglena*. This suggests that the prediction results, especially for chloroplast targeted proteins, need to be carefully assessed in conjunction with enzyme analysis.

3.2.2 Organisation of the central metabolic pathways in *Euglena*

By considering the enzymatic, proteomic and protein targeting prediction results, I generated a model of the distribution of the metabolic pathways for heterotrophic carbon metabolism in *Euglena* (Figure 3.2). Glycolysis, which catalyses the initial breakdown of sugars, is present in the cytosol and chloroplasts, as is commonly found in green plants and algae. The TCA cycle of *Euglena* is located in the mitochondria, as is the norm for eukaryotes. The enzymes of the glyoxylate cycle, commonly associated with microbodies in higher plants, are present in the mitochondria of *E. gracilis*. Unlike plants and green algae, in *Euglena,* the complete PPP is present in the cytosol with a duplicated non-oxidative phase present in the plastid.

The subcellular location of some amino acid biosynthesis pathways is ambiguous, though most of them appear to be distributed between the mitochondria and cytosol. A plant-like pathway for aromatic amino acid biosynthesis was detected in the plastids when

grown under autotrophic conditions (Reinbothe et al., 1994), alongside a cytosolic pathway, similar to that found in non-photosynthetic eukaryotes. Unlike plants, the α-aminoadipate pathway for lysine synthesis was detected in *E. gracilis* (Hasan et al., 2017), similar to that found in fungi (Torruella et al., 2009, Xu et al., 2006), although the subcellular location of the α-aminoadipate pathway has not been reported. Lipid can be synthesised in various subcellular locations, such as wax ester formed in mitochondria and photosynthetic glycolipids in the chloroplasts. Tetrapyrroles are synthesised in both the chloroplasts and mitochondria in *Euglena*. These indicate that, aside from the reactions of photosynthesis that are exclusively present in the chloroplast, all central metabolic pathways needed for survival of *Euglena* are found elsewhere in the cell.

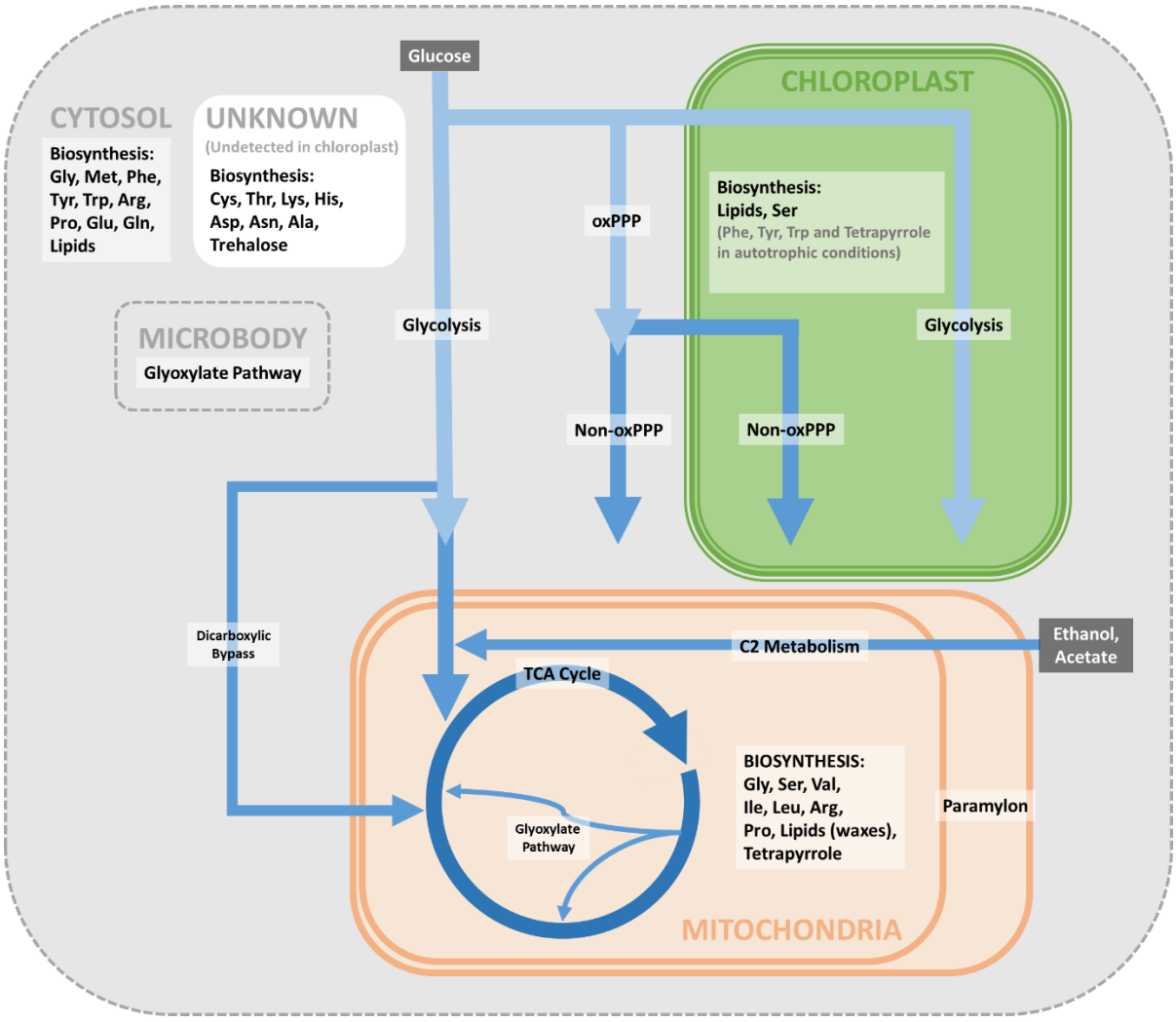


Figure 3.2. Proposed distribution of central metabolic pathways in *Euglena*. The considered subcellular locations consist of cytosol (grey), mitochondria (orange) and chloroplast (green). This model demonstrates that central metabolic pathways essential for survival of *Euglena* are located outside the chloroplasts. This figure was previously published (Inwongwan et al., 2019).

3.3 Construction of the metabolic network model of *E. gracilis*

The process of metabolic network reconstruction consists of three major stages: draft reconstruction, refinement of reconstruction and network evaluation (Thiele and Palsson, 2010). In this section, the processes of drafting and refinement of the network of *Euglena* will be described in detail.

3.3.1 Drafting the metabolic network of *E. gracilis*

Metabolic networks are commonly drafted based on genome annotation (Thiele and Palsson, 2010). Due to the unavailability of complete genome data of *Euglena,* to reconstruct the metabolic network of *E. gracilis*, the available annotated transcriptomes from two reports (Yoshida et al., 2016, O'Neill et al., 2015) were collated as a set of EC numbers and KOs. A total of 203 EC numbers and 552 KOs were obtained and cross-referenced to the KEGG database to acquire a set of 864 metabolic reactions. After removing the duplicated reactions, the draft reconstructed network contained 716 unique reactions (Figure 3.3). The draft reaction network was visualised and analysed by iPath 2.0 (Yamada et al., 2011) (Figure 3.4). The diagram demonstrates that the annotated transcriptomes provided reactions that covered all of the central pathways of carbon metabolism, including glycolysis, the PPP, the TCA cycle, glyoxylate shunt and anaplerotic reactions. The data also provided evidence for other major pathways, such as, amino acids biosynthesis, lipid metabolism and nucleic acid biosynthesis. This shows the potential of using the data set as a basis for reconstructing the metabolic network model of *Euglena*. However, as commonly found draft reconstructions (Thiele and Palsson, 2010, Prigent et al., 2017), some pathways were incomplete due to the

apparent absence of specific reactions (termed a “gap”), and the draft network does not contain any indication of the subcellular location of the pathways.

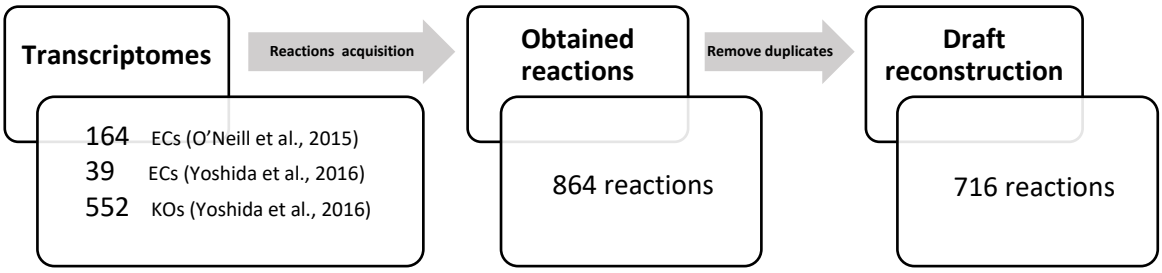


Figure 3.3. A summary of the metabolic network reconstruction of *E. gracilis* from the transcriptomic data. EC is enzyme commission number and KO is KEGG orthology obtained from the reported transcriptomes (Yoshida et al., 2016, O'Neill et al., 2015).

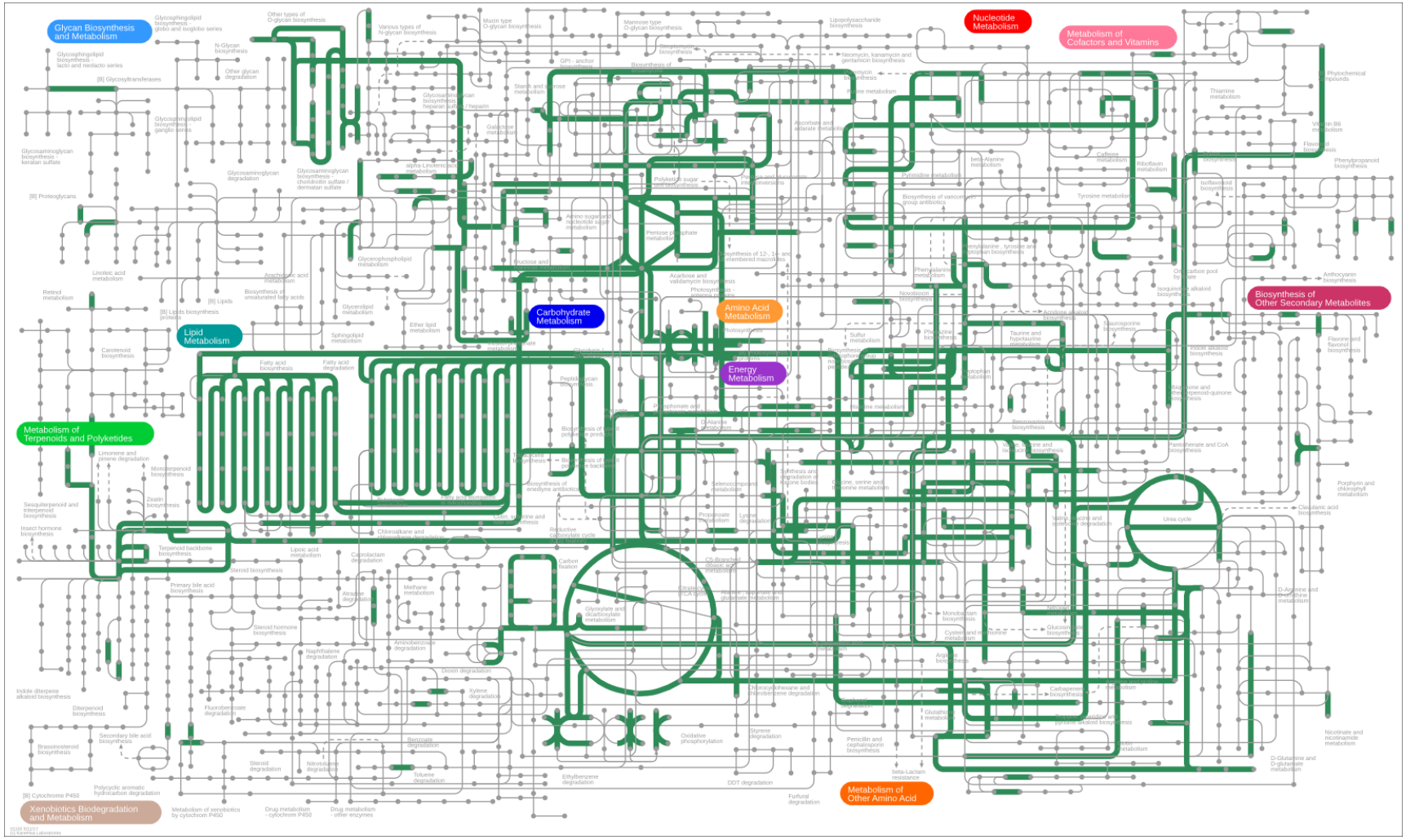


Figure 3.4. Draft of the metabolic network of *E. gracilis* visualised using iPath 2.0 (Yamada et al., 2011). The green highlighted lines demonstrate the reactions evidenced in the transcriptomes (Yoshida et al., 2016, O'Neill et al., 2015).

3.3.2 Model curation

Following construction of the draft metabolic network of *E. gracilis* described in Section 3.3.1, the model was further manually curated. First, the input reactions were added according to the nutrients in the AF6 medium, including the different carbon sources, ammonia (NH_3), sulfate (SO_4), and phosphate (Pi). Reactions allowing free exchange of O_2, H_2O, H^+ and CO_2 with the environment were also included in the model. Secondly, the possible reactions were categorised based on their function in the metabolic pathways. The draft model provides complete central metabolic pathways including glycolysis, PPP, the TCA cycle, glyoxylate shunt and the anaplerotic reactions. Gap-filling, a standard method to complete the incomplete pathways by adding the necessary missing reactions to the pathways (Prigent et al., 2017), was done manually in this study. Gaps in the pathways for the biosynthesis of several amino acids, purine and pyrimidine were identified in the drafted network. The missing reactions were added to the pathways according to the KEGG database to allow the amino acids and nucleic acids to be produced from the input medium. Furthermore, the electron transport chains and ATP synthase were added to allow oxidative phosphorylation.

Thirdly, stoichiometrically imbalanced reactions including the macromolecules which allow the infinite production or consumption of metabolites were checked and removed from the model to stop the production of metabolites requiring an infinite supply of precursors. These include the reactions that contain the macromolecules, cellulose and β-glucan. The reversibility of each reaction involved in the network was set according to information from KEGG, ExPASy and/or BRENDA databases. Subsequently, the reactions in the model were allocated to their subcellular locations according to the analysis presented in Section 3.2.2,

together with the addition of the necessary transport reactions between cytosol and mitochondria or the cytosol and chloroplast. As a result, the final model consisted of 831 reactions and 956 metabolites.

3.4 Formulating biomass synthesis reactions for metabolic flux analysis

Biomass synthesis reactions are essential for constraining the metabolic model as they determine the molar ratio of the precursors in the production of biomass components. The biomass synthesis reaction is generally expressed in the unit of mmol (g cell DW)$^{-1}$. To express the biomass composition in the unit of mmol (g cell DW)$^{-1}$, first, the mol fractional contribution of monomer or precursor to the biomass component (% mol) is multiplied by the anhydrous molecular weight (g mol^{-1}) of the monomer to estimate the contribution in grams of the monomer/precursor per mol of biomass component. The sum of gram monomer/precursors contribution to mol component of the monomers is the estimation of the molecular weight of the component (g mol^{-1}) on average. Dividing the mol contribution of monomer/precursor by the average molecular weight of component provides the mol contribution of monomer per gram of component (mol g^{-1}), which can be calculated into the mol contribution of monomer/precursor to cell dry weight (mol (g cell DW)$^{-1}$) by multiplying with the biomass composition of the component. An example of the calculation of alanine in the biomass of *E. gracilis* grown on glucose is provided below.

The calculation of mol contribution to biomass dry weight of alanine (ALA)

0.102 mol ALA (mol protein)$^{-1}$ × 71.1 g anhydrous ALA (mol ALA)$^{-1}$ = 7.22 g ALA (mol protein)$^{-1}$

Σ (g amino acids (mol protein)$^{-1}$) = 109 g protein (mol protein)$^{-1}$

0.102 mol ALA (mol protein)$^{-1}$ / 109 g protein (mol protein)$^{-1}$ = 0.931 mmol ALA (g protein)$^{-1}$

0.931 mmol ALA (g protein)$^{-1}$ × 0.169 g protein (g cell DW)$^{-1}$ = 0.157 mmol ALA (g cell DW)$^{-1}$

In this section, I illustrate the formulation of biomass synthesis reactions in the unit of mmol (g cell DW)$^{-1}$ using the biomass composition of wild-type *E. gracilis* grown heterotrophically on glucose (Figure 3.5). The biomass was categorised into four groups (protein, carbohydrate, lipid and nucleic acid) to facilitate the calculation. The carbohydrate content was divided into soluble and insoluble carbohydrate, and the nucleic acid is considered to be represented in the unidentified part of the biomass (others). The biomass synthesis reactions of each component were formulated as shown in the following sections.

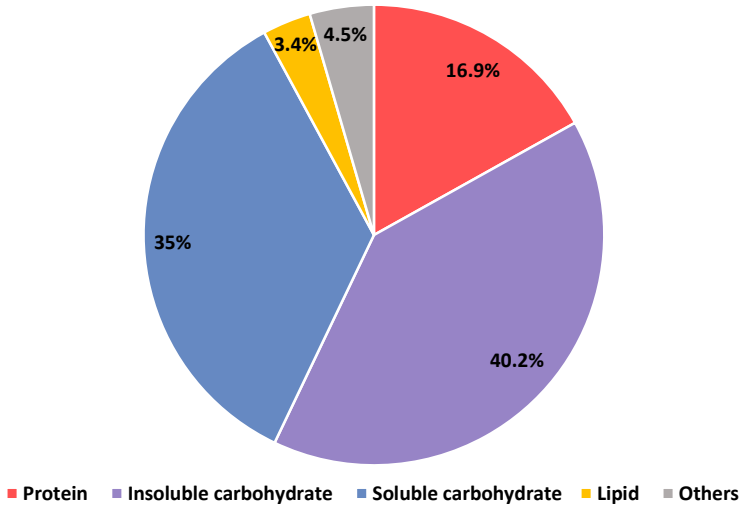


Figure 3.5. **Biomass composition of heterotrophic wild-type *E. gracilis* grown on glucose medium.** The biomass composition was measured from late exponential *E. gracilis* culture grown in the dark as mentioned in Chapter 2. The biomass components consist of protein (16.9 % DW), insoluble carbohydrate (40.2 % DW), soluble carbohydrate (35.0 % DW), lipid (3.4 % DW) and unidentified biomass (4.5 % DW).

3.4.1 Protein

The amino acid mol fractional abundance of *Euglena* protein was reported to be unchanged across diverse growth conditions (Miller and Kempner, 1976). The formulation of the amino acid contribution to biomass in the unit of mmol (g cell DW)$^{-1}$ is shown in Table 3.1. Considering the condensation of amino acids in peptide polymerisation, the anhydrous molecular weight (MW) of each amino acid was used in this calculation. A molecule of H_2O at the C-terminus of the peptide was not taken to account as it negligibly alters the final calculated value. Protein content of 0.169 g (g cell DW)$^{-1}$ was used in this calculation.

Table 3.1. Formulating amino acid contribution of biomass of heterotrophic wild-type *E. gracilis* grown on glucose. Amino acid mol fractional abundance of *Euglena* used in this calculation was obtained from a published data (Miller and Kempner, 1976).

Amino acid	Anhydrous MW ($g\ mol^{-1}$)	Mol fractional abundance	Amino acid content ($g\ mol^{-1}$)	Amino acid content in mmol $(g\ protein)^{-1}$	Amino acid contribution $(mmol\ (g\ cell\ DW)^{-1})$
Alanine (ALA)	71.1	0.1016	7.224	0.931	0.157
Arginine (ARG)	156.2	0.0576	8.997	0.528	0.089
Aspartate (ASP)	115.1	0.0398	4.581	0.365	0.062
Asparagine (ASN)	114.1	0.0398	4.541	0.365	0.062
Cysteine (CYS)	103.2	0.0040	0.4128	0.037	0.006
Glycine (GLY)	57.1	0.0820	4.682	0.752	0.127
Glutamate (GLU)	129.1	0.0592	7.636	0.542	0.092
Glutamine (GLN)	128.14	0.0592	7.579	0.542	0.092
Histidine (HIS)	137.2	0.0209	2.867	0.192	0.032
Isoleucine (ILE)	113.2	0.0433	4.902	0.397	0.067
Leucine (LEU)	113.2	0.0972	11.00	0.891	0.151
Lysine (LYS)	128.2	0.0653	8.371	0.599	0.101
Methionine (MET)	131.2	0.0112	1.469	0.103	0.017
Phenylalanine (PHE)	147.2	0.0354	5.211	0.325	0.055
Proline (PRO)	97.1	0.0684	6.642	0.627	0.106
Serine (SER)	87.1	0.0565	4.921	0.518	0.088
Threonine (THR)	101.1	0.0561	5.672	0.514	0.087
Tryptophan (TRP)	186.2	0.0048	0.8938	0.044	0.007
Tyrosine (TYR)	163.2	0.0279	4.553	0.256	0.043
Valine (VAL)	99.1	0.0699	6.92709	0.641	0.108

3.4.2 Carbohydrate

Insoluble carbohydrate

β-glucan in the form of paramylon, the insoluble storage carbohydrate of *Euglena*, has glucose as the monomer. The mol contribution of the precursor of β-glucan (glucose 6-

phosphate (G6P)) to the cell biomass was calculated similarly to the calculation of amino acid contribution in biomass. The insoluble carbohydrate composition is 0.402 g (g cell DW)$^{-1}$. The monomer factional composition of 1.0 and the anhydrous molecular weight of glucose were used in this calculation to makeup of the long chain beta-glucan.

Soluble carbohydrate

Mannitol and trehalose were reported as the main soluble carbohydrates in heterotrophic *E. gracilis* grown in the dark, and the ratio of trehalose to mannitol varied in different osmotic conditions (Dwyer, 1986). The molar ratio of trehalose to mannitol was estimated to be 0.17 : 0.83 (Dwyer, 1986) for use in this calculation. The soluble carbohydrate content of the biomass was 0.35 g (g cell DW)$^{-1}$. In this case, the molecular weights of mannitol and trehalose were directly used as these two are mono- and disaccharides, respectively. The precursor content of carbohydrate biomass in the unit of mmol (g cell DW)$^{-1}$ is shown in Table 3.2.

Table 3.2. Precursors contribution of carbohydrate biomass of wild-type *E. gracilis* grown on glucose.

Carbohydrate	Compound	Precursor	Mol fractional abundance	Precursor contribution (mmol (g cell DW)$^{-1}$)
Insoluble	Paramylon	G6P	1.0	2.479
Soluble	Mannitol	Fructose 6-Phosphate (F6P)	0.83	1.387
	Trehalose	2xG6P	0.17	0.284

3.4.3 Lipid

The lipid composition of *Euglena* was reported to vary widely across different conditions (Constantopoulos and Bloch, 1967). The lipid composition of *Euglena* used in this

calculation was estimated from several reports of lipid composition in *E. gracilis* grown under heterotrophic conditions (Table 3.3). Lipids of *Euglena* were categorised into four major groups, specifically free fatty acids, phospholipids, triglycerides and wax-esters. In heterotrophic conditions, neutral lipids constitute up to 50% of the total lipid weight (Rosenberg, 1963), which consists of about 14.5% triglycerides while the rest is mostly wax-esters. In growth medium that contained both carbon and nitrogen sources, the proportion of phospholipids were reported to constitute about 30% of total lipid weight (Regnault et al., 1995). The total lipid content in *Euglena* of 0.034 g (g cell DW)$^{-1}$ was separated by lipid composition to calculate the contribution of each lipid to cell dry weight.

Table 3.3. The estimated lipid composition of *E. gracilis* grown heterotrophically

Lipid	Lipid composition (%)	Lipid content (g (g DW)$^{-1}$)
Free fatty acid (Regnault et al., 1995)	20.0	0.0068
Phospholipid (Regnault et al., 1995)	30.0	0.0102
Triglyceride (Rosenberg, 1963)	14.5	0.0049
Wax-ester (Rosenberg, 1963, Wang et al., 2018)	35.5	0.0121

Free fatty acids

The fatty acid composition of heterotrophic *Euglena* cells was reported to be 9.0, 7.9, 46.5, 19.6 and 12.9 % wt for C12, C13, C14, C16 and C18 fatty acids, respectively (Rosenberg, 1963). The fatty acid composition was calculated as mol fractional composition of total fatty acid by multiplying the molecular weight of each compound with its composition, and then dividing by the average molecular weight of total fatty acid (sum of fatty acid content (g mol^{-}

^{1})). The calculation of each lipid was done similarly to that previously described, and precursors contribution of fatty acids were deduced (Table 3.4).

Table 3.4. Precursors contribution of fatty acids in biomass of wild-type *E. gracilis* grown on glucose

Free fatty acid	Precursors	Mol fractional abundance	Precursor contribution (mmol (g cell DW)$^{-1}$)
C12	6x acetyl coenzyme A (AcCoA)	0.110	0.0180
C13	5xAcCoA, propionyl CoA	0.091	0.0015, 0.003
C14	7xAcCoA	0.500	0.0980
C16	8xAcCoA	0.188	0.0400
C18	9xAcCoA	0.111	0.0270

Phospholipids

With the same fatty acid composition as described in the previous section, the precursor contributions to cell biomass of each phospholipid were calculated in the same way as for free fatty acids. The number of precursors were considered for phospholipids consisting of two fatty acid chains and one headgroup. The reported fractional composition of the headgroup of phospholipid in *Euglena* (Shibata et al., 2018) were also accounted for and the precursor contribution in biomass of phospholipids with the headgroups are reported in Table 3.5.

Table 3.5. Precursor contribution of phospholipids in biomass of wild-type *E. gracilis* grown on glucose

Phospholipid	Precursors	Mol fractional abundance	Precursor contribution (mmol (g cell DW)$^{-1}$)
C12	Glyceraldehyde 3-phosphate (Glyc3P), 12xAcCoA	0.106	0.002, 0.024

C13	Glyc3P, 10xAcCoA, 2xPpnCoA	0.089	0.001, 0.01, 0.002
C14	Glyc3P, 14xAcCoA	0.497	0.008, 0.112
C16	Glyc3P, 16xAcCoA	0.192	0.003, 0.048
C18	Glyc3P, 18xAcCoA	0.116	0.002, 0.036
Headgroup			
Glycolipid	G6P	0.234	0.0037
Phosphatidylglycerol	Glyc3P	0.015	0.0002
Phosphatidylcholine & phosphatidylethanolamine	3PGA	0.751	0.0120

Triglycerides

With the same fatty acid composition, the precursor contributions in cell biomass of triglycerides were calculated in the same way as other lipids, but taking into account that one molecule of triglyceride consists of three fatty acid chains and one glycerol. The results are reported in Table 3.6.

Table 3.6. Precursors contribution of triglycerides in biomass of wild-type *E. gracilis* grown on glucose

Triglyceride	Precursors	Mol fractional abundance	Precursor contribution (mmol (g cell DW)$^{-1}$)
C12	Glyc3P, 18xAcCoA	0.1094	0.001, 0.018
C13	Glyc3P, 15xAcCoA, 3xPpnCoA	0.0901	0.001, 0.015, 0.003
C14	Glyc3P, 21xAcCoA	0.4995	0.003, 0.063
C16	Glyc3P, 24xAcCoA	0.1886	0.001, 0.024
C18	Glyc3P, 27xAcCoA	0.1124	0.001, 0.027

Wax-ester

The vast majority of the wax-ester was reported to be esters of C14 and C16 fatty acids and fatty alcohols (Wang et al., 2018). Considering this, the precursor contribution of wax-ester in cell biomass was calculated based on C14 and C16 hydrocarbon chains. The result is shown in Table 3.7.

Table 3.7. Precursors contribution of wax-ester in biomass of wild-type *E. gracilis* grown on glucose

Wax-ester	Precursor	Mol fractional abundance	Precursor contribution (mmol (g cell DW)$^{-1}$)
14:0 – C16:0 or 16:0 – C14:0	15*AcCoA	1	0.405

3.4.4 Nucleic acid

The nucleic acid content of *Euglena* was low compared to other cell components and assumed to be present in the unidentified component of the biomass of *E. gracilis*. The nucleic acid content of *Euglena* was estimated from a previous study (Kempner and Miller, 1965a) to be approximately 0.005 g (g cell DW)$^{-1}$ for DNA and 0.031 g (g cell DW)$^{-1}$ for RNA. The nucleotide fractional composition of DNA and RNA were obtained from a previous study (Brawerman et al., 1962). The nucleotide contribution is shown in Table 3.8 for DNA and Table 3.9 for RNA.

Table 3.8. Precursors contribution of DNA in biomass of wild-type *E. gracilis* grown on glucose

Nucleotide	Precursor	Mol fractional abundance	Precursor contribution (mmol (g cell DW)$^{-1}$)
dAMP	Inosine monophosphate (IMP)	0.245	0.004
dGMP	IMP	0.248	0.004
dCMP	Uridine monophosphate (UMP)	0.26	0.0042
dTMP	Thymidine monophosphate (TMP)	0.247	0.004

Table 3.9. Precursors contribution of RNA in biomass of wild-type *E. gracilis* grown on glucose

Nucleotide	Precursor	Mol fractional abundance	Precursor contribution (mmol (g cell DW)$^{-1}$)
AMP	IMP	0.19	0.0182
GMP	IMP	0.322	0.0309
CMP	UMP	0.292	0.028
UMP	UMP	0.196	0.0188

3.4.5 The biomass synthesis reactions of *E. gracilis* grown on glucose

From the contributions of each precursor for each biomass component demonstrated previously, the biomass synthesis reactions of wild-type *E. gracilis* grown on glucose are shown below, the coefficient relates to the relative contribution of each precursor to the biomass in the unit of mmol (g cell DW)$^{-1}$.

***Protein biomass*:** 0.157*ALA + 0.089*ARG + 0.062*ASN + 0.062*ASP + 0.006*CYS + 0.127*GLY + 0.092*GLU + 0.092*GLN + 0.032*HIS + 0.067*ILE + 0.151*LEU + 0.101*LYS + 0.017*MET + 0.055*PHE + 0.106*PRO + 0.088*SER + 0.087*THR + 0.007*TRP + 0.043*TYR + 0.108*VAL --> biomass_P

***Carbohydrate biomass*:** 3.047*G6P + 1.387*F6P --> biomass_C

***Lipid biomass*:** 0.98*AcCoA + 0.02324*Glyc3P + 0.008*PpnCoA + 0.004*G6P + 0.012*3PGA --> biomass_L

***Nucleic acids biomass*:** 0.0571*IMP + 0.051*UMP + 0.004*TMP --> biomass_NA

***Total biomass*:** biomass_P + biomass_C + biomass_L + biomass_NA --> Biomass

3.5 Testing the stoichiometric model of the metabolic network in *E. gracilis*.

The metabolic network model of *E. gracilis* was validated by applying FBA to heterotrophic carbon metabolism in *Euglena*. The ability to utilise a range of organic carbon sources as the growth substrate for *E. gracilis* was tested by constraining the model with biomass synthesis reactions for *E. gracilis* grown heterotrophically on glucose, acetate, ethanol or pyruvate (Appendix B). FBA was conducted using COBRApy (Ebrahim et al., 2013) with a fixed influx of the carbon substrates at 100 mol per unit time. Objectives of maximising the biomass production and minimising total fluxes through the network were used in the analysis. The summary of the FBA results (Table 3.10) shows that the biomass components were successfully produced from the supplied medium and substrates, resulting in predictions of flux for biomass production (objective function). For all substrates, the model predicted that, to produce the biomass, carbon substrate, O_2, Pi, NH_3 and SO_4 were consumed from the environment (in fluxes). Protons, water (H_2O) and CO_2 were predicted to be released from the cells (out fluxes) during the metabolism of the carbon substrate. The results demonstrate success in predicting fluxes of *E. gracilis* grown on different substrates in heterotrophic conditions based on the biomass constraint. This indicates the potential use of the model as the basis for FBA in *E. gracilis*.

Table 3.10. The summary of FBA results of *E. gracilis* grown on different carbon substrates. The data represent mol of fluxes per 100 mol of carbon substrate intake.

In fluxes		Out fluxes		Objective
Metabolites	Fluxes	Metabolites	Fluxes	flux
Glucose	100.00	H^+	1.38E+03	13.01
Pi	62.37	H_2O	321.73	
O_2	34.30	CO_2	106.50	
NH_3	28.96			
SO_4	0.30			

In fluxes		Out fluxes		Objective
Metabolites	Fluxes	Metabolites	Fluxes	flux
Acetate	100.00	H^+	1.97E+03	1.99
O_2	57.05	H_2O	346.85	
Pi	9.81	CO_2	130.71	
NH_3	1.46			
SO_4	0.01			

In fluxes		Out fluxes		Objective
Metabolites	Fluxes	Metabolites	Fluxes	flux
Ethanol	100.00	H^+	2.57E+03	3.78
O_2	68.22	H_2O	380.04	
Pi	19.7	CO_2	56.03	
NH_3	4.84			
SO_4	0.08			

In fluxes		Out fluxes		Objective
Metabolites	Fluxes	Metabolites	Fluxes	flux
Pyruvate	100.00	H^+	1.41E+03	4.10
O_2	39.05	H_2O	174.89	
Pi	20.45	CO_2	147.64	
NH_3	5.88			
SO_4	0.05			

3.6 Discussion

3.6.1 The metabolic pathway distribution of *E. gracilis* is substantially different from other photosynthetic organisms, which results in its ability to survive in the absence of a plastid.

Based on multiple strands of evidence, I propose a model of the subcellular distribution of the reactions of central carbon metabolism in *Euglena* (Figure 3.5). The model shows that the subcellular location of PPP differs from those found in other photosynthetic organisms. In *Euglena*, the complete PPP is present in the cytosol with a duplicated non-oxidative phase present in the plastid, but not the oxidative phase. This is different from that found in green plants in which the complete PPP is located in the chloroplasts (Kruger and von Schaewen, 2003; Schwender et al., 2003).

The distribution of the metabolic pathways also indicates that, aside from the reactions of photosynthesis, all the central metabolic pathways needed for cell growth in heterotrophic conditions are located in the cytosol and mitochondria of the cells. This includes the biosynthesis of isoprenoids, as the mevalonate pathway is present in the cytosol. The ability to survive on a simple carbon source when their chloroplasts have been destroyed, well-documented in *E. gracilis* but not in other plastid containing organisms (Shao et al., 2019), can be rationalised from the subcellular localisation and duplication of these critical pathways.

3.6.2 The reconstruction of the metabolic network model of *E. gracilis* provides the basis for flux analysis under heterotrophic conditions.

In this study, I reconstructed the metabolic network of *E. gracilis* based on transcriptomic data (O'Neill et al., 2015, Yoshida et al., 2016) and a review of central metabolic pathways of *E. gracilis* (Inwongwan et al., 2019). Although the culture conditions for these transcriptomes are not exactly matched with the growth conditions in this thesis, the draft model demonstrates the complete central metabolic pathways, which is sufficient for the scope of this study. The network model was curated, mainly focusing on the metabolic pathways involved in heterotrophic metabolism in *Euglena*. The pathways or reactions of photosynthesis were not extensively curated or tested as it is beyond the scope of this study. The ability of *E. gracilis* to survive stably without its chloroplast indicates that, without the complete plastidic pathways, the model could represent the operation of the metabolic network of *E. gracilis* under heterotrophic, but not autotrophic or mixotrophic conditions.

The integrity of model was tested using FBA with the objectives of maximising biomass production and minimising the numbers of fluxes through the network. The prediction of fluxes shows that the model is able to predict the biomass production of *E. gracilis* grown on the organic carbon substrates. This indicates that this model can provide a basis for the study of the metabolic network of *E. gracilis* under heterotrophic conditions, and could also potentially be used as a foundation to develop flux prediction in autotrophic conditions. This will lead to the ability to predict the metabolic phenotypes of *Euglena* under various growth conditions. The application of the model will be discussed in later chapters.

3.7 Conclusion

Other than the reactions of photosynthesis, pathways present in the chloroplast are duplicated in the cytosol or mitochondria, justifying the ability of this species to survive with the complete loss of chloroplasts. Transcriptomic data provides the basis of the reconstruction of the metabolic network of *E. gracilis* under heterotrophic conditions. The network model was reconstructed and tested for FBA with the biomass synthesis reactions derived from biomass composition of *E. gracilis* grown on various carbon substrates. The results demonstrate the potential use of the model to investigate the metabolic phenotype of *E. gracilis* under heterotrophic conditions.

Chapter 4 | Metabolic phenotypes of wild-type and aplastidic *E. gracilis* grown on glucose

4.1 Introduction

4.1.1 Glucose metabolism in *Euglena*

Euglena can grow in a diverse set of conditions, such as autotrophic, heterotrophic or mixotrophic (Takeyama et al., 1997), aerobic or anaerobic, and over a wide range of pH (Yamane et al., 2001). In heterotrophic conditions, *Euglena* has the ability to utilise a range of organic substrates as the sole source of carbon and energy; these include sugars, short-chain alcohols, short-chain fatty acids and amino acids. As the most abundant monosaccharide in nature, glucose has been used extensively as a substrate for biosynthesis of compounds of commercial value by *Euglena*. For instance, glucose was demonstrated to be the optimal substrate for tocopherol (vitamin E) production in *Euglena*, compared to ethanol, glycerol and galactose (Ogbonna et al., 1998). The utilisation of glucose as a growth substrate by *E. gracilis* has been known for decades (Wolken, 1961). The pathways of glucose metabolism in *E. gracilis* are expected to be similar to those found in other eukaryotes, since many enzymes necessary for glycolysis, the PPP, the TCA cycle and gluconeogenesis have

been characterised in *Euglena* (Buetow, 1989, Hurlbert and Rittenberg, 1962). However, the metabolic phenotype of glucose metabolism in *E. gracilis* has not been studied.

4.1.2 Chloroplasts of *Euglena*

Euglena is one of the few plastid-containing organisms which can survive complete loss of the chloroplasts (Sato, 2011). Streptomycin (SM) can be used to destroy the chloroplasts in *E. gracilis* (Moriber et al., 1963, Ben-Shaul and Ophir, 1970). It was demonstrated that eliminating the chloroplasts of *Euglena* using SM does not affect cell viability (Provasoli et al., 1948). The mechanism of SM action on chloroplasts is still unclear, but it has been suggested that this antibiotic either inhibits the 16s ribosome subunit or directly affects the DNA of the chloroplasts of *Euglena* (Shao et al., 2019). Little work has been carried out on the fine structure of SM bleached *Euglena* chloroplasts (Ben-Shaul and Ophir, 1970). Some residue of the chloroplast genome was detected in certain ofloxacin and SM treated *E. gracilis* strains (Wang et al., 2004). The loss of parts of the plastid genome affects the cell function and encourages the formation of non-functional enzymes essential for chloroplast formation (Shao et al., 2019).

The evolution of *Euglena* is complicated by the multiple secondary endosymbiosis events that happened during its evolutionary history. This complex origin provides *Euglena* with genetic resources from multiple diverse sources which may contribute to the metabolic flexibility of *E. gracilis*, as can be seen from the broad range of its habitable environments. However, the influence of the endosymbiosis on genome sequence has not been thoroughly investigated because of the complex nature as a hybrid of a non-photosynthetic host cell and a secondary endosymbiont (Maruyama et al., 2011b). A recent genome study identified at

least 14 genes acquired through cryptic endosymbiotic gene transfer from algal sources different from the green alga that provided its current plastid (Ebenezer et al., 2017, Novák Vanclová et al., 2020). These transferred genes are involved in the integration and functioning of the secondary plastid of *Euglena*, and demonstrate the complexity of its genome evolution.

The chloroplast genome of *E. gracilis* has been sequenced and investigated (Hallick et al., 1993). The plastid genome of *Euglena* is found to be substantially similar to higher plants for its gene content, although the structure and evolution are different (Turmel et al., 2009). The study of the overall functional organisation of the *E. gracilis* plastid genome using microarray transcriptome analysis highlights that the expressed genes are unchanged although their relative expression levels vary across the stress conditions including UV light, temperature, antibiotics, herbicide and heavy metal exposure (Geimer et al., 2009). The rate of protein synthesis by the *Euglena* plastid in the dark was found to be extremely low compared to that in the light (Miller et al., 1983). One of the limiting factors of *E. gracilis* proplastid protein synthesis was suggested to be energy provision, as supplementing isolated proplastids with ATP and Mg^{2+} significantly increased the rate of protein synthesis (Malek et al., 1984, Cushman and Price, 1986).

Our recent review of the subcellular localisation of pathways in *Euglena* suggests a highly limited role of chloroplasts in heterotrophic metabolism (Inwongwan et al., 2019). Despite knowing that *Euglena* can survive the loss of its chloroplasts, the effect of the absence of chloroplasts on the metabolism of carbon substrates remains unknown.

4.1.3 Objectives

The major aims of the research work in this chapter are: (i) to define the metabolic phenotype of *E. gracilis* grown heterotrophically on glucose; and (ii) to investigate the effect of chloroplast loss on glucose metabolism in *E. gracilis*. These aims were addressed by exploiting ^{13}C MFA based on the metabolic model developed in Chapter 3, and the flux maps generated were used to define the metabolic phenotype of *E. gracilis* grown on glucose. To investigate the contribution of chloroplasts to heterotrophic carbon metabolism in *Euglena*, the metabolic phenotype of aplastidic cells was compared to that of the wild-type. FBA was incorporated in the analysis to assess the maintenance costs and the metabolic efficiency of the two strains.

4.2 Physiological characterisation of wild-type and aplastidic *E. gracilis* grown on glucose

4.2.1 Anatomy of *E. gracilis* strains

To investigate the effect of SM-induced chloroplast disruption on the cell structures of *E. gracilis*, the anatomy of 5-days-old wild-type *E. gracilis* cells grown under 40 µmol m^{-1} s^{-1} light, wild-type *E. gracilis* grown in the dark, and SM-treated *E. gracilis* grown in the dark on glucose-supplemented medium were examined using a TEM (Figure 4.1). The ultra-structure of wild-type *E. gracilis* observed in this study was similar to that described previously (Wolken, 1961).

The TEM images demonstrate that the elongated spherical *E. gracilis* cells are covered with the proteinaceous ridged wall called the pellicle. *Euglena* cells contained a nucleus, and several paramylon and mitochondria. The cells size is approximately 10 µm long and 5-7 µm wide. However, by comparing the cell structure of wild type grown in the light, wild type grown in the dark and aplastidic cells grown in the dark, it can be seen that the chloroplast structures are very different.

The TEM images demonstrate that the conventional/functional chloroplast structures (C) are clearly defined in the wild-type cell grown in the light (Figure 4.1.1 and 4.1.2). The chloroplasts are large and the layers of thylakoid membrane are clearly visible. When grown in the dark (Figure 4.1.3 and 4.1.4), wild-type cells contain compressed chloroplasts which are visually different from the functioning chloroplasts. The compressed chloroplast structures are smaller compared to the chloroplast observed in light-grown *E.*

gracilis and the thylakoid membranes are not clearly defined. On the other hand, recognisable chloroplasts are not observed in the SM-treated cells (Figure 4.1.5 and 4.1.6).

As the heterotrophic wild-type *E. gracilis* cells contain compressed chloroplasts, the metabolic pathways located in the plastid cannot be completely excluded from the metabolic network model of wild-type cells under heterotrophic conditions. In this study, total elimination of chloroplasts is assumed in SM-treated *E. gracilis* based on the TEM images and previous reports that SM treatment results in loss of the plastid genome (Wang et al., 2004). The loss of plastid genes in aplastidic strains does not affect cell growth in the dark (Hadariová et al., 2017), indicating that the plastid genes are not essential for the survival of *E. gracilis*. Although I have not tested for the residual plastid DNAs of aplastidic cells to ensure the total elimination of all plastid function, the physiological role of residual plastid DNAs in heterotrophic conditions is still poorly investigated (Qin et al., 2020).

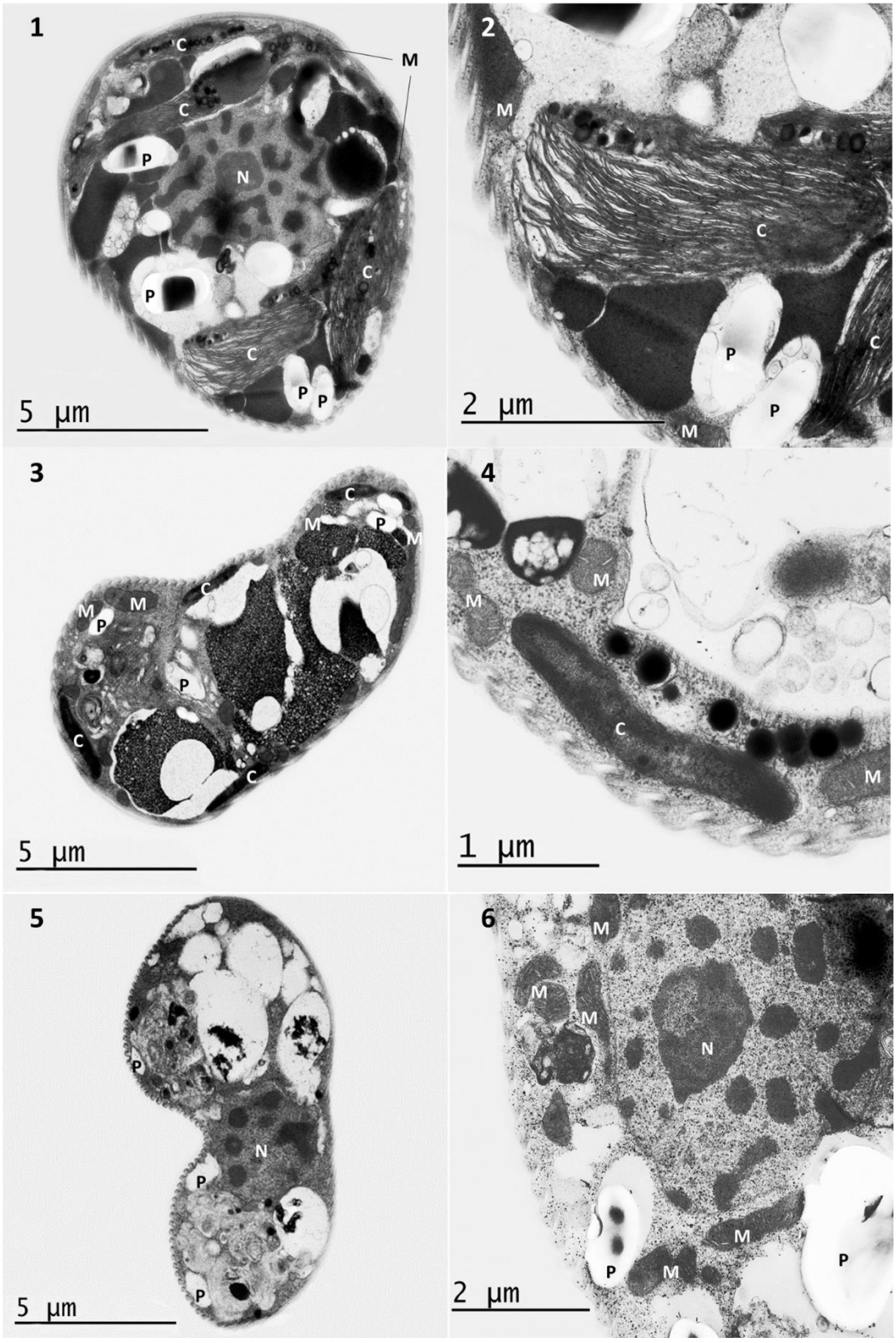
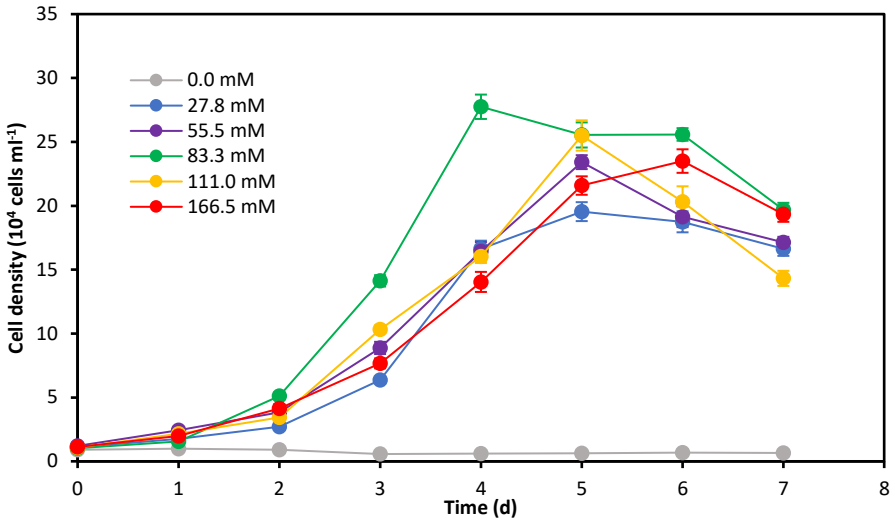


Figure 4.1. The anatomy of *E. gracilis* strains. The images of late exponential (~5 days old) wild-type *E. gracilis* grown in the light (1&2), wild-type *E. gracilis* grown in the dark (3&4) and

aplastidic *E. gracilis* grown in the dark (5&6) were taken using TEM; C = chloroplast, M = mitochondria, P = paramylon granule and N = nucleus. The images show the presence of compressed chloroplast in wild-type cells grown in the dark, and demonstrate the loss of chloroplasts in aplastidic strain.

4.2.2 Growth of wild-type *E. gracilis* grown heterotrophically on glucose

To identify an optimal concentration of glucose for growth, wild-type *E. gracilis* was cultivated on 30 ml of AF6 medium supplemented with up to 166.5 mM glucose. The cultures were grown in the dark with four replicates and cell density for each concentration was monitored until reaching stationary phase (Figure 4.2). The growth of *E. gracilis* required the presence of glucose in the AF6 medium. On the different concentrations of glucose*, E. gracilis* showed a similar growth pattern, containing three phases of growth: lag phase, exponential phase and stationary phase. *E. gracilis* cultures with 27.8, 55.5 and 111.0 mM of glucose showed a similar growth pattern. The cultures had same duration of lag phase of two days and reached the stationary phase on day five with cell densities of $19.5 \pm 0.7 \times 10^4$, $23.4 \pm 0.5 \times 10^4$ and $25.5 \pm 1.2 \times 10^4$ cells ml^{-1} respectively. With the same duration of lag phase, *Euglena* cultures supplemented with 166.5 mM glucose reached the peak on day six with $23.5 \pm 0.9 \times 10^4$ cells ml^{-1}. *Euglena* cultures supplemented with 83.3 mM glucose had the fastest growth and the highest cell density at peak, at about $27.8 \pm 1.0 \times 10^4$ cells ml^{-1} on day four with one day of lag phase. Thus, the glucose concentration of 83.3 mM (15 g L^{-1}) was chosen for the further analyses.



	[glucose] (mM)					
	0.0	27.8	55.5	83.3	111.0	166.5
Maximum cell density (10^4 cells ml^{-1})	1.0 ± 0.1	19.5 ± 0.7	23.4 ± 0.5	27.8 ± 1.0	25.5 ± 1.2	23.5 ± 0.9
Lag phase (d)	-	2	2	1	2	2

Figure 4.2. Glucose concentration dependence of heterotrophic wild-type *E. gracilis* growth. The cell density values are the mean ± SD of four independent biological replicates, monitored daily by counting cell number using haemocytometer. Lag phase was estimated by the period before the extrapolated exponential phase.

4.2.3 Comparison of growth characteristics of wild-type and aplastidic *E. gracilis*

Aplastidic and wild-type *E. gracilis* cells were cultured heterotrophically in AF6 medium supplemented with 83.3 mM glucose, and the cell density of each strain was monitored until cell growth reached stationary phase. Growth of the aplastidic *E. gracilis* was similar to that of the wild-type cells. Both strains showed one day of lag phase before entering the exponential growth phase and reached a peak cell density at about 25×10^4 cells ml^{-1} on day six (Figure 4.3).

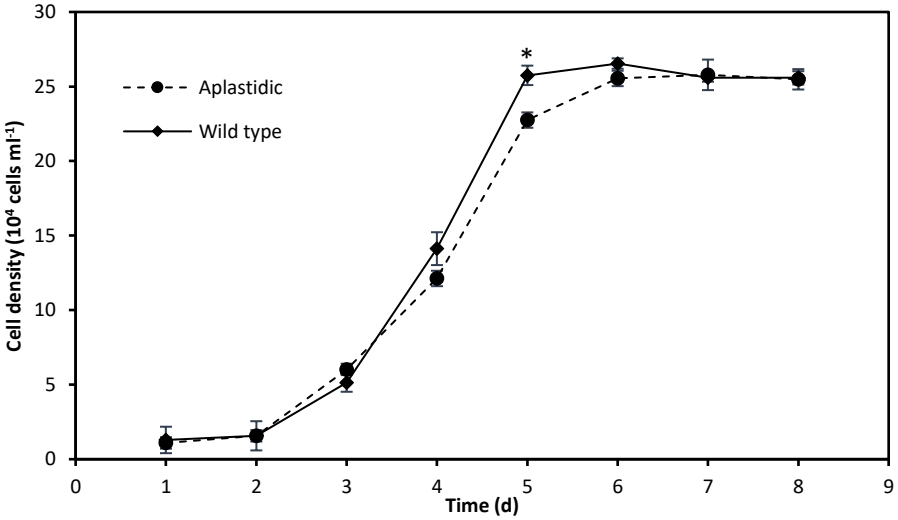


Figure 4.3. Growth of wild-type and aplastidic *E. gracilis* grown heterotrophically on AF6 medium at optimal glucose concentration. The data presented are mean ± SD of independent biological replicates, cell density was monitored daily using haemocytometer. Lag phase was estimated from the period before the extrapolated exponential phase. The growths of both strains are similar across the culture period, except on day 5 when the cell density of wild-type cells was significantly higher than that of aplastic cells, as indicated by * (t-test, $P<0.05$).

Specific growth rates ± SD (Levasseur et al., 1993) during the exponential phase of wild type and aplastidic strains were 0.95 ± 0.13 and 0.89 ± 0.09 d^{-1} respectively (Figure 4.4), which were not significantly different ($p > 0.05$). Comparison of the growth characteristics of the two strains suggested that the process of chloroplast bleaching using SM has no significant effect on growth of *E. gracilis* on glucose medium under heterotrophic conditions.

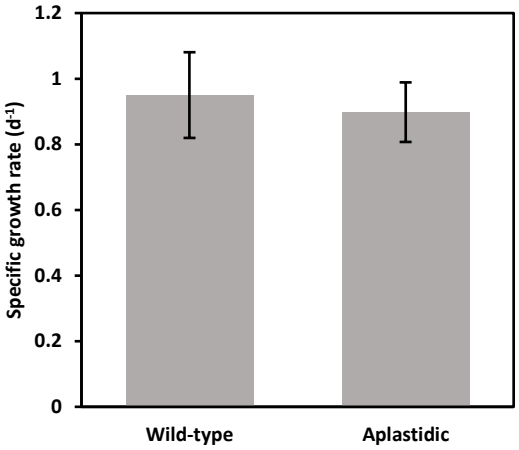


Figure 4.4. **Specific growth rate during exponential growth phase of wild-type and aplastidic *E. gracilis* grown heterotrophically on glucose medium.** The specific growth rate was calculated from cells density of day 2 to day 5 using the established formula (Levasseur et al., 1993). The data shown in the figure are the mean ± SD of four biological replicates. Specific growth rates of wild-type and aplastidic cells are not significantly different at 0.95 ± 0.13 and 0.89 ± 0.09 d^{-1} respectively (t-test, $p > 0.05$).

The rates of glucose uptake of both strains were analysed using a ^{14}C labelling technique (Figure 4.5). Glucose uptake by wild-type and aplastidic *E. gracilis* cultures during exponential growth phase showed an unexpected initial lag phase during the first hour of

incubation. This might be the response of the cells to altered conditions during the analysis. Based on the data from 1 h onwards, the rate of glucose uptake by wild-type *E. gracilis* cells was 8.15 ± 0.63 µmol h^{-1} 10^{-6} cells, while the rate of uptake by aplastidic cells was 7.79 ± 0.58 µmol h^{-1} 10^{-6} cells. The rates of glucose consumption by the two *E. gracilis* strains were not significantly different ($p > 0.05$).

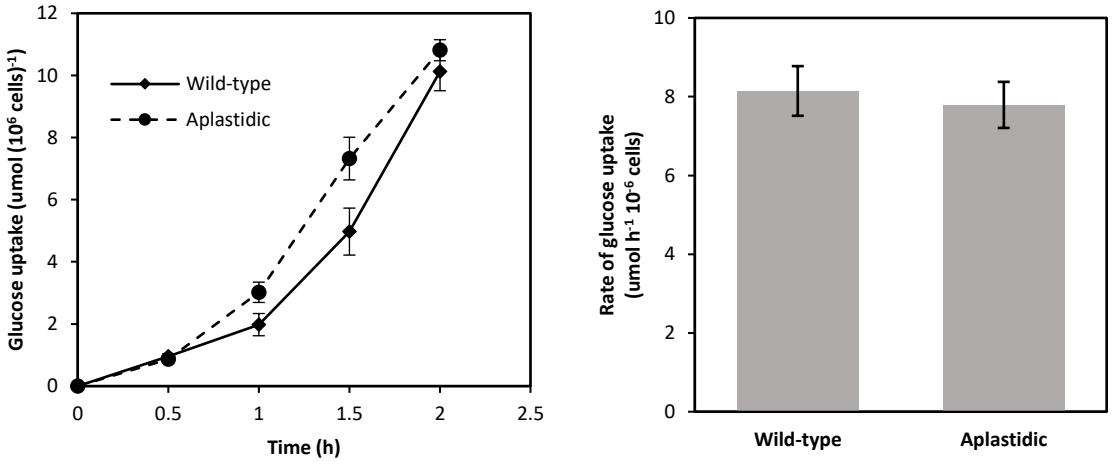


Figure 4.5. ^{14}C Glucose uptake (left) and rate of uptake (right) of *E. gracilis* strains in the exponential growth phase. Data are the mean ± SD of four biological replicates. The uptake rates of wild-type and aplastidic cells are not significantly different (t-test, $p > 0.05$).

4.2.4 Biomass composition of *E. gracilis* grown heterotrophically on glucose

In this study, four main components of the biomass of wild-type and aplastidic *E. gracilis* grown heterotrophically on glucose medium were determined (Table 4.1). The components consist of protein, insoluble carbohydrate, soluble carbohydrate and lipid. The analysed biomass components of wild-type and aplastidic *E. gracilis* constitute about 95.5%

and 93.8% of dry cell mass, respectively, and the residual components are likely to consist predominantly of mineral ions and minor organic compounds, including nucleic acids.

The protein contents of wild-type and aplastidic *Euglena* cells grown in glucose were 16.9 ± 1.1% and 17.2 ± 2.0% of dry cell mass respectively, which were not significantly different to each other ($p > 0.05$). Wild-type *E. gracilis* accumulated 40.2 ± 0.4% insoluble carbohydrate, which is significantly lower than aplastidic cells that contained 58.8 ± 2.1% insoluble carbohydrate. In contrast, the total carbohydrate content of wild-type and aplastidic *E. gracilis* were not significantly different, at 75.2 ± 4.2% in wild-type and 71.8 ± 4.9% in aplastidic cells. It can be seen that total carbohydrate content exceeds the amount of insoluble carbohydrate for both *E. gracilis* strains. Insoluble carbohydrate content was subtracted from total carbohydrate to estimate the proportion of soluble carbohydrate, the results showing that wild-type and aplastidic *E. gracilis* biomass contained about 35% and 18% of soluble carbohydrate, respectively. Total lipid analysis shows that the biomass of wild-type and aplastidic cells grown on glucose contained relatively low proportion of lipid compared to other components. The lipid contents of the two strains were significantly different, at 3.4 ± 0.3% and 4.9 ± 0.6%, respectively.

Overall, the main component of biomass for both strains is carbohydrate which makes up over 70% of dry weight. However, the ratio of insoluble and soluble carbohydrate differed between the two strains. The second most abundant component was protein which constituted about 17% of dry cell mass. Lipid contributes the lowest proportion of the biomass compared to other components. About 4.5% of wild-type and 6.2% of aplastidic

biomass are not identified in this study (Figure 4.6), which might include nucleic acids, other soluble metabolites, such as organic acids, lactic acid, carotenoids and minerals.

Table 4.1. Biomass composition of wild-type and aplastidic *E. gracilis* grown on glucose medium

Biomass component	Biomass composition (% DW)		P value
	Wild-type	Aplastidic	
Protein	16.9 ± 1.1	17.1 ± 2.0	0.89
Insoluble carbohydrate	40.2 ± 0.4	53. 8± 2.1*	0.01
Total carbohydrate	75.2 ± 4.2	71.8 ± 4.9	0.33
Lipid	3.4 ± 0.3	4.9 ± 0.6*	0.03
Residue	4.5	6.2	

Biomass compositions of *E. gracilis* strains are the mean ± SD of four biological replicates; * indicated the data that is significantly different between the two strains (t-test, $p<0.05$).

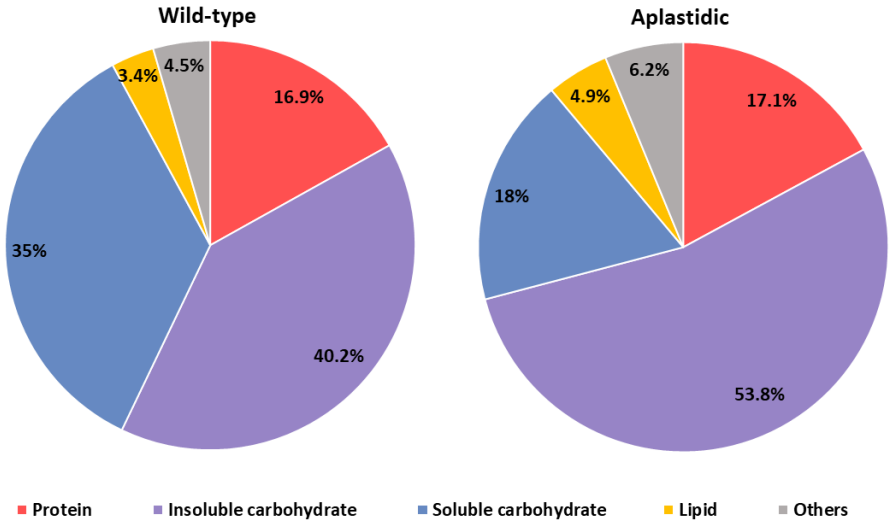


Figure 4.6. Biomass composition of wild-type and aplastidic *E. gracilis* grown heterotrophically on glucose.

4.3 ^{13}C MFA of *E. gracilis* grown on glucose

4.3.1 Flux map of wild-type *E. gracilis* grown on glucose

^{13}C MFA was performed on wild-type *E. gracilis* cultures grown heterotrophically and in parallel on AF6 medium supplied with three different ^{13}C labelled sources of glucose: [1-^{13}C]glucose, [2-^{13}C]glucose and 20% [$^{13}C_6$]glucose. The MIDs of amino acids obtained from protein hydrolysate of cell extract were used to generate the flux map that describes the production of biomass components constrained by the biomass synthesis reactions defined in the metabolic model constructed in Chapter 3. At the confidence level of 95%, the analysis showed that the SSR is statistically accepted (SSR = 492.2, expected SSR = 609.6 - 754.2) at the degree of freedom of 680, indicating that the measured MIDs are comparable to those of the simulation.

The flux map of wild type *E. gracilis* grown heterotrophically on glucose shows that *E. gracilis* metabolises glucose through the conventional cytosolic central metabolic pathways, including glycolysis, the PPP and the TCA cycle (Figure 4.7). During glucose metabolism, the plastidic pathways (the nonoxPPP and glycolysis) carried significantly lower fluxes compared to the same pathways in the cytosol, and the flux values of these plastidic pathways are negligibly small. Of the glucose taken up, about 41% of G6P generated was the direct precursor for biomass components, while 52% was converted to F6P by phosphoglucoisomerase (EC 5.3.1.9). An amount of F6P equivalent to 18% of glucose intake was withdrawn to produce biomass, and the remainder (equivalent to 34% of glucose uptake) together with F6P produced from the PPP was further metabolised via glycolysis. Only about 7% of the glucose was metabolised via the PPP. Flux through the glyoxylate shunt was limited

in this condition at about 1 mol (100 mol glucose)$^{-1}$ [0.08 µmol h^{-1} 10^{-6} cells]. Malic enzyme (EC 1.1.1.40) and PEPC (EC 4.1.1.31) were active in the precursor provision of the TCA cycle for biomass production. PEPC contributed 20 mol oxaloacetate (100 mol glucose)$^{-1}$ [1.63 µmol h^{-1} 10^{-6} cells] and about 9 mol malate (100 mol glucose)$^{-1}$ [0.73 µmol h^{-1} 10^{-6} cells] was converted to pyruvate by malic enzyme.

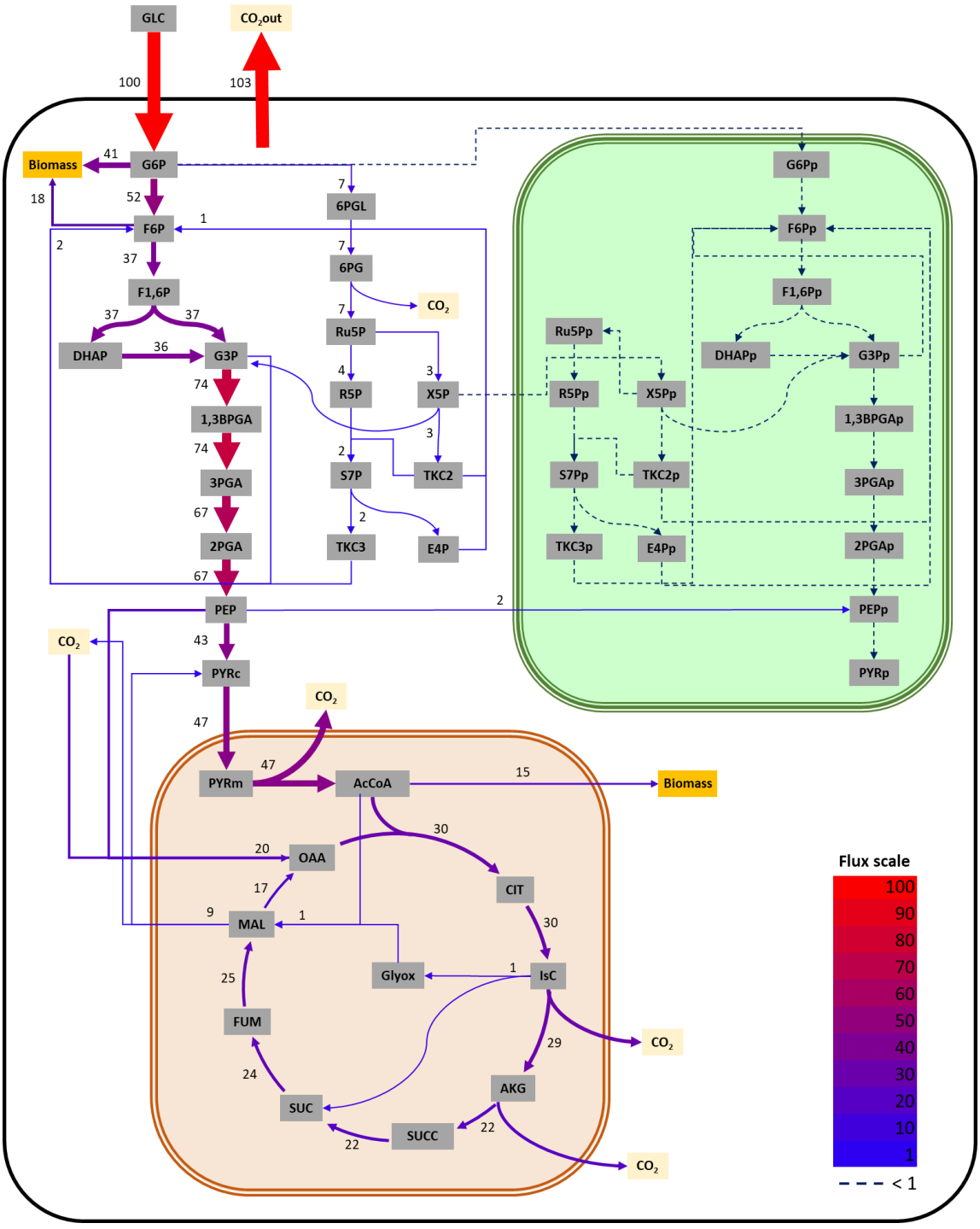


Figure 4.7. Metabolic flux map of wild-type *E. gracilis* grown heterotrophically on glucose. The thickness of arrows and the numbers represent mol of fluxes per 100 mol of glucose input.

4.3.2 Flux map of aplastidic *E. gracilis* grown on glucose

As with wild-type *E. gracilis*, ^{13}C MFA was conducted on aplastidic *E. gracilis* cultures grown heterotrophically on AF6 medium using the same methods and labelling strategies. The central metabolic network model of *E. gracilis* was modified for the analysis of aplastidic cells by omitting all plastidic reactions. The result showed the statistically accepted SSR of 541.7 compared to the expected SSR of 640.0 - 787.8 at the confidence level of 95% and degree of freedom of 712. The flux values were visualised as a flux map (Figure 4.8).

As with the wild-type cells, aplastidic *E. gracilis* metabolised glucose predominantly via central metabolic pathways, including glycolysis, the PPP, anaplerotic reactions and the TCA cycle. Of the glucose intake, 37% of G6P and 7% of F6P produced were directly withdrawn to produce biomass, whilst about 11% was metabolised through the PPP. F6P, equivalent to 51% of glucose taken up, was subsequently metabolised via glycolysis. About 20 mol oxaloacetate (100 mol glucose)$^{-1}$ [1.56 µmol h^{-1} 10^{-6} cells] was produced from PEP by PEPC, and about 11 mol malate (100 mol glucose)$^{-1}$ [0.86 µmol h^{-1} 10^{-6} cells] was converted to pyruvate by malic enzyme. The glyoxylate shunt of aplastidic cells showed negligible flux.

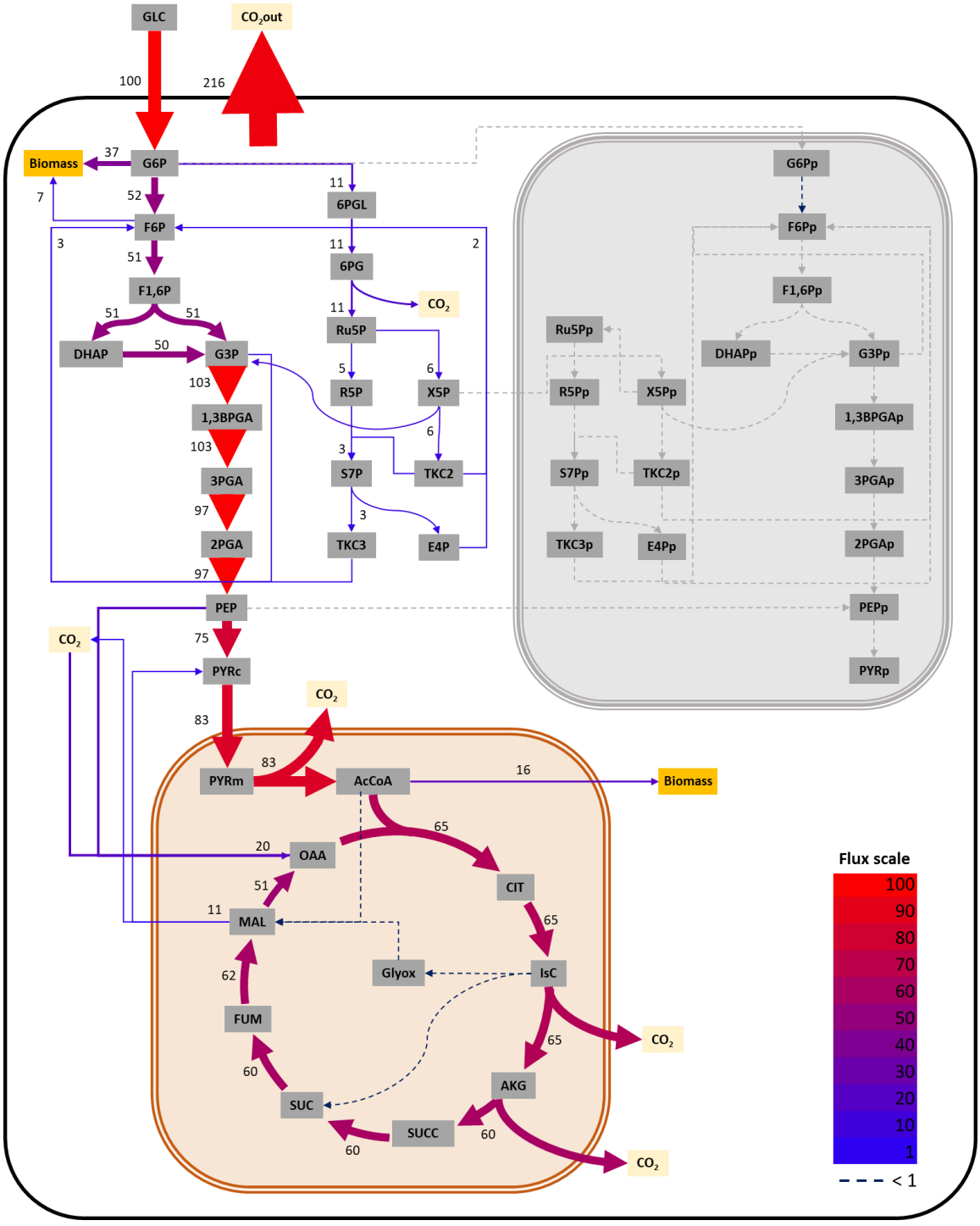


Figure 4.8. Metabolic flux map of aplastidic *E. gracilis* grown heterotrophically on glucose. The thickness of arrows and the numbers represent mol of fluxes per 100 mol of glucose input.

4.3.3 Comparing flux maps of wild-type and aplastidic *E. gracilis* grown on glucose

Even though the growth characteristics of the two strains were similar, their metabolic phenotypes differed substantially. The fluxes through the central metabolic pathway with the upper- and lower-bound of confidence limits are shown in Appendix C. In detail, fluxes through glycolysis, the PPP and the TCA cycle were significantly higher in the aplastidic strain. Less G6P and F6P were diverted to produce biomass in the aplastidic strain, leading to higher fluxes through the following TCA steps where about 65 mol citrate (100 mol glucose)$^{-1}$ was generated and further metabolised through the TCA cycle, more than double the amount generated in wild-type cells (about 30 mol citrate (100 mol glucose)$^{-1}$). A smaller proportion of the intermediates of the TCA cycle were used to produce biomass in aplastidic cells, corresponding with a lower rate of biomass synthesis. PEPC catalysed an equal flux of 20 mol (100 mol glucose)$^{-1}$, while 11 and 9 mol pyruvate (100 mol glucose)$^{-1}$ were produced by malic enzymes for aplastidic and wild-type cells respectively. In both strains, the glyoxylate shunt was shown to be barely active during the growth on glucose.

4.3.4 CCE of wild-type and aplastidic *E. gracilis* grown on glucose

In wild-type cells, the flux map showed that about 8.39 µmol CO_2 h^{-1} 10^{-6} cells was released from the cell which is about 17% of total carbon taken up. Released CO_2 was mainly produced from pyruvate dehydrogenase (PDH) (3.83 µmol h^{-1} 10^{-6} cells) and the TCA cycle (4.16 µmol h^{-1} 10^{-6} cells), while only 0.73 and 0.57 µmol h^{-1} 10^{-6} cells of CO_2 was produced by malic enzyme and the oxPPP respectively. The remaining CO_2 was produced via reactions of amino acid synthesis and biomass production. About 1.63 µmol h^{-1} 10^{-6} cells CO_2 was fixed by PEPC. CCE can be calculated as the proportion of carbon atoms used for biomass production

(Alonso et al., 2007). Based on the flux of CO_2 release, the CCE of wild-type *E. gracilis* grown on glucose was about 83%.

For aplastidic cells, the flux map showed about double total CO_2 releases to that of wild-type cells, at about 36% of total carbon intake (16.83 µmol h^{-1} 10^{-6} cells). As with the wild-type strain, the majority of CO_2 was produced via the TCA cycle (9.74 µmol h^{-1} 10^{-6} cells). PDH released 6.46 µmol h^{-1} 10^{-6} cells of CO_2. The oxPPP and malic enzyme produced the same amount of CO_2 at about 0.85 µmol h^{-1} 10^{-6} cells. The remainder of the CO_2 production was generated from amino acids and biomass synthesis. About 1.56 µmol h^{-1} 10^{-6} cells of CO_2 was used by PEPC to produce oxaloacetate. CCE based on flux of CO_2 released from aplastidic *E. gracilis* was lower than that of wild-type cells, at about 64%.

Using an independent approach to corroborate the fluxes measured by ^{13}C MFA, ^{14}C labelled glucose was fed to the cultures of both strains, and the CCE was calculated from the radioactivity incorporated into *E. gracilis* cells and in released CO_2. With this approach, CCEs for wild-type and aplastidic *E. gracilis* grown on glucose were measured as 70.4% (57.4 – 83.4%) and 44.47% (34.6 – 54.4%) respectively, with lower and upper standard confidence limits are shown in brackets. The CCEs estimated from the ^{14}C labelling approach and ^{13}C MFA were compared (Figure 4.9). For both strains, the CCEs measured by the ^{14}C approach were lower than those from ^{13}C MFA, this could be the result of the underestimation of CO_2 release in MFA that did not account for degradation of unlabelled metabolites in the cell resulting from the natural turnover of cellular components. However, despite this quantitative discrepancy, CCEs from both methods indicate the same trend that the CCE of aplastidic *E.*

gracilis was less than that of wild-type cells when exploiting glucose as the sole respiratory substrate.

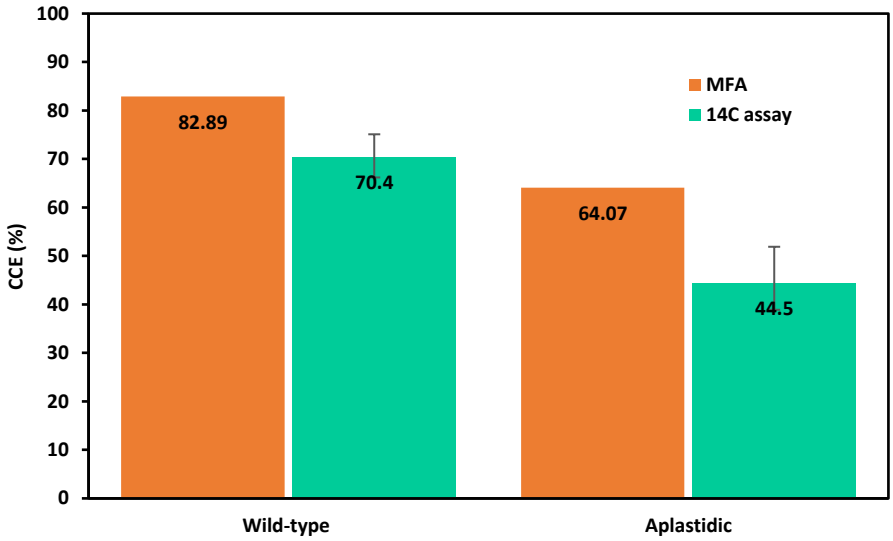


Figure 4.9. Comparison of CCE of *E. gracilis* strains calculated from ^{13}C MFA and ^{14}C labelling approach. The error bars shown in the figure are lower and upper standard confidence limits.

4.3.5 Cofactor provision in wild-type and aplastidic *E. gracilis* grown on glucose

The production and consumption of cofactors, ATP, NADH, $FADH_2$ and NADPH, in glucose heterotrophic metabolism in wild-type *E. gracilis* (Figure 4.10) were obtained from the deduced flux maps based on stoichiometries shown in Appendix D. In heterotrophic conditions, wild-type *E. gracilis* grown on glucose produced ATP through substrate-level phosphorylation in glycolysis and the TCA cycle at the rates of 9.64 and 1.80 µmol h^{-1} 10^{-6} cells, respectively. This ATP production is only about 50% of the estimated rate of ATP consumption needed to support biosynthesis (22.06 µmol h^{-1} 10^{-6} cells). NADH was produced at similar rates via the TCA cycle and glycolysis, at 5.59 and 6.03 µmol h^{-1} 10^{-6} cells respectively. 1.93 µmol $FADH_2$ h^{-1} 10^{-6} cells was also produced by the TCA cycle. The

consumption of NADH was estimated as 5.25 µmol h^{-1} 10^{-6} cells, which is half the rate of NADH production. $FADH_2$ and the remaining NADH can be used to produce ATP via the electron transport chain to meet the ATP demand of the cells. The majority of NADPH was produced via $NADP^+$-dependent PDH (3.85 µmol h^{-1} 10^{-6} cells), and only 1.10 and 0.76 µmol NADPH h^{-1} 10^{-6} cells were produced from the oxPPP and malate dehydrogenase, respectively. The total NADPH production is higher than the consumption estimated from the MFA, at 2.54 µmol h^{-1} 10^{-6} cells.

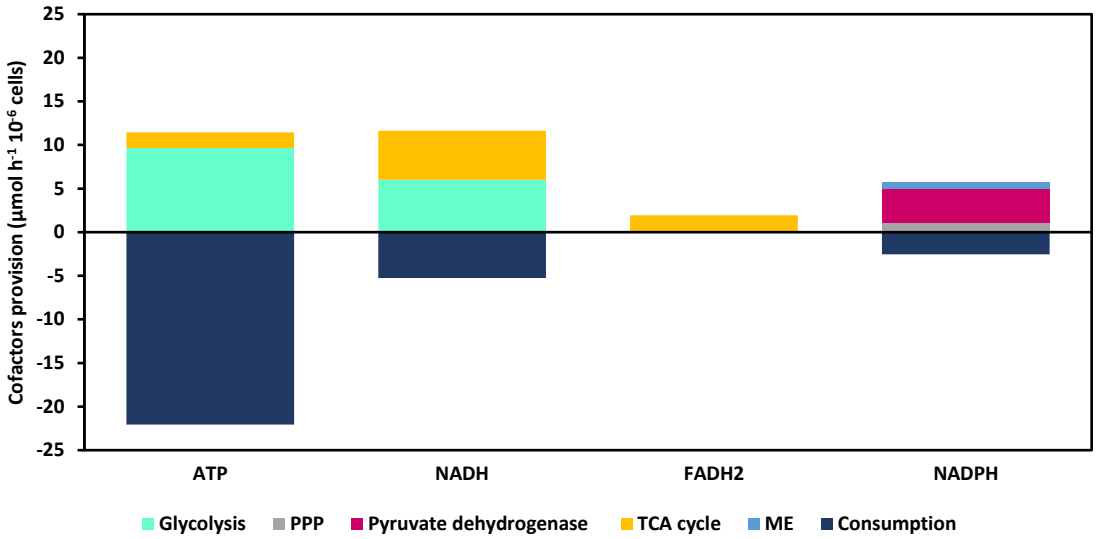


Figure 4.10. Production and consumption of cofactors from ^{13}C MFA of wild-type *E. gracilis* grown on glucose medium under the heterotrophic condition. The considered cofactors consist of ATP, NADH, $FADH_2$ and NADPH. The production of cofactors was categorised by their sources, including glycolysis, the PPP, pyruvate dehydrogenase, the TCA cycle and malic enzyme (ME).

The rates of ATP, NADH, $FADH_2$ and NADPH production by the central metabolic network of aplastidic *E. gracilis* grown on glucose was determined using the same approach

(Figure 4.11). In heterotrophic conditions, the substrate-level phosphorylation of glycolysis and the TCA cycle provided about 13.93 and 4.69 µmol ATP h^{-1} 10^{-6} cells respectively. This route of ATP production is insufficient to support the estimated ATP consumption of 19.03 µmol ATP h^{-1} 10^{-6} cells. The majority of NADH was produced via the TCA cycle at about 13.74 µmol h^{-1} 10^{-6} cells, while glycolysis produced a further 8.06 µmol NADH h^{-1} 10^{-6} cells. In contrast, the rate of NADH consumption was only 3.79 µmol h^{-1} 10^{-6} cells, allowing the surplus NADH to be used to produce ATP via oxidative phosphorylation to meet the ATP demand of the cells. About 4.70 µmol $FADH_2$ h^{-1} 10^{-6} cells were produced via the TCA cycle. As with wild-type cells, the majority of NADPH was produced via $NADP^+$-dependent PDH. This reaction produced 6.45 µmol NADPH h^{-1} 10^{-6} cells, and 1.69 and 0.87 µmol h^{-1} 10^{-6} cells NADPH were produced via the oxPPP and malic enzyme respectively. The combined rate of NADPH production is much higher than the demand for NADPH consumption by biosynthesis which is 2.18 µmol h^{-1} 10^{-6} cells.

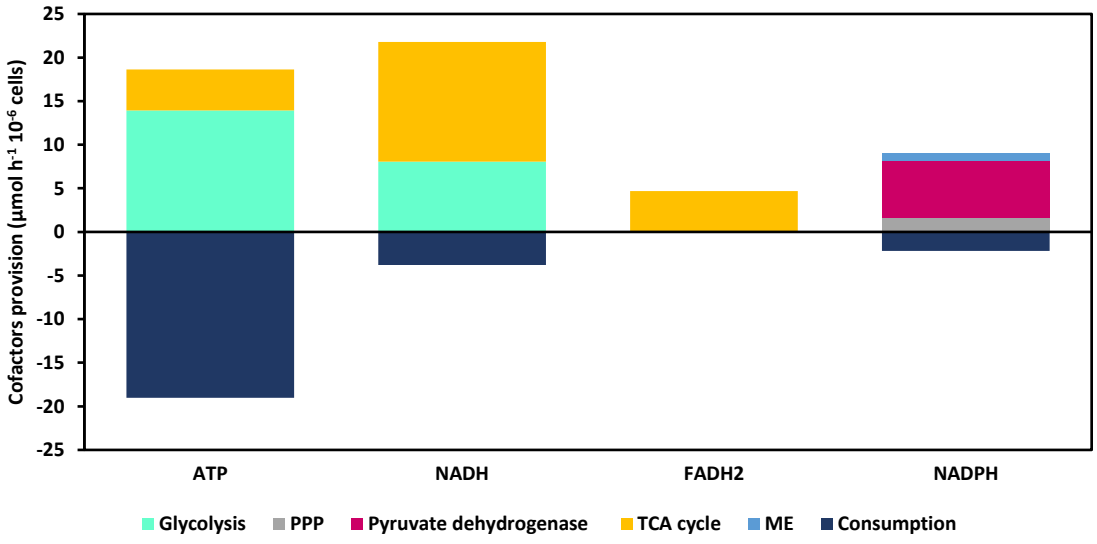


Figure 4.11. Production and consumption of cofactors estimated by ^{13}C MFA of aplastidic *E. gracilis* grown on glucose medium under the heterotrophic condition. The considered cofactors consist of ATP, NADH, $FADH_2$ and NADPH. The production of cofactors was categorised by their sources, including glycolysis, the PPP, pyruvate dehydrogenase, the TCA cycle and malic enzyme (ME).

Overall, the results demonstrate that the rate of ATP production from substrate-level phosphorylation is insufficient to support the ATP demands for biosynthesis in either strain. The ATP generated from this process contributes about 52.8% and 97.8% of ATP requirement for biosynthesis in wild-type and aplastic strains, respectively. However, the production of NADH and $FADH_2$ exceeds the immediate metabolic requirements for these cofactors. The surplus reductant, in the form of NADH and $FADH_2$, can be used for ATP production via the electron transport chain. The ATP production from this process was estimated at about 1.8 and 127.0 times the remaining ATP requirement for biosynthesis (18.82 and 52.08 μmol h^{-1} 10^{-6} cells) of wild-type and aplastidic strains, respectively. In comparison between the two strains, the central metabolic pathways of aplastidic cells produced almost twice as much of each cofactor relative to the wild-type cells, whereas the estimated rates of cofactors consumption in both strains were comparable (Table 4.2). Considering that the growth characteristics of wild-type and aplastidic strains are similar, this implies a higher maintenance demand in the aplastidic strain.

Table 4.2. ^{13}C MFA estimated the overall production and consumption of cofactors in wild-type and aplastidic *E. gracilis* grown heterotrophically on glucose

Cofactors	Production (μmol h^{-1} 10^{-6} cells)		Consumption (μmol h^{-1} 10^{-6} cells)	
	wild-type	aplastidic	wild-type	aplastidic
ATP	11.44	18.62	22.06	19.03
NADH	11.62	21.80	5.25	3.79
$FADH_2$	1.93	4.70	0	0
NADPH	5.71	9.01	2.54	2.18

4.4 FBA of wild-type and aplastidic *E. gracilis* grown on glucose

4.4.1 FBA of *E. gracilis* strains grown on glucose with biomass constraint

The reconstruction of the metabolic network model generated based on the transcriptomes of *E. gracilis* for FBA was described in Chapter 3. In this section, the flux solutions for aplastidic and wild-type strains predicted by FBA using the model were investigated in detail. The original model was used for the wild-type strain, and the model was modified for FBA in aplastidic cells by excluding all chloroplastic reactions. The biomass composition of each strain was processed into biomass synthesis reactions as demonstrated in Section 3.3, and these reactions were used to constrain fluxes to produce the precursors for biomass with the specific ratios (mol precursors (g cell DW)$^{-1}$) determined from the biomass composition. The FBA was conducted using COBRApy (Ebrahim et al., 2013) with the objectives of maximising the biomass production and minimising total fluxes through the network, referred from hereon as the biomass constraint.

The FBA results of the *E. gracilis* strains demonstrate that the models of the metabolic network of wild-type and aplastidic *E. gracilis* constrained with only the biomass composition of each strain predicted very similar flux distributions. These results are visualised in Figure 4.12 (wild-type) and Figure 4.13 (aplastidic). FBA predicted that the chloroplastic pathways in the wild-type model were inactive. For both strains, glucose was predicted to be metabolised through the conventional routes involving glycolysis and the TCA cycle with substantially similar rates. Fluxes of the non-oxidative section of the PPP and malic enzyme were also predicted to be similar in both strains, while the oxPPP, PEPC and the glyoxylate shunt were predicted to be inactive. With the same rate of glucose intake, compared to flux maps by ^{13}C MFA, FBA with the biomass constraint of both strains predicted a higher proportion of G6P and F6P to be initially withdrawn for biomass synthesis, leading to substantially lower fluxes through glycolysis and the TCA cycle.

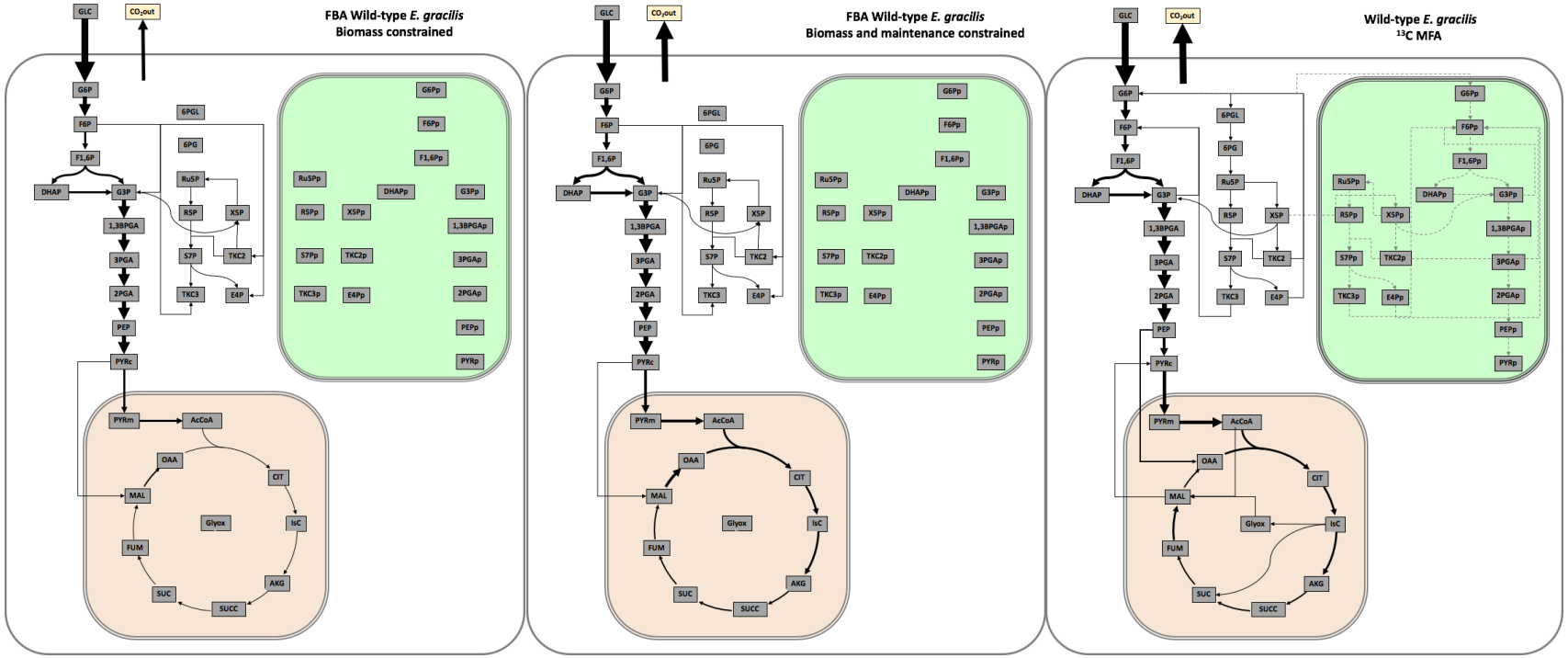


Figure 4.12. **Comparison of flux maps of wild-type *E. gracilis* grown heterotrophically on glucose from different approaches**; FBA with biomass constraint (left), FBA with biomass constraint and maintenance included (middle) and ^{13}C MFA (right). Thickness of the arrows indicates the fluxes relative to the rate of substrate input.

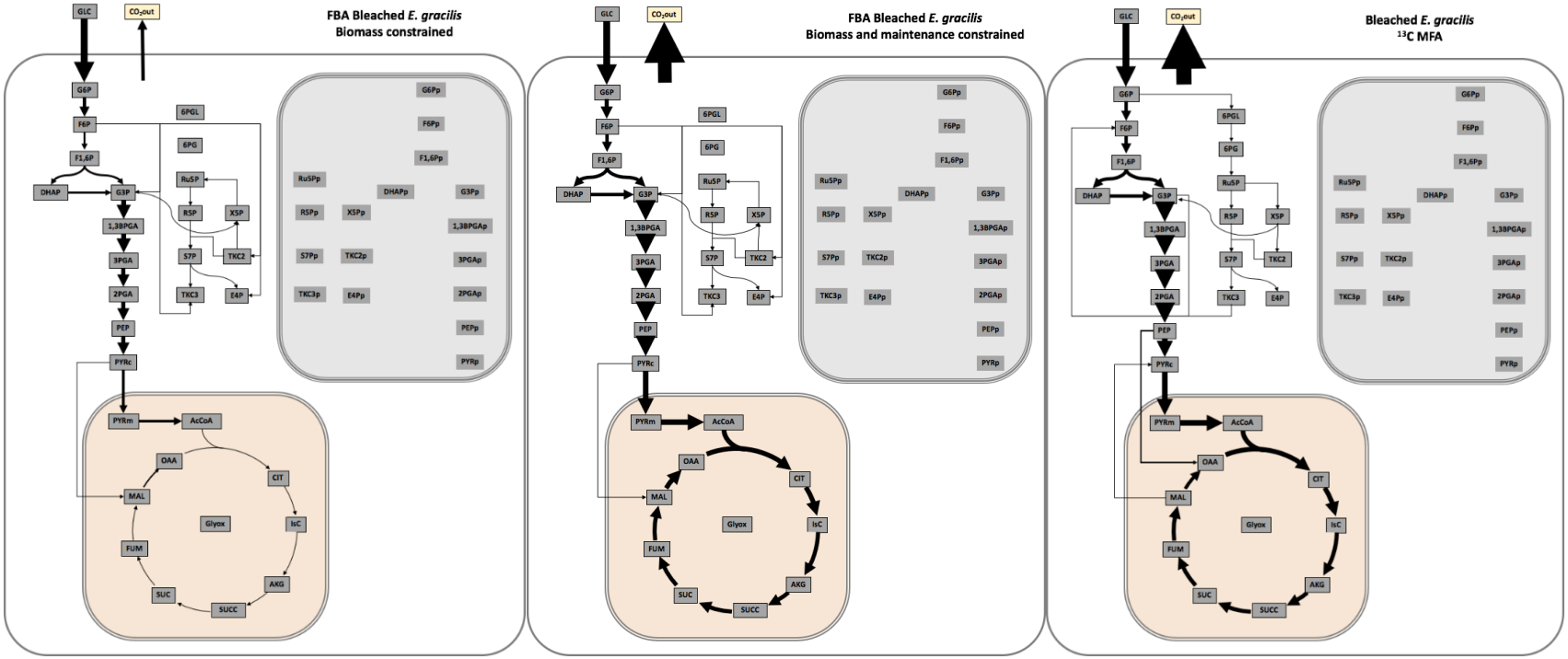


Figure 4.13. **Comparison of flux maps of aplastidic *E. gracilis* grown heterotrophically on glucose from different approaches**; FBA with biomass constraint (left), FBA with biomass constraint and maintenance included (middle) and ^{13}C MFA (right). Thickness of the arrows indicates the fluxes relative to the rate of substrate input.

4.4.2 Estimating cell maintenance of *E. gracilis* strains grown on glucose

Cell maintenance costs for FBA of plant networks has been estimated based on varying parameters that leads FBA solutions to match the experimental data. To improve flux predictions of FBA, the flux of an ATPase was varied until the consumption rate of carbon substrates matched the rate measured *in vivo*, with the surplus produced ATP considered as the maintenance cost (Williams et al., 2010). Furthermore, the trade-off ratios of ATP and NADPH production for maintenance were studied using P-areto optimality analysis. It has previously been shown that predicting and accounting for the cell maintenance costs as a specific ratio of ATP and NADPH consumption substantially improved flux prediction (Cheung et al., 2013).

To estimate the maintenance cost in *E. gracilis*, instead of using indirect methods to predict the maintenance cost of the cells, I undertook the calculation of the maintenance cost by comparing the cofactor provisions of the cells analysed by two independent approaches, FBA and ^{13}C MFA. This calculation is based on the assumptions that: (i) FBA with only the biomass constraint provides the minimal potential cofactors provision for biomass synthesis (no maintenance required); and (ii) the majority of cofactors used in the cells are produced in the central metabolic pathways and can be quantified using ^{13}C MFA. Normalised by glucose consumption rates, the differences between the cofactors provision measured by ^{13}C MFA and that predicted by FBA provides the estimated potential cofactors available for maintenance (Table 4.3).

Table 4.3. Cofactors production estimated from FBA and MFA results normalised to glucose uptake rate of wild-type and aplastidic *E. gracilis* grown on glucose

Cofactors	Wild-type (mol (mol glucose) $^{-1}$)			Aplastidic (mol (mol glucose) $^{-1}$)		
	FBA	^{13}C MFA	Available for maintenance	FBA	^{13}C MFA	Available for maintenance
ATP	1.24	1.40	0.16	1.32	2.39	1.07
NADH	1.05	1.54	0.49	1.03	2.89	1.86
$FADH_2$	0.07	0.24	0.17	0.05	0.60	0.55
NADPH	0.42	0.72	0.30	0.49	1.17	0.68

The NADH and $FADH_2$ produced for maintenance were assumed to be oxidised to generate ATP through the electron transport chain at a stoichiometry of 2.5 and 1.5 molecules of ATP produced per molecule of NADH and $FADH_2$ respectively (Foo et al., 2020). The total rate of ATP and NADPH production for maintenance in *E. gracilis* strains are depicted in Figure 4.14. The result shows that, relative to 1 mol of glucose intake, wild-type *E. gracilis* required 1.63 mol ATP and 0.30 mol NADPH for maintenance, corresponding to about 29% and 42% of the total produced respectively. In comparison to those in wild-type cells, aplastidic *E. gracilis* required about 4- and 2-times higher rates of ATP and NADPH for maintenance at 6.54 and 0.68 mol respectively. The maintenance of aplastidic cell is about 62% of total ATP produced and 58% of total produced for NADPH. This demonstrates that, to maintain the same growth rate, aplastidic *E. gracilis* requires substantially higher rates of cofactor production to meet the high demand for cell maintenance. The estimated

maintenance from this section was further introduced to FBA of both strains as demonstrated in the next section.

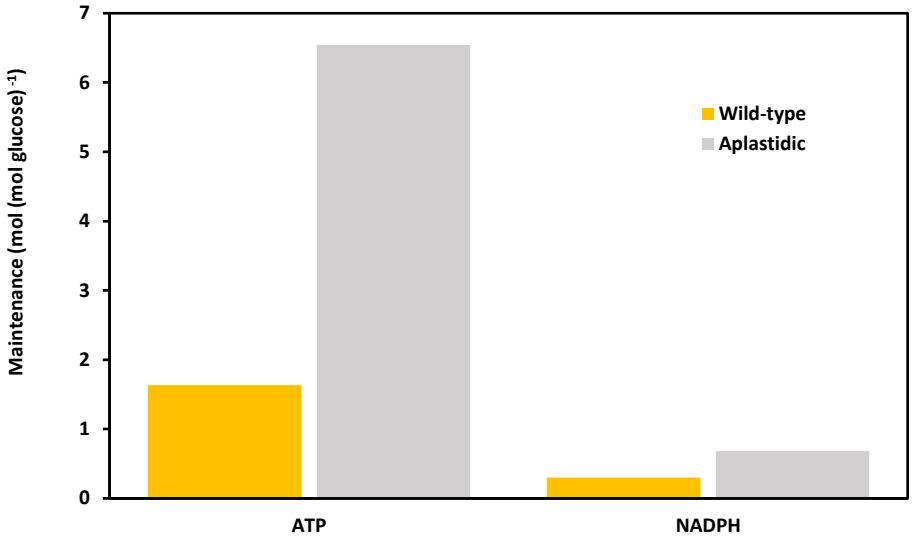


Figure 4.14. Estimated ATP and NADPH maintenance cost of wild-type and aplastidic *E. gracilis* grown heterotrophically on glucose. The maintenance was calculated by comparing reductants provisions estimated by ^{13}C MFA and FBA, the excess NADH and $FADH_2$ were assumed to be oxidised to generate ATP through the electron transport chain.

4.4.3 Accounting for maintenance in FBA of wild-type and aplastidic *E. gracilis*

The comparison of fluxes from FBA with the biomass constraint, FBA with maintenance included and ^{13}C MFA of the wild-type and aplastidic *E. gracilis* is shown in Figure 4.12 and 4.13 respectively. For both strains, FBA with only biomass constraint underestimated the fluxes through central metabolic pathways (glycolysis, the PPP, anaplerotic reactions and the TCA cycle) compared to ^{13}C MFA. Incorporating potential ATP and NADPH costs associated with maintenance into the FBA greatly improved the predicted

fluxes through glycolysis and the TCA cycle to match with ^{13}C MFA results, while fluxes through the PPP and anaplerotic reactions were still underestimated. The results demonstrate that incorporating the maintenance costs to the FBA resulted in substantial improvements in the prediction of fluxes through the central metabolic pathways.

The match between flux solutions of ^{13}C MFA and FBA of both strains were measured using RMSD. RMSDs between ^{13}C MFA and FBA with only the biomass constraint or FBA with maintenance accounted were calculated based on the flux solutions of free fluxes in ^{13}C MFA and the same reactions in FBA solution, the results are shown in Figure 4.15. RMSDs show that accounting for maintenance improved the flux prediction of FBA for both strains, but with different levels of improvement. In wild-type *E. gracilis*, the RMSD was lowered from 11.18 to 10.20 when maintenance was included, corresponding to about an 8.7% improvement. On the other hand, including maintenance greatly lowered the RMSD for aplastidic strain from 32.92 to 11.44 which is about 65.3% improved. In wild-type cells, the matching between the flux solution of ^{13}C MFA and FBA constrained with only biomass is relatively high compared to that in aplastidic cell, and also comparable to that of FBA with maintenance for aplastidic cell. This indicates that, relative to aplastidic cells, heterotrophic glucose metabolism in wild-type *E. gracilis* closely achieves the theoretical efficiency based on the stoichiometric model.

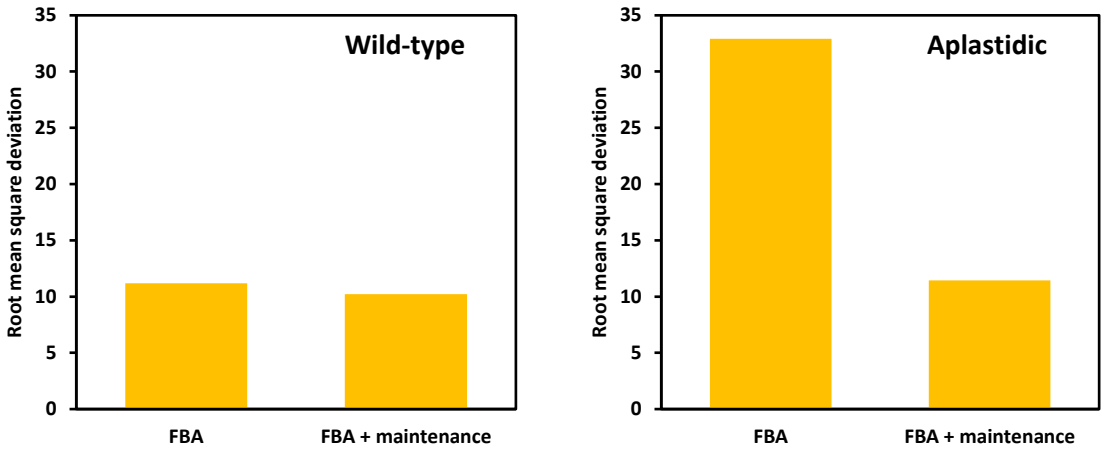


Figure 4.15. **RMSD comparing fluxes of FBA with biomass constraint and FBA with maintenance included for wild-type (left) and aplastidic (right) *E. gracilis* grown on glucose.** RMSD was calculated based on the free fluxes of ^{13}C MFA results and the same reactions in FBA solutions to assess the match between FBA and MFA fluxes.

4.4.4 Metabolic efficiency of *E. gracilis* strains grown on glucose based on CCE

With the CO_2 release of wild-type and aplastidic *E. gracilis* predicted by FBA, the maximal potential CCE of each strain was calculated as 91.76% and 91.55% respectively. As expected, CCEs calculated from FBA with the biomass constraint are higher compared to the CCE obtained from ^{13}C MFA for both strains, and accounting for maintenance costs lowered CCEs to 84.45% and 66.04% in wild-type and aplastidic cells respectively, which are comparable to those from ^{13}C MFA (83% and 64%).

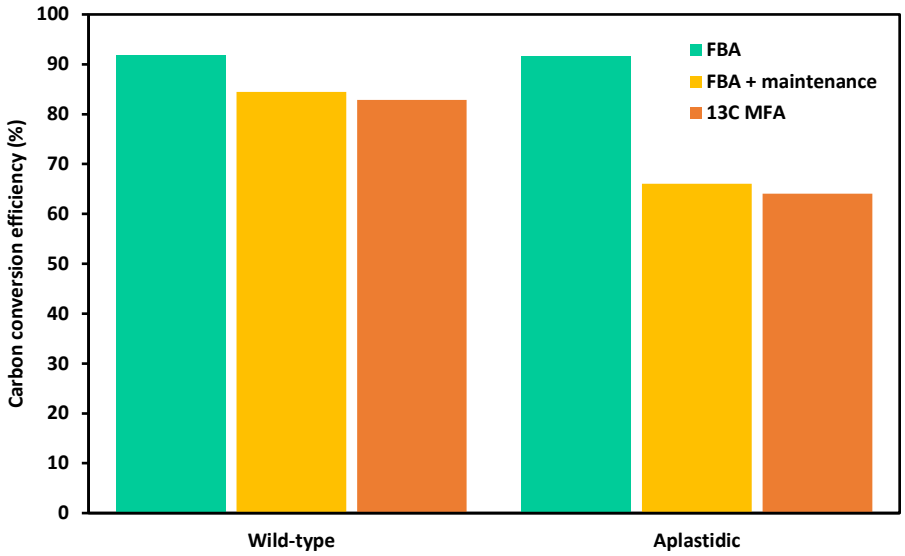


Figure 4.16. Comparison of CCE of wild-type and aplastidic *E. gracilis* grown heterotrophically on glucose by different approaches; FBA (green), FBA with maintenance accounted (yellow) and ^{13}C MFA (orange).

Metabolic efficiency based on CCE of each strain can be calculated as the proportion of CCE from ^{13}C MFA of the maximum potential CCE from FBA without maintenance included (Figure 4.17). Wild-type *E. gracilis* achieved about 90% of the theoretical metabolic efficiency based on CCE, while 70% was achieved in aplastidic cells. The result provides further demonstration that bleaching chloroplasts decreases the metabolic efficiency of glucose metabolism in *E. gracilis*.

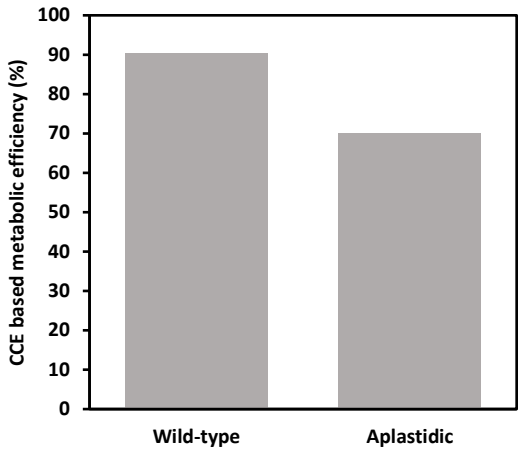


Figure 4.17. Metabolic efficiency based on carbon conversion efficiency of wild-type and aplastidic *E. gracilis* grown heterotrophically on glucose. The metabolic efficiency is calculated as the proportion of observed CCE (^{13}C MFA) of the estimated maximum potential CCE (FBA), indicating efficiency of carbon utilisation of the cells relative to the theoretical performance determined by the stoichiometry of the metabolic network.

4.5 Discussion

4.5.1 The limited contribution of plastidic metabolic processes, the oxPPP and the glyoxylate shunt in heterotrophic glucose metabolism in *E. gracilis*.

^{13}C MFA was successfully conducted to investigate the metabolic phenotype of glucose metabolism in *E. gracilis*. Flux maps of wild-type and aplastidic *E. gracilis* confirm that glucose was metabolised predominantly through glycolysis and the TCA cycle as is the norm in other eukaryotes (Hurlbert and Rittenberg, 1962).

In contrast to plants, the ^{13}C MFA flux map of wild-type *E. gracilis* demonstrated an insignificant role of the plastidic reactions in glucose metabolism. In plants, the oxPPP is located in the chloroplast, providing the majority of NADPH for the cells (Kruger and von

Schaewen, 2003; Schwender et al., 2003). In heterotrophic conditions, plastidic glycolysis in *Arabidopsis* and green algae operates in conjunction with the oxPPP in chloroplasts to produce the precursors needed for biomass production and to respond to the cofactors demand of the cell (Masakapalli, 2011, Masakapalli et al., 2014, Masakapalli et al., 2013, Masakapalli et al., 2010, Xiong et al., 2010). Unlike plants, glycolysis of *E. gracilis* mainly takes place in the cytosol with negligible fluxes through plastidic glycolysis.

The PPP is considered to be an important source of NADPH via its oxidative phase and to provide the precursors for nucleic acid biosynthesis through the reactions of the nonoxidative phase (Patra and Hay, 2014). ^{13}C MFA flux maps demonstrate that the PPP of *E. gracilis* operates at significantly lower rates, at about only 7% and 11% of glucose uptake in wild-type and aplastidic cells respectively, compared to those of 24% in *E. coli* (Gonzalez et al., 2017), 36-58% in *Saccharomyces cerevisiae* (Hayakawa et al., 2018), 33% in *Chlorella protothecoides* (Xiong et al., 2010) and 29% in *A. thaliana* (Masakapalli et al., 2013). It has been suggested that the operation of the PPP is regulated by the NADPH demand of these cells (Bowsher et al., 2007, Masakapalli et al., 2013), whist NADPH in *E. gracilis* is mainly produced through the $NADP^+$-dependent PDH (Nakazawa et al., 2017) at about 67% and 72% of the total NADPH production in wild-type and aplastidic strains respectively. This demonstrates the decreased contribution of the PPP in NADPH provision in *E. gracilis.*

The fluxes through glyoxylate shunt of *E. gracilis* was shown to be negligible during the growth on glucose in the dark. The glyoxylate shunt is known for its function in acetate and lipid metabolism (Chew et al., 2019, Ahn et al., 2016), and is required to enable conversion of acetate to hexose or pentose sugars via gluconeogenesis (Dolan and Welch,

2018). The glyoxylate shunt is reported to operate for this purpose in many groups of organisms, including eubacteria, plants, fungi and lower animals (Schnarrenberger and Martin, 2002). A study in *E. coli* and mycobacteria also reported the partitioning of flux between the TCA cycle and glyoxylate shunt during growth on acetate, but not when grown on glucose (Murima et al., 2016).

4.5.2 Chloroplast disruption increases cell maintenance costs and decreases the metabolic efficiency of *E. gracilis* during growth on glucose.

The almost identical growth characteristics of wild-type and aplastidic strains demonstrates that *E. gracilis* has the capacity to maintain growth and biomass production in the absence of chloroplasts, as the growth rate and glucose uptake rate are not significantly different, and the biomass composition of both wild-type and aplastidic strains are very similar. However, flux maps of wild-type and aplastidic *E. gracilis* demonstrate the changes in metabolic operation when the chloroplasts were eliminated, as disrupting chloroplasts decrease the rate of biomass synthesis and introduced the greater fluxes through the central metabolic pathways.

While the rates of cofactors consumption were approximately the same in both strains, cofactors were produced at higher rates in the aplastidic strain as a result of greater fluxes through glycolysis and the TCA cycle. Comparing the minimal potential cofactors production predicted by FBA with that required for the flux distributions defined by ^{13}C MFA showed that aplastidic *E. gracilis* requires about 4- and 2-times higher rates of ATP and NADPH for maintenance than wild-type cells. This demonstrates that, to retain the same growth rate, aplastidic *E. gracilis* requires much higher rates of cofactor production for cell

maintenance. These maintenance processes are likely to include turnover of membrane components and cycling of paramylon, preventing dissipation of ion gradients through membrane leakage and combatting reactive oxygen species (Seaver et al., 2012). In addition, a greater proportion of maintenance to the total cofactors production may indicate stress responses (Cheung et al., 2013).

The increase the demand for reductants in aplastidic *E. gracilis* leads to greater fluxes through central metabolic pathways and increased CO_2 production, causing the CCE to decrease. CCE provides an indication of metabolic efficiency as a basis for comparing the metabolising of carbon substrates into biomass (Chen and Shachar-Hill, 2012). FBA with only the biomass constraint predicts metabolic operation at maximal potential efficiency. In this study, the metabolic efficiency of each strain was calculated by expressing CCE calculated from ^{13}C MFA as a proportion of the maximum potential CCE obtained from FBA predictions. The results demonstrate that wild-type *E. gracilis* achieved 90% of the theoretical metabolic efficiency, while 70% was achieved in aplastidic cells, suggesting that glucose metabolism in wild-type *E. gracilis* achieves almost the maximum theoretical efficiency, whereas destroying chloroplasts of *E. gracilis* decreased the metabolic efficiency of glucose metabolism by about 20%. However, the differences between CCEs of the two strains do not correspond with the similar rates of glucose uptake and growth based on cell density. It is possible that, although the cell densities are not different, the biomass production per cell of the wild-type strain is greater than that of aplastic cells. In this thesis, the absolute values of fluxes are expressed in the unit of $\mu mol\ h^{-1}\ 10^{-6}$ cells, based on the assumption that *E. gracilis* cells grown on different substrates have a similar cell mass. Measuring the cell mass of *E. gracilis* grown on different

substrates would help confirm this assumption. However, growing *E. gracilis* heterotrophically on minimal media produces a very low biomass which makes it difficult to estimate the cell mass. To do this, larger growth facilities for heterotrophic *Euglena* culture are required.

The decreased metabolic efficiency (based on CCE and indicated by the maintenance in aplastidic *E. gracilis*) might be the result of the incomplete loss of chloroplast related components, that has been reported in aplastidic *Euglena* (Moriber et al., 1963). This may cause the increased accumulation of membrane components as indicated by the higher lipid content in aplastidic cells (Figure. 4.6). Furthermore, the turnover of the abortive plastidic nuclear-encoded proteins could potentially effectuate more maintenance cost in aplastidic cells. It is also possible that the loss of parts of the plastid genome affects the formation of non-functional enzymes for synthesis of chloroplast components (Shao et al., 2019).

4.6 Conclusion

The pathways involved in glucose metabolism in *E. gracilis* are well characterised and expected to be similar to those in other eukaryotes. Nevertheless, the metabolic phenotype of *E. gracilis* during glucose metabolism has not been previously investigated. This chapter describes the successful implementation of ^{13}C MFA to investigate the metabolic phenotypes of wild-type and aplastidic *E. gracilis* and examines the effect of chloroplast bleaching in *E. gracilis*. Flux maps of *E. gracilis* demonstrate the limited contribution of the chloroplast to glucose metabolism. The flux maps of both strains highlight the quantitative significance of $NADP^+$-dependent PDH in NADPH provision. The work also demonstrates that, even though

the aplastidic cells were able to maintain the same growth characteristics as wild-type cells, the metabolic phenotypes of the two strains differed substantially. Bleaching of chloroplast elevated fluxes of central pathways, responding to the increase in cofactor demand for cell maintenance. The decrease of CCE in aplastidic cells indicates that bleaching chloroplasts of *E. gracilis* decreased the metabolic efficiency during glucose metabolism in *E. gracilis*.

Chapter 5 | Metabolic phenotypes of *E. gracilis* grown on C2 and C3 substrates

5.1 Introduction

The carbon sources used to grow *Euglena* have marked effects on various physiological characteristics including, growth rate, respiration, energy metabolism, circadian rhythm, stress responses and biosynthesis (Buetow, 1989). When other conditions are constant, the growth rate of *Euglena* can be varied depending on the source of carbon. For example, acetate and succinate supported growth rates of *Euglena* approximately twice the growth rates when grown on malate and fumarate (Wilson et al., 1959). In a comparison of six different carbon sources, the growth rate of non-photosynthetic WZSL mutant *E. gracilis* decreased in the order: acetate > glucose > glutamate > succinate > pyruvate > ethanol (Kusmic et al., 1998). In addition, growth media with mixed carbon sources gave a higher growth rate compared to media with a single carbon source. It was suggested that a diversity of compounds in the complex source could decrease the requirement for some metabolic reaction steps (Blum and Wittels, 1968).

Carbon sources have a marked effect on the relative proportions of compounds produced and accumulated in *Euglena*. For example, the paramylon content of *E. gracilis* grown on glucose exceeded that when grown on pyruvate which, in turn, was higher than that grown on successively, succinate, acetate, glutamate and ethanol (Barsanti et al., 2000).

Similarly, growth of *Euglena* on glucose gave the highest yield of tocopherol (vitamin E) in comparison to ethanol, glycerol and galactose (Ogbonna et al., 1998). Although changes in biomass composition and growth of *Euglena* with different carbon sources have been intensively investigated, the effect of carbon source on the relative activities of different metabolic pathways that generate these products and support the growth and survival of *Euglena* is still poorly understood.

5.1.1 Acetate and ethanol metabolism in *E. gracilis*

The ability to utilise acetate and ethanol as growth substrates by *Euglena* is well documented (Barsanti et al., 2000, Blum and Wittels, 1968, Buetow, 1989, Feldman and Bruce, 1972, Kusmic et al., 1998, Ogbonna et al., 1998). The glyoxylate shunt has an essential role in assimilating C2 compounds in a diverse range of organisms including bacteria, fungi, algae and plants. The shunt is involved in provision of precursors for anaplerosis and gluconeogenesis and allows the cell to use other respiratory substrates when sugars are not available (Nakazawa, 2017). The activities of isocitrate lyase and malate synthase, enzymes unique to the glyoxylate shunt, have been detected in the mitochondrial cell fraction from *E. gracilis* grown on ethanol (Ono et al., 2003).

Ethanol can enter the cell by simple diffusion, and is first oxidized to acetaldehyde by ADH. Acetaldehyde is subsequently oxidised by acetaldehyde dehydrogenase (ALDH) to produce acetate. Both enzymes are found in the mitochondria of *E. gracilis* (RodrÍGuez-Zavala et al., 2006, Yoval-Sánchez et al., 2011, Ono et al., 1995). The activities of these enzymes are only detected in the presence of ethanol in *Euglena* culture (RodrÍGuez-Zavala et al., 2006). The acetate taken up by simple diffusion, or the ALDH-generated acetate, can be converted

to acetyl-CoA by acetyl-CoA synthetase located in *Euglena* mitochondria (Graves Jr and Becker, 1974), and then metabolised through the TCA cycle or the glyoxylate shunt.

The glyoxylate shunt bypasses the decarboxylation steps of isocitrate dehydrogenase and α-ketoglutarate dehydrogenase in the TCA cycle, and converts two molecules of acetyl-CoA to one molecule of malate. Malate is further converted by MDH to oxaloacetate (OAA) which can be converted to PEP by phosphoenolpyruvate carboxykinase (PEPCK). The generated PEP is subsequently used to produce carbohydrates via gluconeogenesis (Nakazawa, 2017). The activity of PEPCK in *Euglena* was reported to be cytosolic and could not be detected when cells were grown in autotrophic conditions (Miyatake et al., 1984b). Furthermore, the activity of PEPCK was detected in *E. gracilis* cultured with acetate or ethanol, but not with glucose (Peak et al., 1972), indicating the significance of PEPCK in gluconeogenesis.

As described above, *Euglena* metabolises acetate and ethanol through the same metabolic pathways involving the same enzymes, with the exception that both ADH and ALDH are additionally needed for metabolism of ethanol. Nevertheless, the metabolism of these two substrates has not been systematically studied. Investigating the metabolic phenotypes of *E. gracilis* during growth on acetate or ethanol can provide an understanding of how the cells metabolically adjust growth on different substrates with the same carbon composition but different chemical oxidation states.

5.1.2 Pyruvate metabolism

Pyruvate has been studied as one of the growth substrates of *Euglena* for decades (Barsanti et al., 2000, Blum and Wittels, 1968, Buetow, 1989, Feldman and Bruce, 1972, Kusmic et al., 1998, Ogbonna et al., 1998). As the end-product of glycolysis, pyruvate is transported into mitochondria where it is further converted to acetyl-CoA by PDH and subsequently metabolised by the TCA cycle to generate reduced cofactors and ultimately ATP or precursors for biomass production (Gray et al., 2014). Pyruvate can also be converted by pyruvate carboxylase (PC) to OAA which can be used in the TCA cycle, as a precursor for biosynthesis or channelled to gluconeogenesis. In *Euglena*, the activity of PC was detected in the cytosol of the cells grown under heterotrophic conditions supplemented with glucose, but not acetate (Yokota et al., 1982). Even though the enzymes/pathways contributing to the metabolism of pyruvate are well characterised enzymatically in *Euglena*, the metabolic phenotype during the metabolism of pyruvate of *E. gracilis* has never been investigated in detail. Performing flux analysis can provide insights into the metabolism of pyruvate in *E. gracilis*.

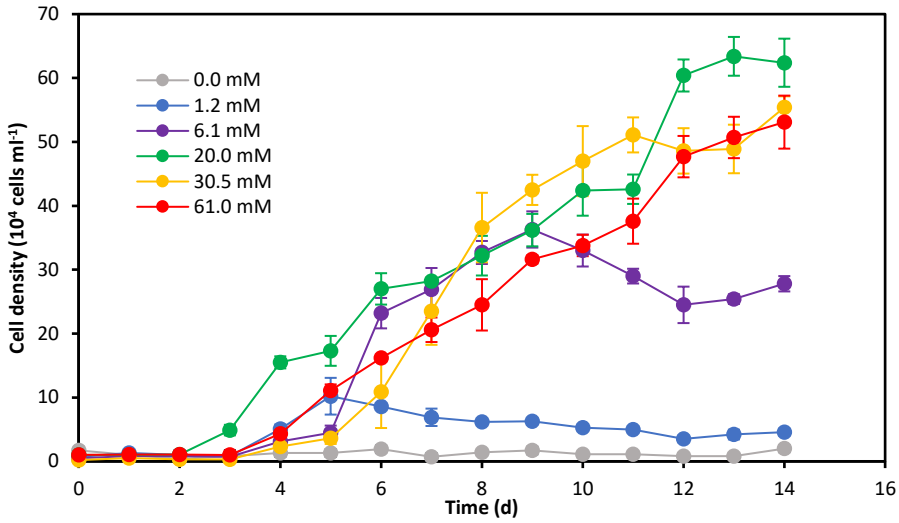
5.1.3 Objectives

This chapter aims to quantify the metabolism of various carbon sources to investigate the metabolic flexibility of the network of *E. gracilis.* In this study, ^{13}C MFA and FBA were exploited to compare the metabolism of ethanol, acetate and pyruvate as growth substrates for *E. gracilis*. Flux maps produced under these conditions will provide insights on how *Euglena* cells respond metabolically to different carbon substrates, which enter the network at similar points but provide different redox states to the network.

5.2 Physiological characterisation of *E. gracilis* grown on acetate, ethanol and pyruvate

5.2.1 Growth of heterotrophically *E. gracilis* cultures on acetate

To establish the optimal acetate concertation for ^{13}C MFA, *E. gracilis* was cultivated on 30 ml of AF6 medium supplemented with up to 61.0 mM acetate. The cultures were grown in the standard conditions previously described in Chapter 2. The cell density of each culture was counted to monitor the growth of *E. gracilis* on different acetate concentrations (Figure 5.1). There was no growth of *E. gracilis* in the dark with no acetate in the medium. *E. gracilis* cultures with 20.0 mM acetate showed the highest peak cell density (see Table in Figure 5.1). The durations of lag phase of *E. gracilis* on acetate varied, with the shortest lag phase of two days observed in the culture grown on 20.0 mM acetate, and three or five days for other cultures. Thus, the concentration of 20.0 mM acetate was chosen for the further analyses.



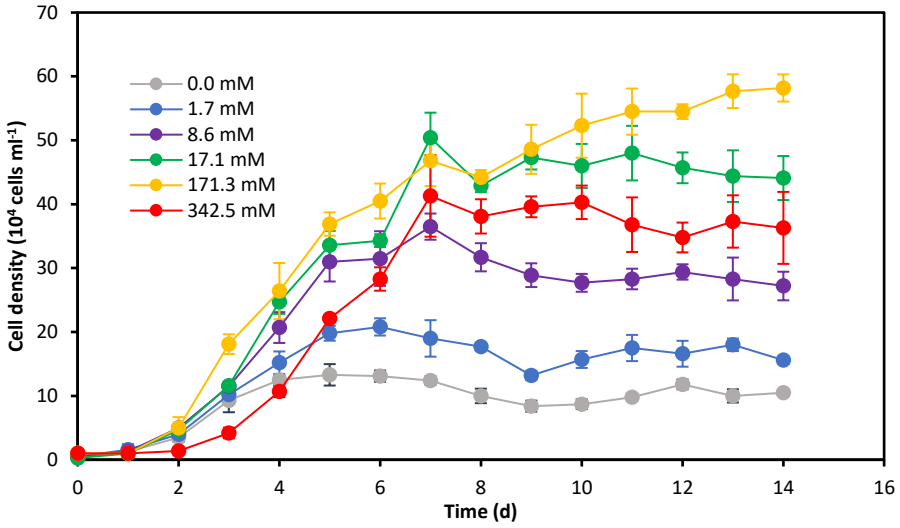
	[Acetate] (mM)					
	0.0	1.2	6.1	20.0	30.5	61.0
Maximum cell density (10^4 cells ml^{-1})	1.9 ± 0.6	10.2 ± 2.9	36.3 ± 2.8	63.4 ± 3.0	55.4 ± 1.7	53.1 ± 4.1
Lag phase (d)	-	3	5	2	5	3

Figure 5.1. Acetate concentration dependence of heterotrophic *E. gracilis* growth. The cell density values are the mean ± SD of four independent biological replicates, monitored daily by counting cell number using haemocytometer. Lag phase was estimated by the period before the extrapolated exponential phase.

5.2.2 Growth of heterotrophically *E. gracilis* cultures on ethanol

E. gracilis was grown on 30 ml of AF6 medium supplemented with up to 342.5 mM ethanol under the standard growth conditions. The cell density for each culture on ethanol was measured until reaching stationary growth phase (~14 days) to assess the optimal ethanol concentration for *E. gracilis* cultivation (Figure 5.2). The culture supplemented with 171.3 mM ethanol provided the highest cell density with a peak of about 58.2 ± 2.1 × 10^4 cells ml^{-1} and one day of lag phase. Apart

from the culture with 342.5 mM ethanol that had two days of lag phase, one day of lag phase was observed at all other ethanol concentrations studied. Although the culture on 171.3 mM ethanol showed the fastest cell growth, the growth pattern and cell densities throughout the growth period were essentially equivalent to the culture with 17.1 mM ethanol. Considering the expense of ^{13}C-labeled ethanol used in isotopic labelling experiments for ^{13}C MFA, the ethanol concentration of 17.1 mM was chosen as the experimental concentration.

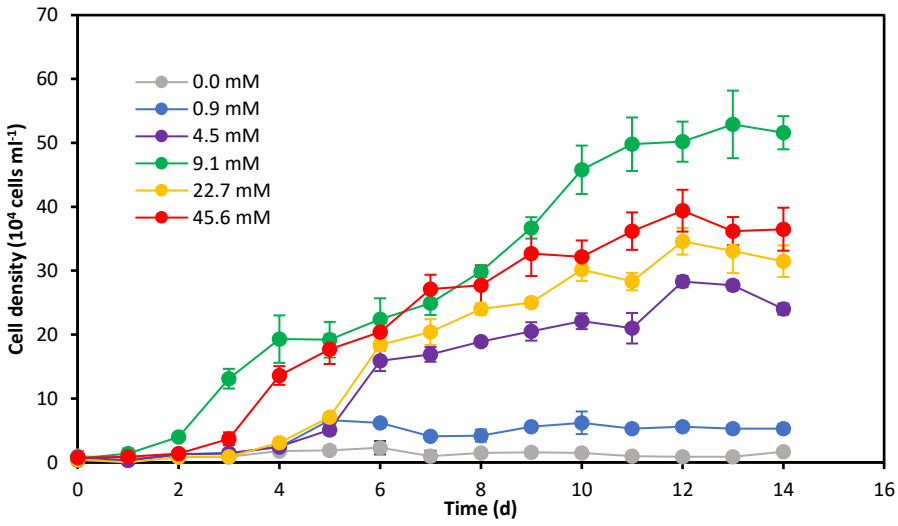


	[Ethanol] (mM)					
	0.0	1.7	8.6	17.1	171.3	342.5
Maximum cell density (10^4 cells ml^{-1})	13.3 ± 1.6	20.8 ± 1.3	36.5 ± 2.0	50.4 ± 3.9	58.2 ± 2.1	41.3 ± 6.4
Lag phase (d)	1	1	1	1	1	2

Figure 5.2. Ethanol concentration dependence of heterotrophic *E. gracilis* growth. The cell density values are the mean ± SD of four independent biological replicates, monitored daily by counting cell number using haemocytometer. Lag phase was estimated by the period before the extrapolated exponential phase.

5.2.3 Growth of heterotrophically *E. gracilis* on pyruvate

Under the standard growth conditions described in Chapter 2, *E. gracilis* was grown on AF6 medium supplemented with up to 45.6 mM pyruvate. Cell densities of the cultures were monitored for 14 days (Figure 5.3). The culture supplemented with 9.1 mM of pyruvate showed the highest cell density at peak of 52.9 ± 5.3 × 10^4 cells ml^{-1} with only one day of lag phase. The duration of the lag phase of these cultures was two days for 45.6 mM, and three days for 0.9, 4.5 and 22.7 mM pyruvate. As a result, a pyruvate concentration of 9.1 mM was chosen for subsequent analyses.



	[Pyruvate] (mM)					
	0.0	0.9	4.5	9.1	22.7	45.6
Maximum cell density (10^4 cells ml^{-1})	1.7 ± 0.2	6.6 ± 1.1	28.3 ± 0.9	52.9 ± 5.3	34.6 ± 2.1	39.4 ± 3.3
Lag phase (d)	-	3	3	1	3	2

Figure 5.3. Pyruvate concentration dependence of heterotrophic *E. gracilis* growth. The cell density values are the mean ± SD of four independent biological replicates, monitored daily

by counting cell number using haemocytometer. Lag phase was estimated by the period before the extrapolated exponential phase.

5.2.4 Comparison of growth characteristics of *E. gracilis* grown on acetate, ethanol and pyruvate

Specific growth rates

The specific growth rates during the exponential growth phase of *E. gracilis* were calculated from the cell densities of day 2 to day 6 of the cultures using an established approach (Levasseur et al., 1993). The growth rates of cells on acetate, ethanol or pyruvate are substantially lower than those observed for cells grown on glucose (Figure 5.4). Among the three 2-3 C substrates, the growth rate of the ethanol cultures was significantly higher than the others ($p < 0.05$) at 0.44 ± 0.00 d^{-1}, whilst the specific growth rates of acetate and pyruvate cultures were not significantly different ($p > 0.05$) at 0.24 ± 0.02 and 0.23 ± 0.02 d^{-1}, respectively.

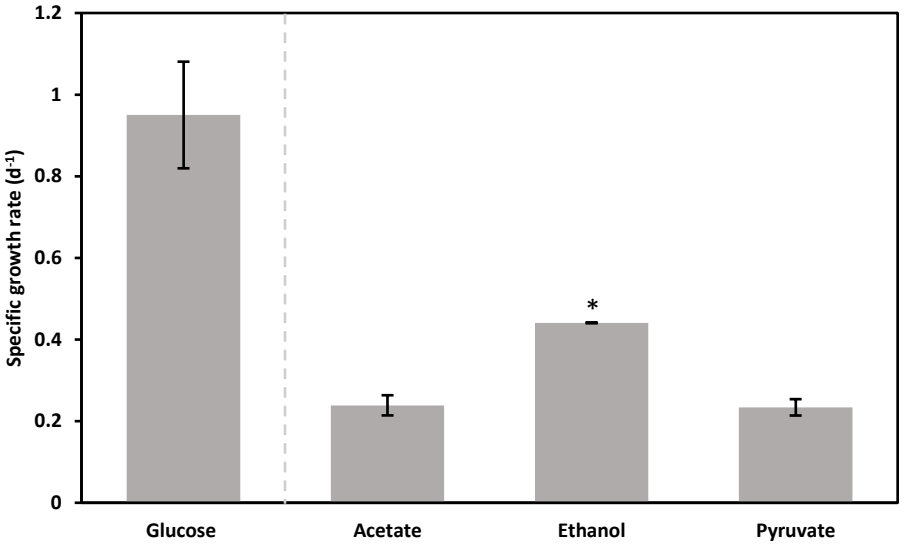


Figure 5.4. **Specific growth rates during exponential phase of *E. gracilis* grown heterotrophically on glucose (Chapter 4), acetate (20 mM), ethanol (17.1 mM) and pyruvate (9.1 mM).** The data shown are the means ± SD of four biological replicates. The growth rate was calculated from the cell densities of day 2 to day 6 of the cultures. The growth rates of cells grown on acetate and pyruvate were not significantly different (t-test, $p > 0.05$), whist growth rate of ethanol grown cells was significantly higher, as indicated by * (Tukey's HSD, $p < 0.05$).

Substrate consumption rate

Substrate concentrations in the growth media of *E. gracilis* cultures grown on acetate, ethanol or pyruvate were monitored every two days for 14 days (Figure 5.5). The concentrations of acetate and pyruvate were measured using GC-MS, and a colorimetric assay was used for ethanol quantification as described in Chapter 2. An initial lag phase in acetate consumption was observed from day 0 to day 2 of cultivation, corresponding to the lag phase of cell growth. After day 2, the concentration of acetate in the growth medium constantly decreased, with a total consumption of 42% of the provided acetate. The concentration of ethanol in the growth medium rapidly decreased from day 2 to day 8 when the culture was in an exponential growth phase. Over this period, about 52% of the supplemented ethanol was consumed. A lag phase and constant depletion of pyruvate were also observed, with a total pyruvate consumption of approximately 35%.

The consumption rate of each substrate was determined using a procedure based on the average method (Görgens et al., 2005). The rates of substrates consumption (Figure 5.6) were calculated based on the substrate concentration and the cell densities from day 2 to

day 6 of the cultivation. The consumption rate of ethanol by *E. gracilis* cells was significantly higher than that of pyruvate ($p < 0.05$) at 0.26 ± 0.06 and 0.11 ± 0.02 µmol h^{-1} 10^{-6} cells, respectively. The rate of acetate consumption was 0.14 ± 0.07 µmol h^{-1} 10^{-6} cells which is intermediate and not significantly different from that of either ethanol or pyruvate ($p > 0.05$). These substrate consumption rates are substantially lower than the glucose consumption rate (8.15 ± 0.63 µmol h^{-1} 10^{-6} cells, data shown in Chapter 4).

For all three substrates, the late-stage growth of heterotrophic *E. gracilis* appeared to continue inconsistently to the plateau of substrate depletion (figure 5.5). This could be a result of the utilisation of storage compound (paramylon) for cell propagation, resulted in the extendedly increase in cell density. In this thesis, growth of *E. gracilis* is measured by cell density based on the assumption that *E. gracilis* cells grown on different substrates have the consistent cell mass throughout the growth period. It is possible that the biomass production per cell changes at the late growth stage. Measuring cell size or mass across the growth period could help confirm this speculation.

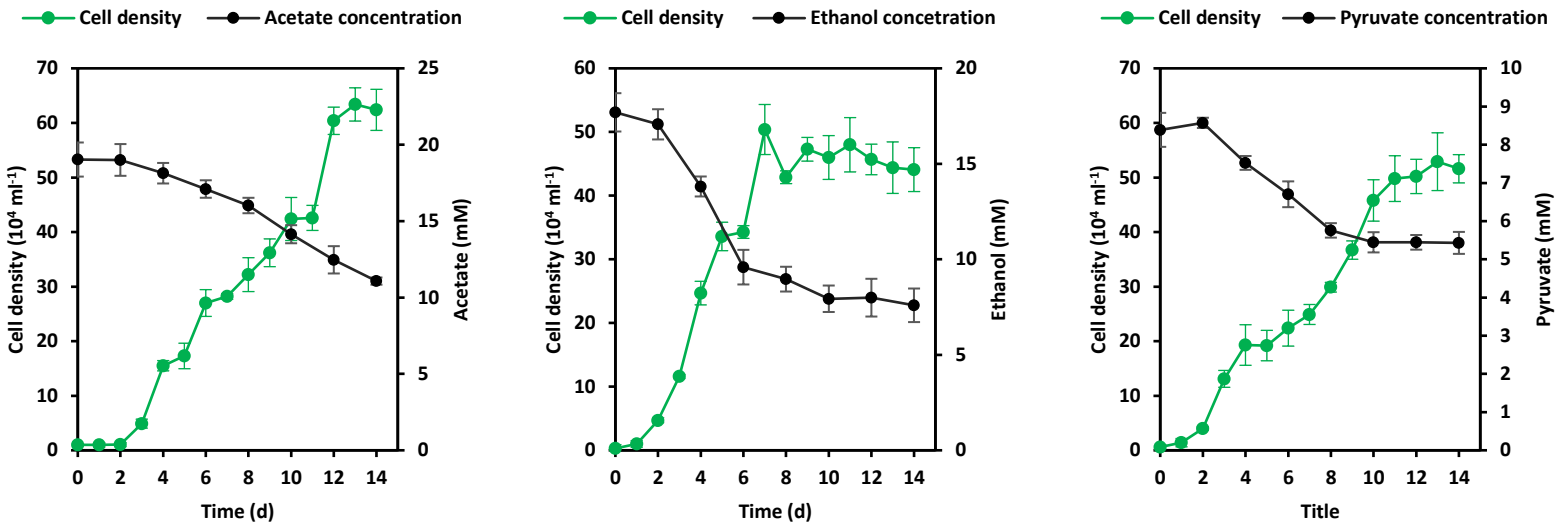


Figure 5.5. Substrate depletion in the culture media during heterotrophic growth (black line) of *E. gracilis* on acetate (left), ethanol (middle) and pyruvate (right). The substrate concentrations in growth media were determined every other day from day 0 to day 14. The substrate concentrations are presented with the growth of *E. gracilis* (green line) grown on the substrates. The data presented are means ± SD of four independent biological replicates.

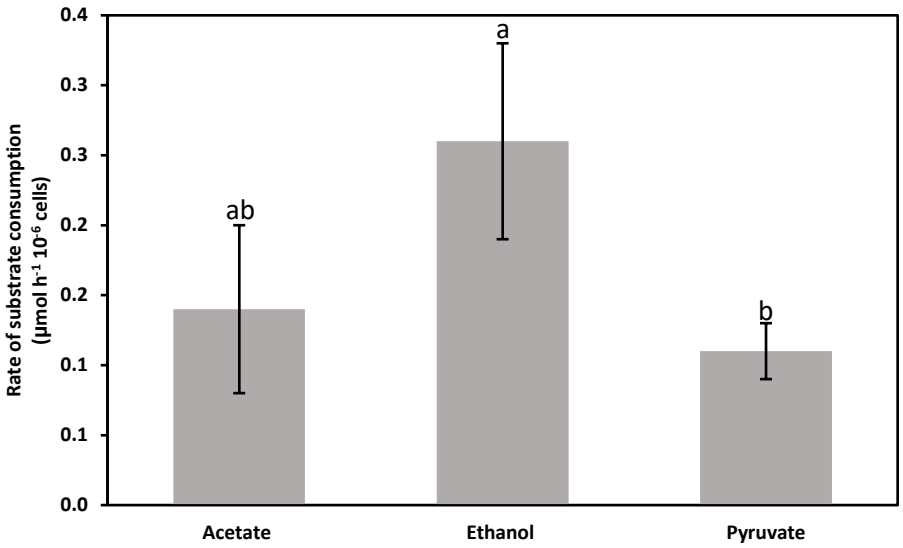


Figure 5.6 Rates of substrate consumption for *E. gracilis* grown on acetate, ethanol and pyruvate during exponential growth phase. The consumption rate was determined using the average method (Görgens et al., 2005) based on the substrate concentration and the cell density from day 2 to day 6 of the cultivation. Data are the means ± SD of four biological replicates. Rates with different letters are significantly different (Tukey's HSD, $p < 0.05$).

5.2.5 Biomass composition of *E. gracilis* grown heterotrophically on acetate, ethanol and pyruvate

The four main components of biomass, comprising protein, insoluble carbohydrate, soluble carbohydrate and lipid, of *E. gracilis* grown heterotrophically on acetate, ethanol and pyruvate were determined (Table 5.1). The analysed biomass components of acetate, ethanol and pyruvate constitute about 85.0%, 92.6% and 90.5% of dry cell mass, respectively. Approximately 7.4 – 15.0 % of the biomass is not identified in this study (Figure 5.7), which might include nucleic acids, other soluble metabolites, such as organic acids, lactic acid, carotenoids and minerals. The nucleic acids content was incorporated into the flux analysis

as described in Chapter 3, and the reminding unidentified biomass components are unlikely to have a substantial impact on the flux maps, which are ultimately derived from the biomass compositions reported in Table 5.1.

The protein content of *Euglena* cells grown on acetate (5.1 ± 0.7%) was significantly lower ($p < 0.05$) than that of cultures grown on ethanol and pyruvate, at 9.1 ± 1.3% and 10.7 ± 0.9%, respectively. In contrast, the biomass proportion of insoluble carbohydrate, total carbohydrate and lipid were not significantly different between the three substrates ($p > 0.05$). *E. gracilis* grown on acetate, ethanol and pyruvate accumulated insoluble carbohydrate as the main biomass component at about 72.9 – 75.4 % of biomass. Insoluble carbohydrate content was subtracted from total carbohydrate to estimate the proportion of soluble carbohydrate. The total carbohydrate contents of *E. gracilis* on the substrates were not much higher than the insoluble carbohydrate content, which results in a soluble carbohydrate content of only about 1.6 - 3% of total biomass, in contrast to the 35 % when grown on glucose. Total lipid content of *E. gracilis* grown on acetate, ethanol and pyruvate was about 4.9 - 5.3% of dry weight biomass.

Table 5.1. Biomass composition of *E. gracilis* grown heterotrophically on medium containing acetate, ethanol and pyruvate

Biomass component	Biomass composition (% DW)		
	Acetate	Ethanol	Pyruvate
Protein	5.1 ± 0.7*	9.1 ± 1.3	10.7 ± 0.9
Insoluble carbohydrate	72.9 ± 0.6	75.4 ± 0.3	73.3 ± 2.3
Total carbohydrate	74.6 ± 2.3	78.4 ± 6.3	74.9 ± 4.5
Lipid	5.3 ± 0.3	5.1 ± 0.2	4.9 ± 0.2

Biomass compositions of *E. gracilis* strains are the mean ± SD of four biological replicates.
* indicates the value that is significantly different (Tukey's HSD, $p < 0.05$)

Overall, *E. gracilis* grown on acetate, pyruvate and ethanol was made up of similar biomass compositions. However, the relative abundances of some biomass components in cultures grown on these substrates are substantially different from those of a glucose-grown culture (Figure 5.7). *E. gracilis* grown on C2 or C3 compounds produced more lipid and less protein compared to cells grown on glucose. Although the total carbohydrate contents were similar (about 72%-75%) on all substrates, *E. gracilis* grown on glucose accumulated less insoluble carbohydrate and contained a higher proportion of soluble carbohydrate. This could be a result of the higher osmotic pressure of a growth medium containing 83.3 mM glucose compared to acetate, ethanol and pyruvate media with substrates at concentrations of 20.0, 17.1 and 9.1 mM, respectively.

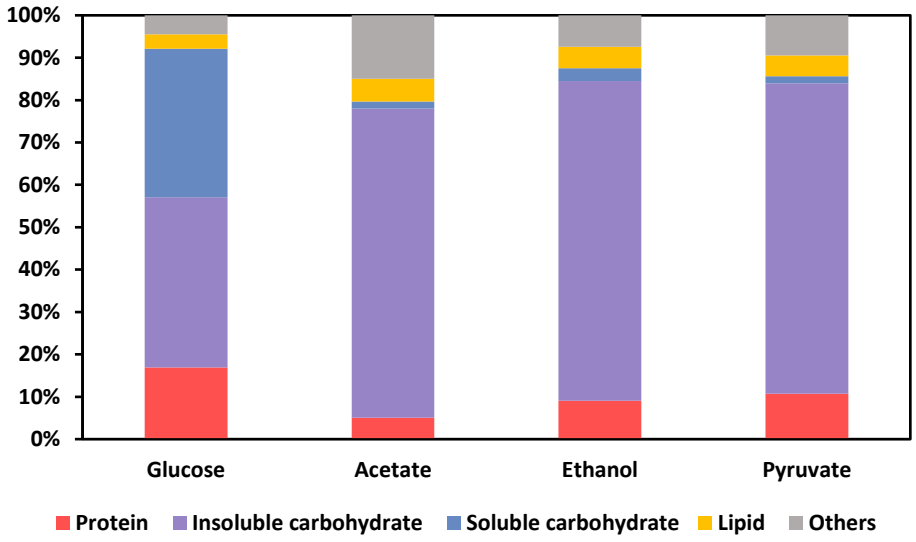


Figure 5.7. Biomass composition of *E. gracilis* grown heterotrophically on glucose (data from Chapter 4), acetate, ethanol and pyruvate.

5.3 ^{13}C MFA of *E. gracilis* grown on acetate, ethanol and pyruvate

5.3.1 Flux map of *E. gracilis* grown on acetate

^{13}C MFA was performed on *E. gracilis* cultures grown on AF6 medium supplemented with three different ^{13}C labelled sources of acetate: [1-^{13}C]acetate, [2-^{13}C]acetate or 20% [$^{13}C_2$]acetate. The flux analysis was conducted using the MIDs of amino acids obtained from the protein hydrolysate of *Euglena* cells. The SSR of this analysis is 553.9, which is lower than the statistically accepted SSR range of 609.6 - 754.2 with 680 degrees of freedom at the 95% confidence level. The flux map of *E. gracilis* grown on acetate (Figure 5.8) shows that of the acetate taken up and converted to acetyl-CoA (0.14 µmol h^{-1} 10^{-6} cells) about 73% entered the TCA cycle and 23% was used by the glyoxylate shunt with the remaining 4% diverted to produce biomass. 23% acetate uptake equivalents of isocitrate was also channelled to the

glyoxylate shunt, resulting in about 50 mol isocitrate (100 mol acetate)$^{-1}$ [equivalent to 0.07 μmol h^{-1} 10^{-6} cells] remaining for further metabolism by isocitrate dehydrogenase.

About 23 mol (100 mol acetate)$^{-1}$ [0.03 μmol h^{-1} 10^{-6} cells] of malate and succinate from the glyoxylate shunt were reintroduced to the TCA cycle, causing the fluxes from succinate to malate to be about 73 mol (100 mol acetate)$^{-1}$ [0.10 μmol h^{-1} 10^{-6} cells], and the flux of MDH to be 96 mol (100 mol acetate)$^{-1}$ [0.13 μmol h^{-1} 10^{-6} cells]. PEPCK converted about 23 mol OAA (100 mol acetate)$^{-1}$ [0.03 μmol h^{-1} 10^{-6} cells] to PEP, which was subsequently directed through gluconeogenesis to produce about 10 mol G6P (100 mol acetate)$^{-1}$ [0.01 μmol h^{-1} 10^{-6} cells] for biomass production. The malic enzyme, PDH, the cytosolic PPP and the plastid pathways (the nonoxPPP and glycolysis) carried negligibly small fluxes in the metabolism of acetate in *E. gracilis*.

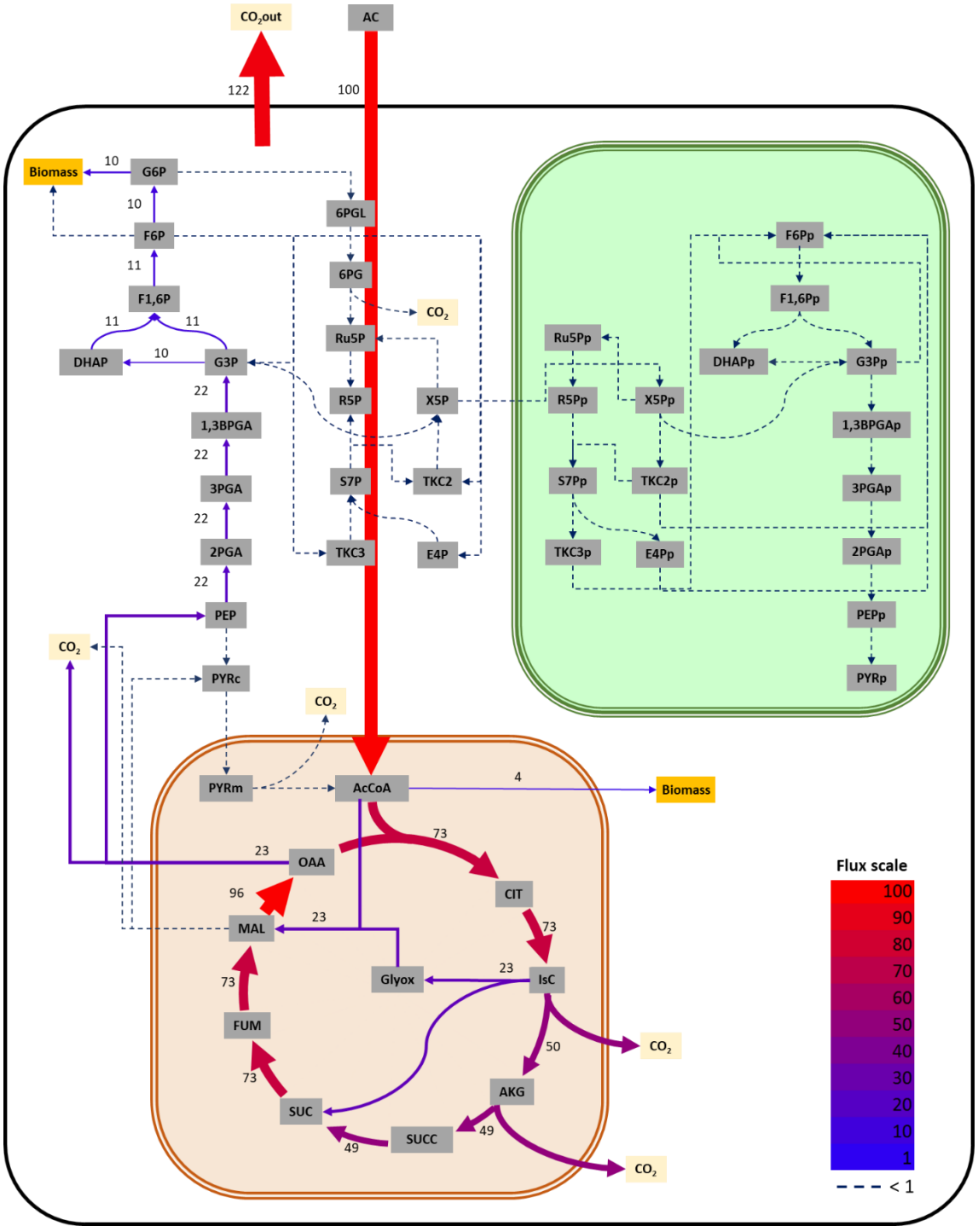


Figure 5.8. Metabolic flux map of *E. gracilis* grown heterotrophically on acetate. The thickness of arrows and the numbers represent mol of fluxes per 100 mol of acetate input.

5.3.2 Flux map of *E. gracilis* grown on ethanol

^{13}C MFA was performed in *E. gracilis* cultures grown on AF6 medium supplied with [1-^{13}C]ethanol, [2-^{13}C]ethanol or 25% [$^{13}C_2$]ethanol in a similar way to that using cells grown on acetate. The MIDs of amino acids obtained from the protein hydrolysate of the cell extracts were used for the flux analysis. At the confidence level of 95%, the analysis provided a statistically accepted fit with an SSR of 617.7, compared to the expected SSR range of 610.6 - 755.2 with 681 degrees of freedom at the 95% confidence level.

The flux results of ^{13}C MFA are visualised as a flux map (Figure 5.9). 100 mol of ethanol intake (0.26 μmol h^{-1} 10^{-6} cells) was converted to acetate by ADH and all the acetate generated was subsequently reduced to acetyl-CoA before entering the central metabolic pathways. Acetyl-CoA equivalent to 5% of ethanol intake was used directly to produce biomass, whilst about 59% entered the TCA cycle, with 40% subsequently diverted through the glyoxylate shunt. An amount equivalent to 18% of ethanol taken up of isocitrate was metabolised by isocitrate dehydrogenase. Malic enzyme converted 38 mol (100 mol ethanol)$^{-1}$ [0.10 μmol h^{-1} 10^{-6} cells] of pyruvate to malate, resulting in an increased flux of MDH, at about 137 mol (100 mol ethanol)$^{-1}$ [0.36 μmol h^{-1} 10^{-6} cells]. 73 mol OAA (100 mol ethanol)$^{-1}$ [0.19 μmol h^{-1} 10^{-6} cells] was withdrawn from the TCA cycle to produce PEP by PEPCK, while about 38 mol PEP (100 mol ethanol)$^{-1}$ was directed to gluconeogenesis to produce 16 mol of G6P (100 mol ethanol)$^{-1}$ [0.04 μmol h^{-1} 10^{-6} cells] for biomass production. The remaining 35 mol (100 mol ethanol)$^{-1}$ of PEP generated by PEPCK was used to produce pyruvate that was subsequently metabolised to malate by malic enzyme. As with acetate grown cells, PDH, the

cytosolic PPP and the plastid pathways (the nonoxPPP and glycolysis) carried negligibly small fluxes during growth on ethanol.

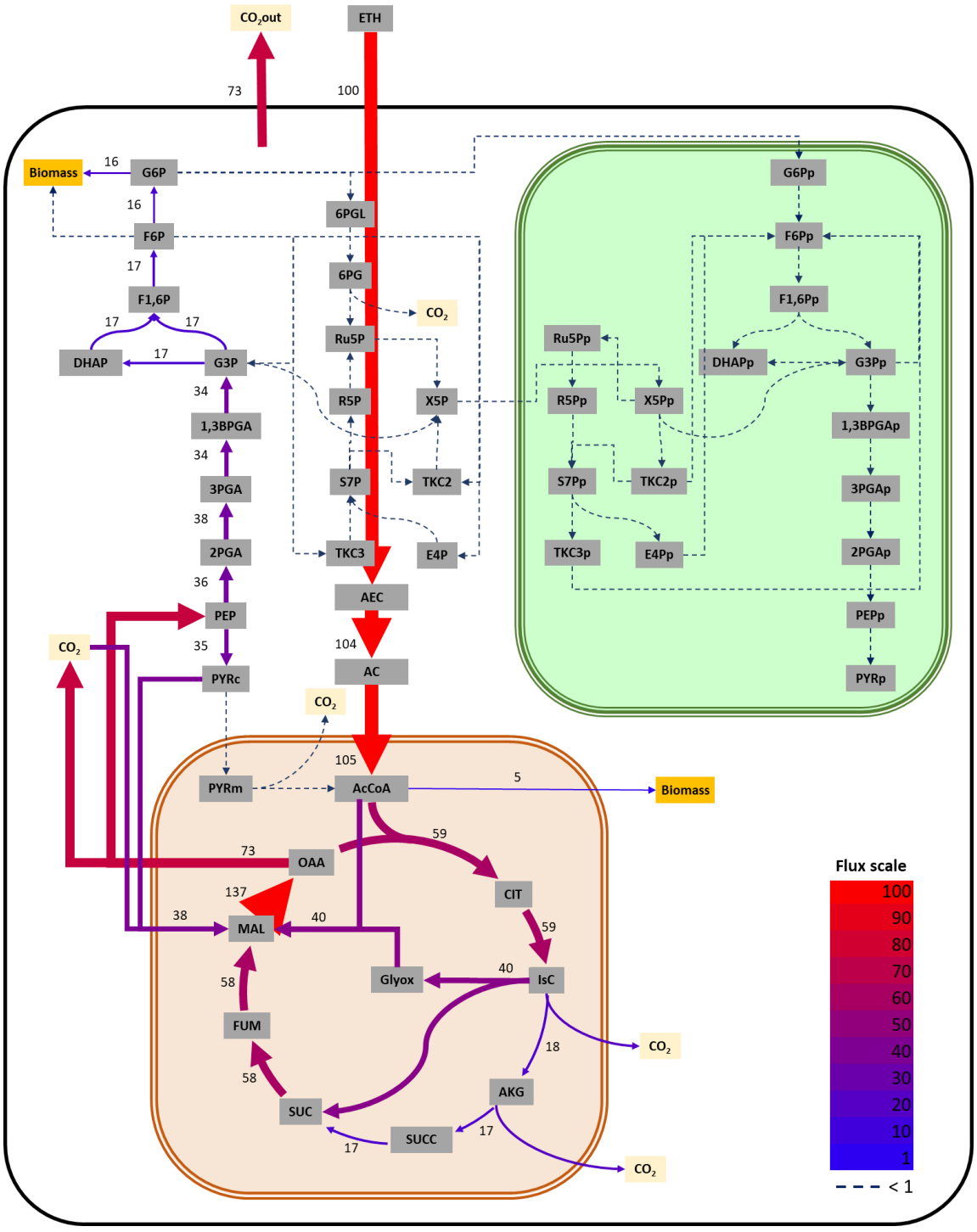


Figure 5.9. Metabolic flux map of *E. gracilis* grown heterotrophically on ethanol. The thickness of arrows and the numbers represent mol of fluxes per 100 mol of ethanol intake.

5.3.3 Flux map of *E. gracilis* grown on pyruvate

Three different ^{13}C labelled sources of pyruvate, consisting of [1-^{13}C]pyruvate, [3-^{13}C]pyruvate and 20% [$^{13}C_3$]pyruvate, were used to conduct ^{13}C MFA of *E. gracilis* grown on pyruvate. The analysis provided a statistically acceptable SSR of 702.9, compared to the expected SSR of 592.6 - 735.2 at the confidence level of 95% and 662 degrees of freedom. The flux values were visualised as a flux map (Figure 5.10).

83% of the pyruvate taken up, 0.11 µmol pyruvate h^{-1} 10^{-6} cells, was transported to mitochondria and converted to acetyl-CoA by PDH, whilst about 29% was metabolised by PC located in the cytosol to produce OAA. 67% equivalent of the pyruvate intake of acetyl-CoA entered the TCA cycle, with 15 mol isocitrate (100 mol pyruvate)$^{-1}$ [0.02 µmol h^{-1} 10^{-6} cells] diverted through the glyoxylate shunt, resulting in the reduced fluxes from isocitrate to succinate of the TCA cycle of about 52 mol (100 mol pyruvate)$^{-1}$. Malic enzyme recycled 11 mol malate (100 mol pyruvate)$^{-1}$ [0.01 µmol h^{-1} 10^{-6} cells] to pyruvate. About 28 mol OAA (100 mol pyruvate)$^{-1}$ [0.03 µmol h^{-1} 10^{-6} cells] was converted to PEP by PEPCK and the generated PEP was then further metabolised via gluconeogenesis to produce about 12 mol G6P (100 mol pyruvate)$^{-1}$ [0.01 µmol h^{-1} 10^{-6} cells] for biomass production. As found for ethanol and acetate cultures, the cytosolic PPP and the plastid pathways contributed negligible flux to the metabolism of pyruvate.

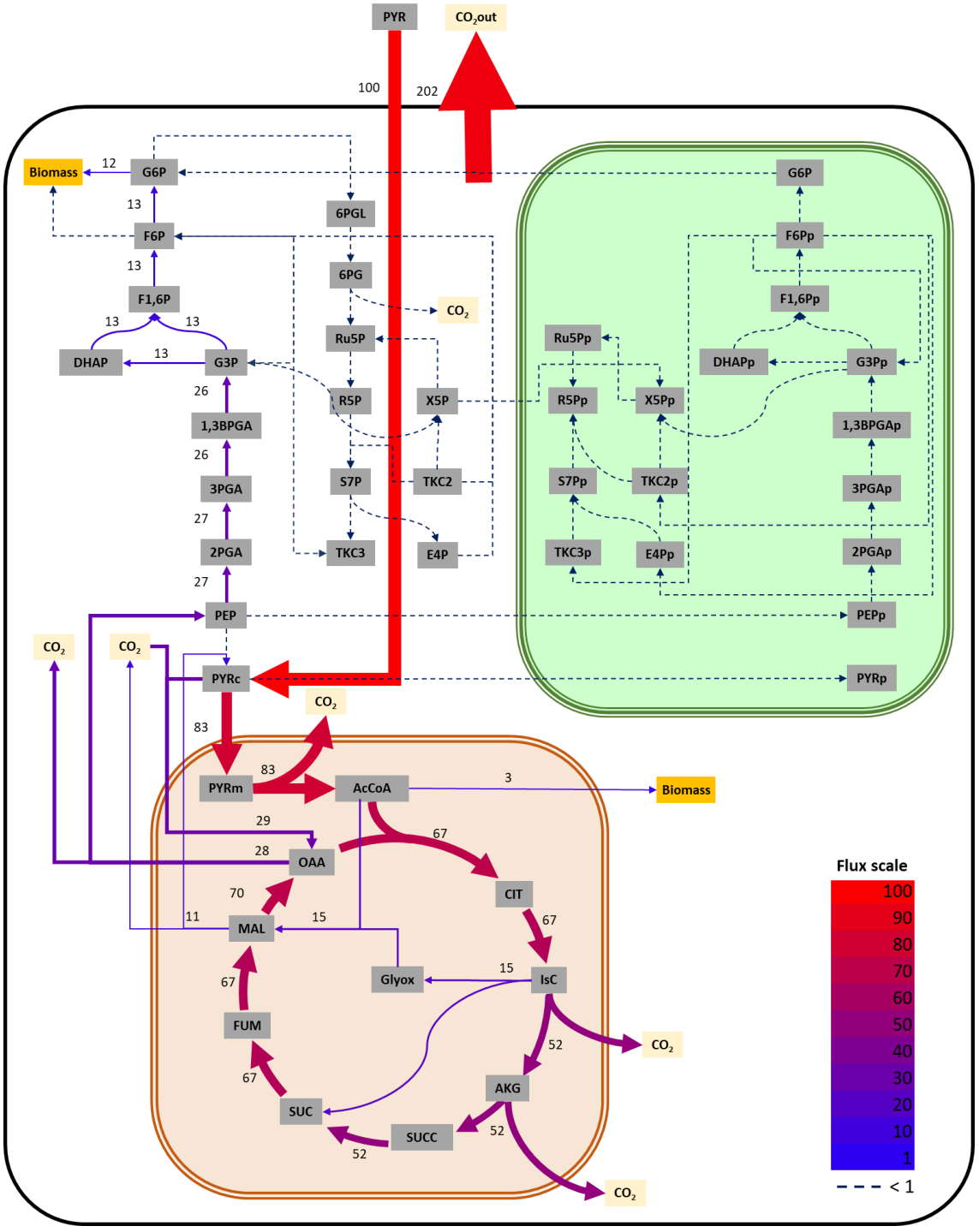


Figure 5.10. Metabolic flux map of *E. gracilis* grown heterotrophically on pyruvate. The thickness of arrows and the numbers represent mol of fluxes per 100 mol of pyruvate intake.

5.3.4 Comparison of ^{13}C MFA flux maps of *E. gracilis* grown on acetate, ethanol and pyruvate

The flux maps show that the metabolism of acetate, ethanol and pyruvate in *E. gracilis* use similar pathways, including the TCA cycle, glyoxylate shunt, anaplerotic reactions and gluconeogenesis. The fluxes through the central metabolic pathway with the upper- and lower-bound of confidence limits are shown in Appendix E. However, the fluxes of *E. gracilis* differed substantially when grown on different substrates. Ethanol and acetate entered the central metabolic pathways after being converted to acetyl-CoA, whilst pyruvate was metabolised by PDH and PC to generate acetyl-CoA and OAA, respectively. Neither PDH nor PC had any measurable flux in the cells grown on ethanol or acetate. For the three substrates, the generated acetyl-CoA was split for use in the glyoxylate shunt and the TCA cycle in different proportions, whilst the fluxes of the glyoxylate shunt were shown to be very small in glucose grown cells (Chapter 4). The aconitase flux in the TCA cycle is highest in the acetate grown cells, followed by pyruvate, ethanol and glucose. Fluxes of the glyoxylate shunt were negligible when cells grown on glucose, these fluxes were substantially higher when cells were grown on ethanol compared to acetate and pyruvate. In acetate and ethanol grown cells, a greatly increased flux of MDH from the contribution of glyoxylate shunt and/or malic enzyme was observed, a feature not present in the pyruvate or glucose grown cells.

PEPCK was active in all three conditions to generate PEP from OAA for gluconeogenesis, in contrast to glucose grown cells (Chapter 4), where PEPC is active instead of PEPCK, and glycolysis is functioning instead of gluconeogenesis. Flux through PEPCK was more than two times higher in ethanol grown cells, compared to acetate and pyruvate.

However, unlike that in acetate and pyruvate grown cells where almost all of the PEP was used in gluconeogenesis, on ethanol, about half of the generated PEP were converted to pyruvate by pyruvate kinase and reintroduced to the TCA cycle by the malic enzyme. G6P was generated from gluconeogenesis at the highest rate in cells grown on ethanol, followed by pyruvate and acetate. As with the analysis on glucose demonstrated in Chapter 4, the flux maps showed that the cytosolic PPP, the plastidic nonoxPPP and plastidic glycolysis contributed negligible flux to the metabolism of acetate, ethanol and pyruvate of *E. gracilis*.

5.3.5 CEEs of *E. gracilis* grown on acetate, ethanol and pyruvate

The net rates of CO_2 release estimated from ^{13}C MFA for *E. gracilis* grown on acetate, ethanol and pyruvate were about 0.17, 0.19 and 0.22 μmol h^{-1} 10^{-6} cells, respectively. In acetate grown cells, CO_2 was mainly produced from the TCA cycle at about 0.14 μmol h^{-1} 10^{-6} cells, and the remainder of about 0.03 μmol h^{-1} 10^{-6} cells was produced from PEPCK. For the ethanol culture, PEPCK produced CO_2 at about 0.19 μmol h^{-1} 10^{-6} cells. However, malic enzyme of ethanol grown cells used CO_2 and pyruvate to produce malate at a rate of 0.10 μmol h^{-1} 10^{-6} cells. Thus, the net rate of CO_2 production from these two anaplerotic enzymes was 0.09 μmol h^{-1} 10^{-6} cells in total. The TCA cycle contributed a further release of 0.09 μmol CO_2 h^{-1} 10^{-6} cells in ethanol grown cells. For acetate and ethanol grown cells, only a negligible amount of CO_2 was generated by PDH. Conversely, in pyruvate grown cells, PDH contributed 0.09 μmol h^{-1} 10^{-6} cells of CO_2 release with a further production of 0.11 μmol h^{-1} 10^{-6} cells CO_2 via the TCA cycle. PEPCK and malic enzyme of *E. gracilis* grown on pyruvate produced CO_2 at about 0.03 and 0.01 μmol h^{-1} 10^{-6} cells, respectively. Unlike that in acetate and ethanol grown

cells, PC used about 0.03 µmol CO_2 h^{-1} 10^{-6} cells to produce OAA. For all three substrates, the oxPPP only contributed minimally to CO_2 production.

The CCE was calculated as described by others (Alonso et al., 2007). Based on the flux of total CO_2 release, the CCEs of *E. gracilis* grown on acetate, ethanol and pyruvate were about 39.1%, 63.5% and 32.7% (Figure 5.11), respectively, which are lower than the value obtained on glucose (82.9%). Using an independent ^{14}C labelling approach to validate the fluxes measured by ^{13}C MFA, the CCE for *E. gracilis* grown on acetate was measured as 23.8 ± 3.05%. As for glucose cultures, the CCEs measured by the ^{14}C approach were lower than those deduced from ^{13}C MFA, which could be the result of an underestimation of CO_2 release in MFA that did not account for degradation of metabolites or the turnover of cellular components. The ^{14}C approach was only conducted with glucose and acetate cultures due to the unavailability of commercial [U-^{14}C]-ethanol and -pyruvate.

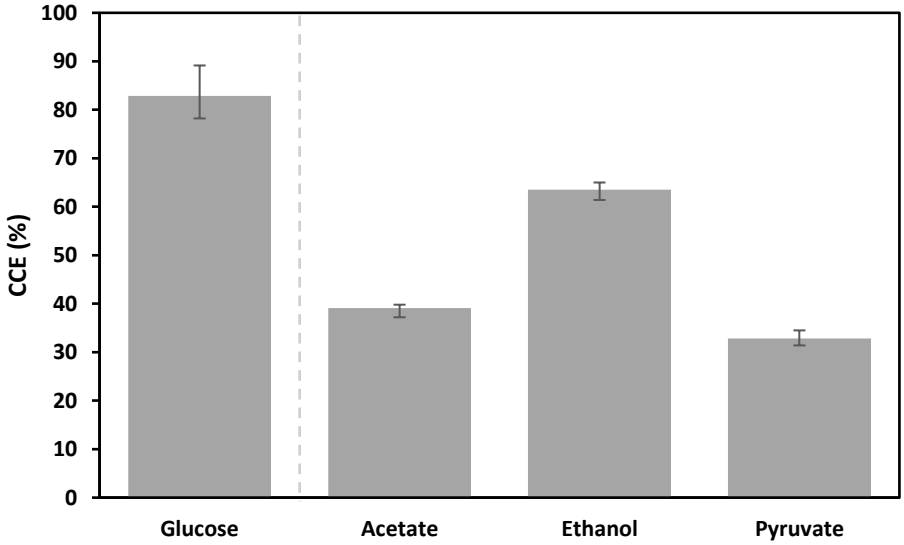


Figure 5.11. Comparison of CCEs of *E. gracilis* grown on glucose (data from Chapter 4), acetate, ethanol and pyruvate, calculated from ^{13}C MFA. The error bars shown in the figure are lower and upper standard confidence limits.

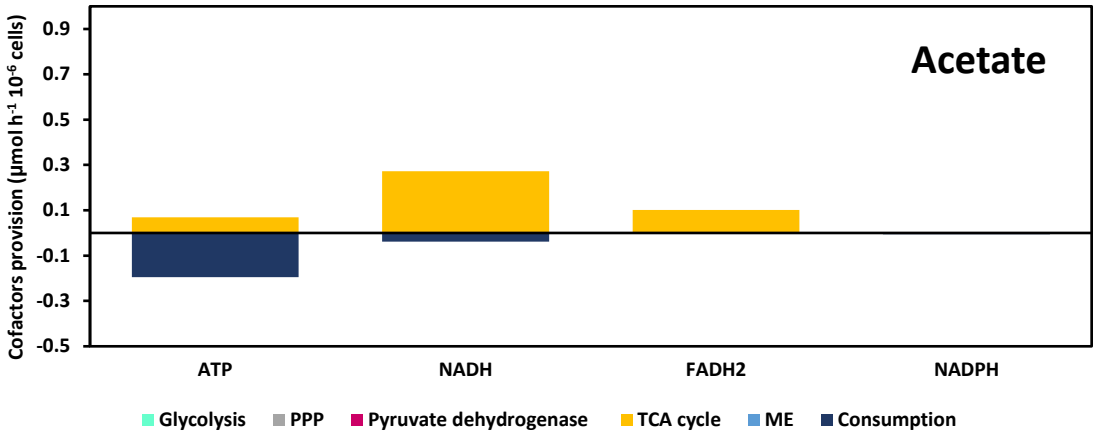
5.3.6 Cofactor production by *E. gracilis* grown on acetate, ethanol and pyruvate

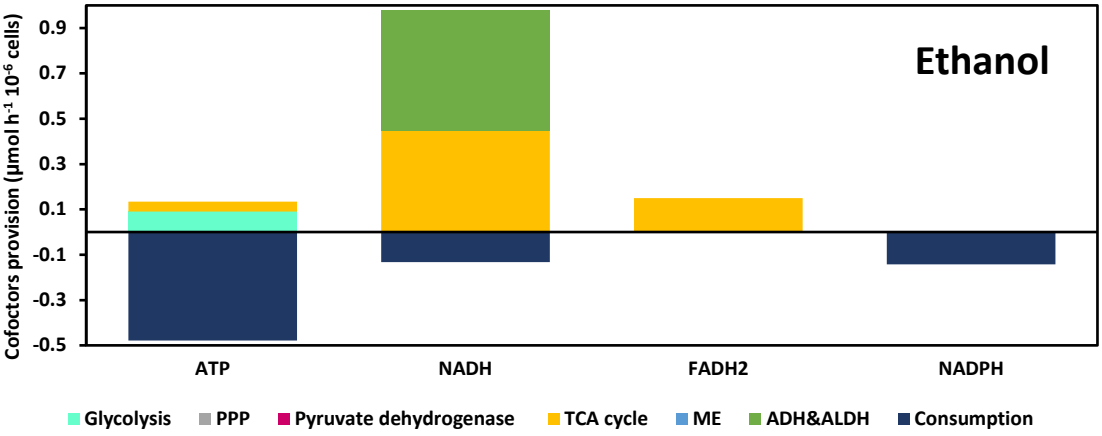
The production and consumption of cofactors (ATP, NADH, $FADH_2$ and NADPH) by *E. gracilis* grown on acetate, ethanol and pyruvate (Figure 5.12) were estimated from the deduced flux maps based on the reaction stoichiometries shown in Appendix D. Negligible amounts of ATP were produced by substrate-level phosphorylation from pyruvate kinase in acetate and pyruvate grown cultures, whilst 0.09 μmol ATP h^{-1} 10^{-6} cells of were produced by this enzyme in ethanol grown cells. ATP was produced by succinyl-CoA synthase at the rates of 0.07, 0.04 and 0.06 μmol h^{-1} 10^{-6} cells in the cells grown on acetate, ethanol and pyruvate, respectively. As expected, the ATP produced from substrate-level phosphorylation was insufficient to meet the estimated demand for ATP consumption for biomass production (0.20, 0.48 and 0.09 μmol h^{-1} 10^{-6} cells for acetate, ethanol and pyruvate grown cells, respectively).

On acetate and pyruvate, NADH was produced almost solely by the TCA cycle at about 0.27 and 0.19 μmol h^{-1} 10^{-6} cells, respectively. ADH and ALDH in ethanol grown cells produced 0.53 μmol NADH h^{-1} 10^{-6} cells and the TCA cycle produced 0.45 μmol NADH h^{-1} 10^{-6} cells. The consumption rates of NADH were estimated as 0.04, 0.13 and 0.04 μmol h^{-1} 10^{-6} cells for acetate, ethanol and pyruvate grown cells, which are lower than NADH production. The consumption rates of NADPH for acetate and pyruvate were substantially lower at 0.01 and

0.01 µmol h^{-1} 10^{-6} cells, respectively. In ethanol grown cells, NADPH was consumed at the rate of 0.14 µmol h^{-1} 10^{-6} cells. NADPH production rates were very low and insufficient to satisfy the consumption rates in *E. gracilis* grown on acetate and ethanol, whilst about 0.09 µmol NADPH h^{-1} 10^{-6} cells was produced by PDH in pyruvate grown cells. The insufficient production of NADPH in acetate and ethanol grown cell could possibly be compensated by the alternative isoenzymes of isocitrate dehydrogenase (NADP-specific, EC 1.1.1.42) (Oda et al., 1981, Graves Jr et al., 1972) and NADP-linked MDH (Isegawa et al., 1984, Miyatake et al., 1985), which have been characterised in *E. gracilis*.

Overall, the central metabolic pathways of ethanol grown cells produced almost twice as much NADH as acetate and pyruvate grown cells, and the rates of cofactor consumption were also higher on ethanol. ATP production from substrate-level phosphorylation is insufficient compared to ATP demands on any substrate. However, the surplus NADH and $FADH_2$ can be used to produce ATP via the electron transport chain to support ATP demands of the cells, and the ATP production from this process was estimated at about 0.72, 1.99 and 0.47 µmol h^{-1} 10^{-6} cells in *E. gracilis* grown on acetate, ethanol and pyruvate, respectively. ATP generated from this process is sufficient compared to the ATP consumption for biomass production.





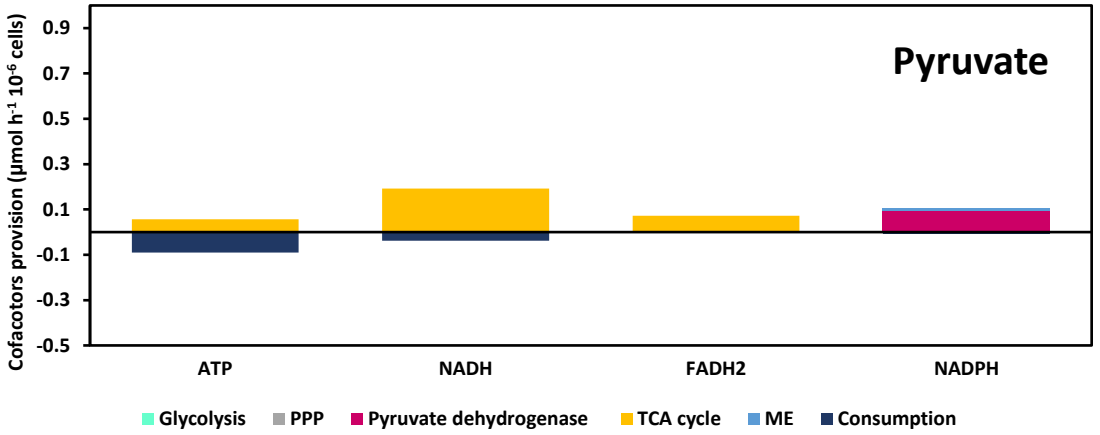


Figure 5.12. Production and consumption of cofactors estimated by ^{13}C MFA flux maps of *E. gracilis* grown on acetate, ethanol and pyruvate under heterotrophic conditions. The

considered cofactors consist of ATP, NADH, $FADH_2$ and NADPH. The production of cofactors was categorised by their sources, including glycolysis, the PPP, pyruvate dehydrogenase, the TCA cycle and malic enzyme (ME).

5.4 FBA of *E. gracilis* grown on acetate, ethanol and pyruvate

5.4.1 FBA of *E. gracilis* grown on acetate, ethanol and pyruvate with biomass constraint

The metabolic network model of *E. gracilis* described in Chapter 3 was constrained with the biomass synthesis reaction formulated from the biomass composition of the cells grown on each substrate. The substrate input was fixed at 100. Objectives of maximising the biomass production and minimising total fluxes through the network were used for FBA.

The results of FBA with the biomass constraint of *E. gracilis* grown on acetate, ethanol and pyruvate are visualised in Figure 5.13, 5.14 and 5.15, respectively. At the maximum potential efficiency based on the stoichiometry of the models, FBA predicted substantially different flux distributions on each substrate. None of the chloroplast pathways was predicted to be active with any substrate. As observed by ^{13}C MFA, FBA predicted that *E. gracilis* metabolised the three substrates using the TCA cycle and PEPCK to support biomass production via gluconeogenesis. In comparison to the ^{13}C MFA of each substrate, with the same rate of substrates intake, FBA with the biomass constraint predicted smaller fluxes through the TCA cycle. The FBA predicted fluxes through the glyoxylate shunt to be higher flux than what observed from ^{13}C MFA for acetate and ethanol grown cells, whist FBA predicted the glyoxylate shunt to be inactive in pyruvate grown cells.

For ethanol grown cells, FBA with only the biomass constraint does not accurately predict the direction of net flux of malic enzyme, shown by ^{13}C MFA to produce malate from pyruvate and CO_2. In addition, the activity of PEPCK and pyruvate kinase was highly underestimated by FBA, showing that FBA cannot predict the CO_2 recycling demonstrated by ^{13}C MFA of *E. gracilis* grown on ethanol. FBA predicted that malic enzyme would be substantially active to introduce pyruvate to the TCA, instead of PC demonstrated in ^{13}C MFA. This demonstrated that FBA with only the biomass constraint of the *Euglena* metabolic model cannot predict the flux through this alternative route specifically used when grown with pyruvate. The inability to prediction the involvement of alternative routes has been emphasised as one of the common limitations of FBA (Cheung et al., 2015).

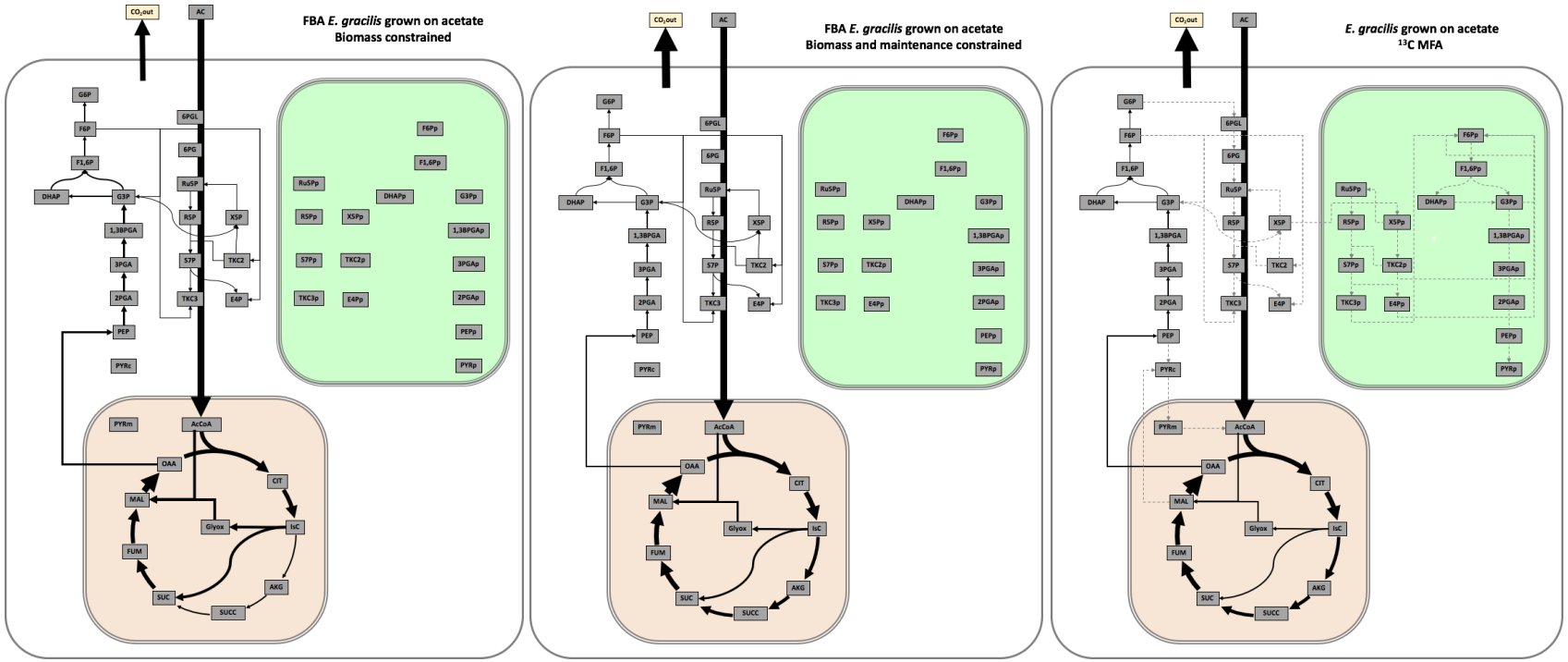


Figure 5.13. Comparison of flux maps of *E. gracilis* grown on acetate; FBA with only the biomass constraint (left), FBA with the biomass constraint and maintenance included (middle) and ^{13}C MFA (right). Thickness of the arrows indicates the fluxes relative to the rate of acetate input.

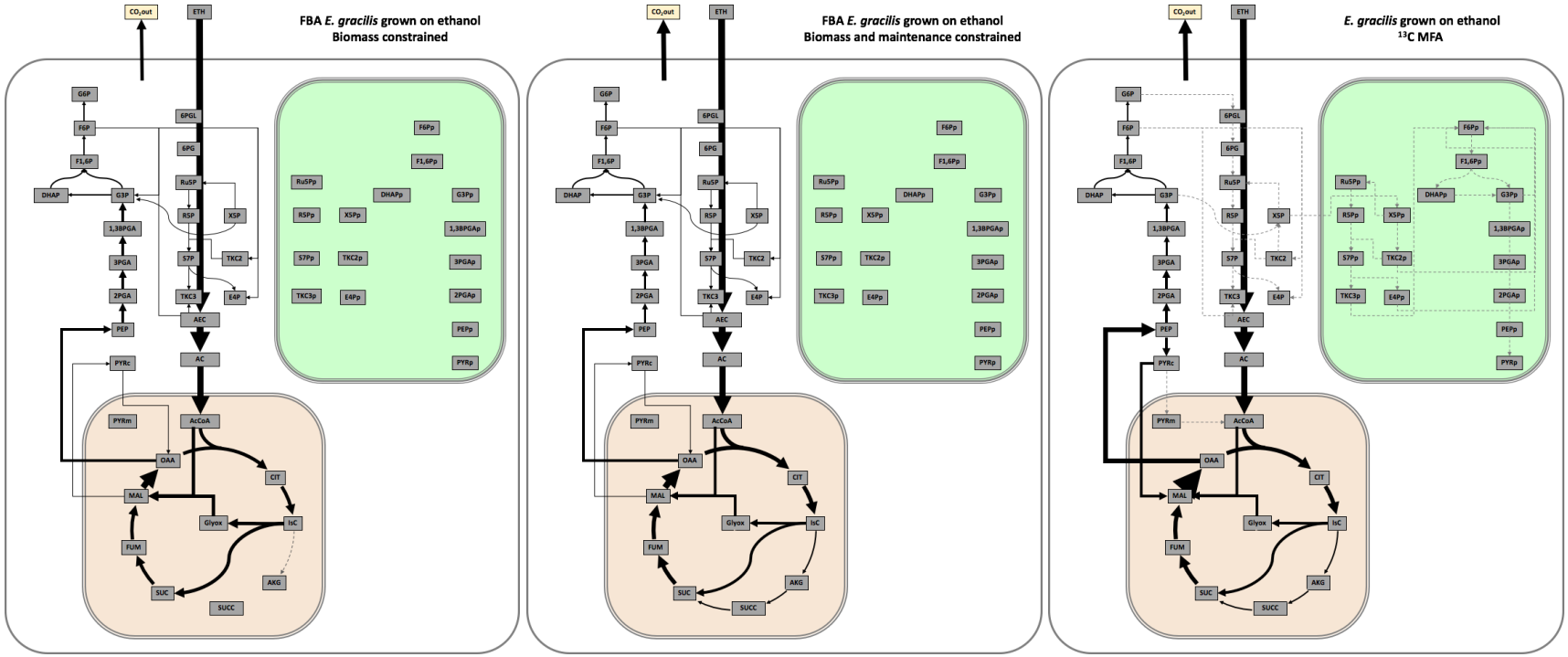


Figure 5.14. Comparison of flux maps of *E. gracilis* grown on ethanol; FBA with only the biomass constraint (left), FBA with the biomass constraint and maintenance included (middle) and ^{13}C MFA (right). Thickness of the arrows indicates the fluxes relative to the rate of ethanol input.

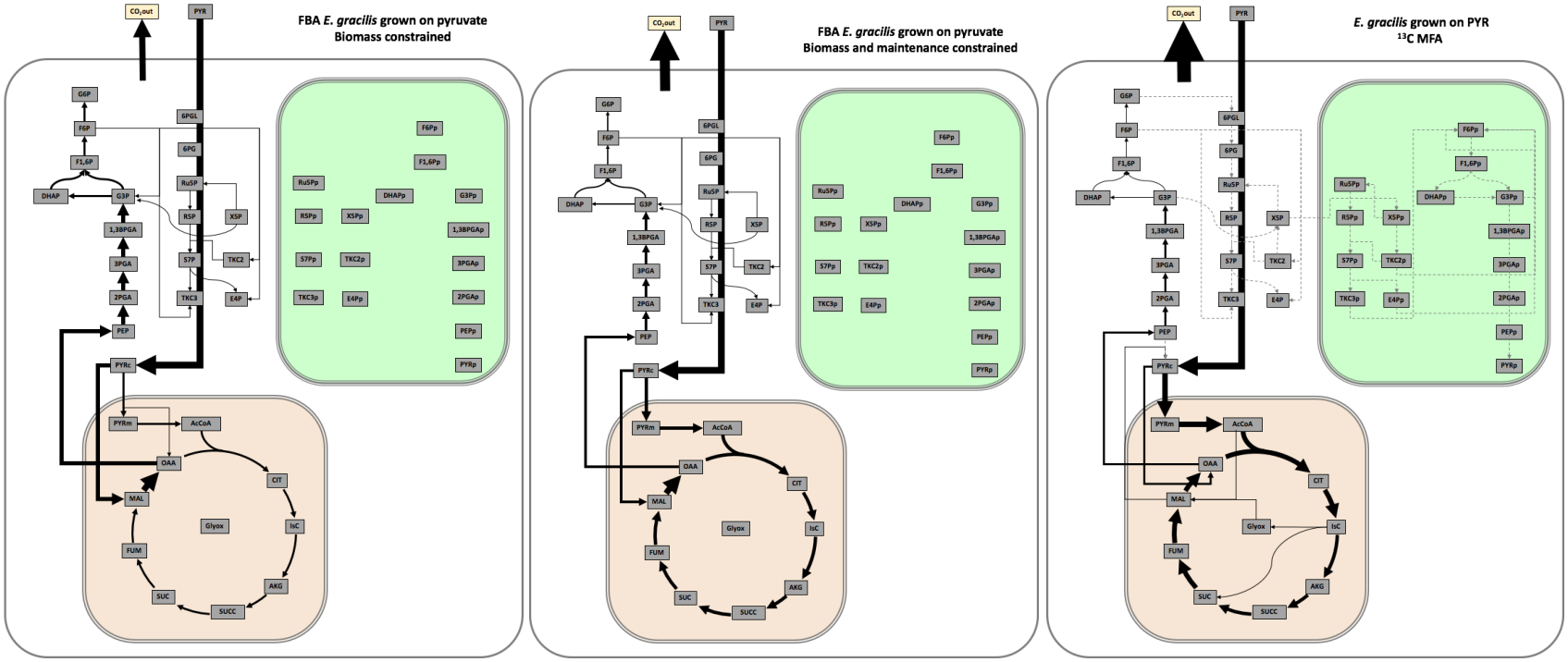


Figure 5.15. Comparison of flux maps of *E. gracilis* grown on pyruvate; FBA with only the biomass constraint (left), FBA with the biomass constraint and maintenance included (middle) and ^{13}C MFA (right). Thickness of the arrows indicates the fluxes relative to the rate of pyruvate input.

5.4.2 Estimating cell maintenance of *E. gracilis* grown on acetate, ethanol and pyruvate

As described in Section 4.4.2, the potential maintenance costs in *E. gracilis* grown on acetate, ethanol and pyruvate were estimated as shown in Table 5.2. Normalised by consumption rates of each substrate, the estimated potential cofactor availability for maintenance was calculated as the differences between the cofactors provision measured by ^{13}C MFA and that predicted by FBA.

Table 5.2. Cofactors production estimated from FBA and MFA results normalised to substrate uptake rate of *E. gracilis* grown on acetate, pyruvate and ethanol

Cofactors	Acetate (mol (mol acetate)$^{-1}$)			Ethanol (mol (mol ethanol)$^{-1}$)			Pyruvate (mol (mol pyruvate)$^{-1}$)		
	FBA	^{13}C MFA	Available for Maintenance	FBA	^{13}C MFA	Available for Maintenance	FBA	^{13}C MFA	Available for Maintenance
ATP	0.21	0.50	0.29	0	0.52	0.52	0.26	0.53	0.27
NADH	1.31	1.96	0.65	2.88	3.78	0.84	1.16	1.77	0.61
$FADH_2$	0.58	0.73	0.15	0.47	0.58	0.11	0.26	0.65	0.39
NADPH	0.06	0.00	-0.06	0.08	0.01	-0.07	0.72	0.99	0.27

FBA estimated a higher NADPH requirement for biosynthesis than the production estimated by ^{13}C MFA in acetate and ethanol grown cells. The insufficient production of NADPH was presumed to be compensated for by the alternative isoenzyme of NADP-specific isocitrate dehydrogenase (Oda et al., 1981, Graves Jr et al., 1972) and/or NADP-linked isozymes of MDH (Isegawa et al., 1984, Miyatake et al., 1985). Thus, the amount of NADH

equivalent to that required for NADPH maintenance was subtracted from the estimated available NADH maintenance before proceeding with further calculations.

NADH and $FADH_2$ maintenance were used to estimate the total ATP maintenance for constraining the FBA with a stoichiometry of 2.5 and 1.5 molecules of ATP produced per molecule of NADH and $FADH_2$ (Foo et al., 2020). The total estimated maintenances in *E. gracilis* are shown in Figure 5.16. The results show that, relative to 1 mol of substrate intake, *E. gracilis* grown on acetate, ethanol and pyruvate required 1.99, 2.59 and 2.40 mol ATP for maintenance, corresponding to about 31%, 27% and 40% of the total ATP produced estimated by ^{13}C MFA, respectively. The ATP maintenances on these substrates are higher than that on glucose (1.63 mol ATP, 29% of total ATP produced). Pyruvate grown cells produce available NADPH for maintenance at about 0.27 mol (mol pyruvate)$^{-1}$, similar to that on glucose (0.30 mol (mol glucose)$^{-1}$). This demonstrates that, on different substrates, *E. gracilis* produce different amounts of cofactors that available for maintenance.

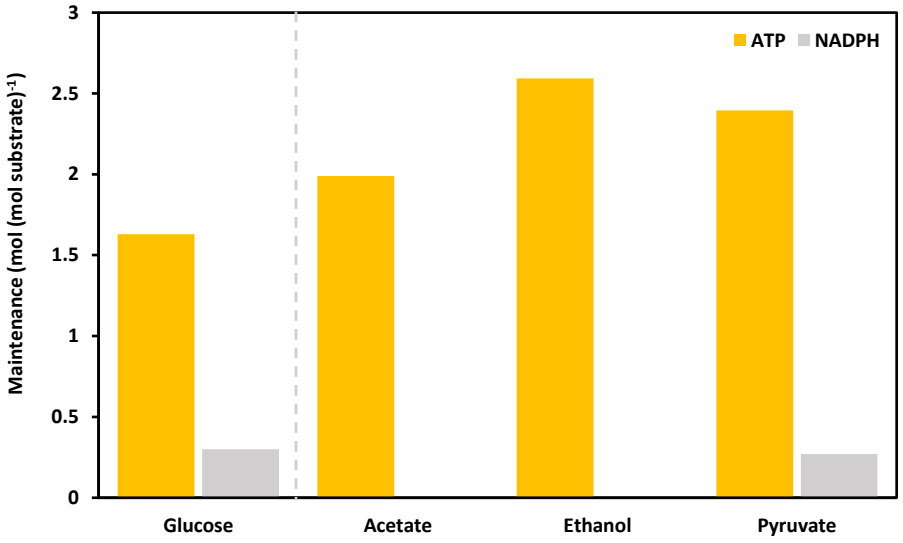


Figure 5.16. Estimated potential ATP and NADPH maintenance cost of *E. gracilis* grown on glucose (Chapter 4), acetate, ethanol and pyruvate. The maintenance was calculated by comparing reductants provisions estimated by ^{13}C MFA and FBA, the excess NADH and $FADH_2$ were assumed to be oxidised to generate ATP through the electron transport chain.

5.4.3 Accounting for maintenance in FBA of *E. gracilis* grown on acetate, ethanol and pyruvate

The fluxes from FBA with maintenance included and ^{13}C MFA of *E. gracilis* grown on acetate, ethanol and pyruvate are shown in Figure 5.13, 5.14 and 5.15 respectively. The flux maps demonstrate that incorporating the estimated potential maintenance costs into the FBA improved the prediction of fluxes through the TCA cycle, glyoxylate shunt and gluconeogenesis for all three substrates. However, the flux of PEPCK and pyruvate kinase was still underestimated by FBA for *E. gracilis* grown on ethanol, and PC was still predicted to be inactive in pyruvate grown cells.

Root mean square deviation (RMSD) was used to compare the match between flux solutions of ^{13}C MFA and FBA with or without inclusion of maintenance considerations (Figure 5.17). These comparisons show that accounting for maintenance improved the flux prediction of FBA with different levels of improvement on the cells grown on different substrates. On acetate, accounting for maintenance greatly improved the prediction, as the RMSD was reduced from 8.98 to 1.34. On the other hand, accounting for maintenance improved flux predictions to smaller extents in ethanol and pyruvate grown cells. This reflected the inability

of FBA to predict the accurate fluxes of anaplerotic reactions crucial for the metabolism of ethanol and pyruvate in *E. gracilis*.

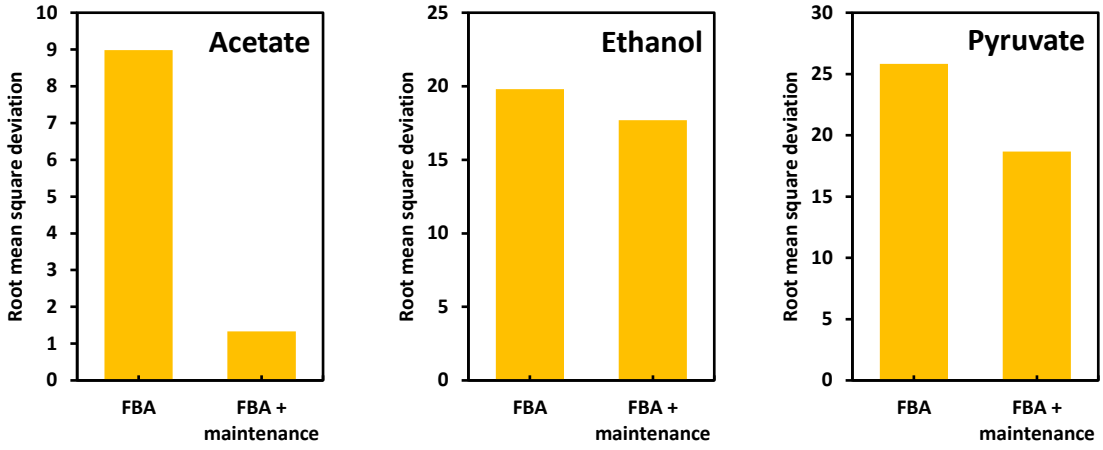


Figure 5.17. RMSD comparing fluxes of FBA with biomass constraint and FBA with maintenance included for *E. gracilis* grown on acetate (left), ethanol (middle) and pyruvate (right). RMSD was calculated based on the free fluxes of ^{13}C MFA results and the same reactions in FBA solution to assess the match between FBA and MFA fluxes.

5.4.4 Metabolic efficiency of *E. gracilis* grown on acetate, ethanol and pyruvate

The maximal potential CCEs predicted by FBA with only the biomass constraint of *E. gracilis* grown on acetate, ethanol and pyruvate are 60.9%, 76.6% and 69.1% respectively. Accounting for maintenance cost lowered CCEs to 42.2% and 63.7% in acetate and ethanol grown cells, which are closely matched with that estimated by ^{13}C MFA on acetate (39.1%) and ethanol (63.5%). Accounting for maintenance lowered the CCE of pyruvate grown cells from 69.1% to 50.8%, which is still substantially higher than the 32.4% estimated by ^{13}C MFA

(Figure 5.18). This is the result of the underestimated fluxes of PDH and malic enzyme by FBA of pyruvate grown cells.

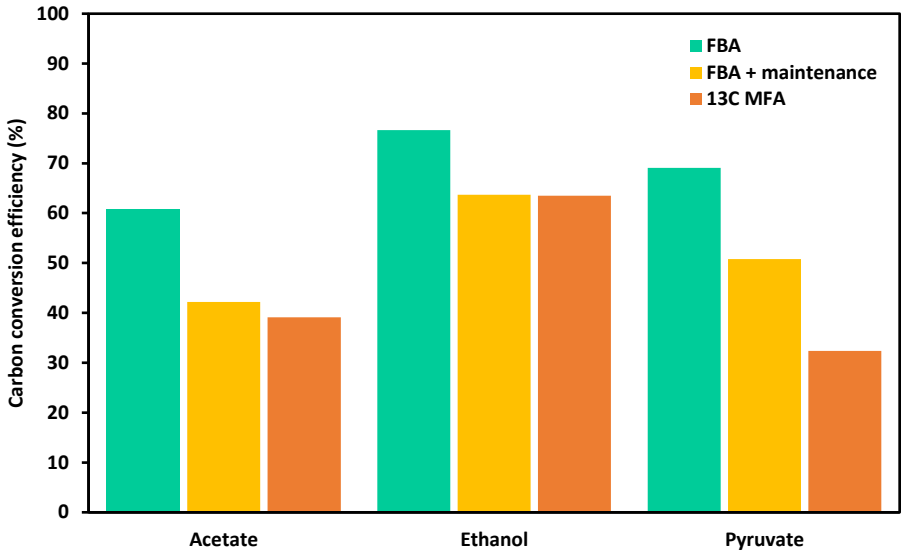


Figure 5.18. Comparison of carbon conversion efficiency of *E. gracilis* grown heterotrophically on acetate, ethanol and pyruvate by different approaches; FBA (green), FBA with maintenance accounted (yellow) and ^{13}C MFA (orange).

The metabolic efficiencies based on CCEs of *Euglena* grown on the three substrates were calculated as the proportion of CCE from ^{13}C MFA of the maximal potential CCE from FBA without accounting for maintenance (Figure 5.19). The results demonstrate that, in comparison to acetate and pyruvate, *E. gracilis* grown on ethanol achieved the highest metabolic efficiency at about 82.9% of the maximal potential CCE, almost as high as the 90.0% achieved when grown on glucose. About 64.2% and 46.8% of the maximal potential CCE were achieved in acetate and pyruvate grown cells, respectively.

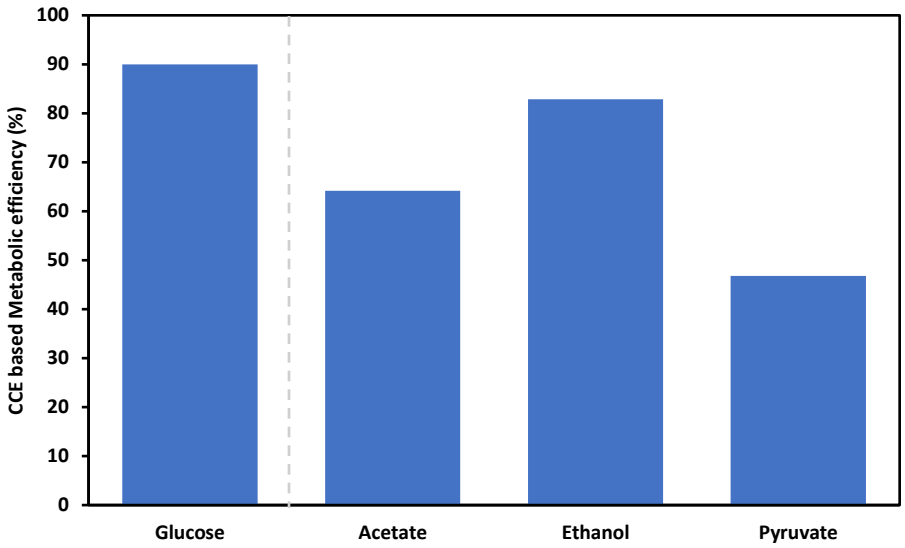


Figure 5.19. Metabolic efficiency based on carbon conversion efficiency of *E. gracilis* grown heterotrophically on glucose (data from Chapter 4), acetate, ethanol and pyruvate. The metabolic efficiency is calculated as the proportion of observed CCE (^{13}C MFA) of the estimated maximum potential CCE (FBA), indicating efficiency of carbon utilisation of the cells relative to the theoretical performance determined by the stoichiometry of the metabolic network.

5.5 Discussion

5.5.1 ^{13}C MFA flux maps reveal the metabolic flux phenotypes of *E. gracilis* and emphasise the limited contribution of plastid metabolic processes and the oxPPP in heterotrophic metabolism.

^{13}C MFA was successfully exploited to determine the flux distribution in the metabolism of acetate, ethanol and pyruvate by *E. gracilis*. Supporting the general understanding of C2 compound metabolism in *E. gracilis* (Nakazawa, 2017), the flux maps of *E. gracilis* grown on acetate and ethanol demonstrate a highly active glyoxylate shunt to

enable gluconeogenesis. On these three substrates, but not on glucose, PEPCK is shown to be key to producing PEP required for gluconeogenesis to facilitate the production of precursors for biomass synthesis. PEPCK has previously been detected in *Euglena* grown on acetate and ethanol, but not glucose (Peak et al., 1972) and had significantly increased activity when grown on ethanol (Miyatake et al., 1984a), indicating the significance of this enzyme in gluconeogenesis in *E. gracilis*.

The flux maps show that *E. gracilis* metabolises the majority of pyruvate via PDH and the TCA cycle, in line with the general understanding of pyruvate metabolism in other eukaryotes (Gray et al., 2014). In prokaryotes, fungi and mammals, pyruvate can also be metabolised by PC to supply OAA to the cells (Utter and Keech, 1960, Wexler et al., 1994, Jitrapakdee et al., 2008, Yang et al., 2011). The flux map of *E. gracilis* on pyruvate demonstrates that PC operates when grown on pyruvate, but not on other substrates, to produce OAA for the TCA cycle and gluconeogenesis. Compared to acetate and ethanol grown cells, fluxes of the glyoxylate cycle in pyruvate-grown cells are smaller. This suggests the importance of PC in supporting anaplerosis in *E. gracilis* when grown on pyruvate.

As found for glucose-grown cells, flux maps of *E. gracilis* grown on acetate, ethanol and pyruvate emphasise the limited contribution of chloroplast pathways in heterotrophic metabolism in *Euglena* and of the oxPPP in NADPH provision. All the enzymes necessary for *E. gracilis* central metabolic pathways are present in the cytosol or mitochondria, with some duplication in the chloroplasts (Inwongwan et al., 2019). This might allow the metabolic network of *E. gracilis* to operate independently of the chloroplasts during heterotrophic metabolism. The limited contribution of plastid pathways in the heterotrophic metabolism of

acetate was also observed in *C. reinhardtii* (Boyle et al., 2017). In contrast to plants (embryophytes), in which the oxPPP in plastids provide the majority of NADPH for the cells under non-photosynthetic conditions (Kruger and von Schaewen, 2003; Schwender et al., 2003), the oxPPP in *E. gracilis* contributes a negligible amount of NADPH to the cells. The NADPH provision of *E. gracilis* grown on acetate, ethanol and pyruvate will be discussed in detail in Section 5.5.3.

5.5.2 Despite the same carbon composition and similar metabolic network entry points, the metabolic phenotypes of *E. gracilis* grown on acetate and ethanol are substantially different.

The flux maps show that *E. gracilis* metabolises acetate and ethanol using the same pathways. In brief, ethanol is converted to acetate by ADH and ALDH. Acetate is then converted to acetyl-CoA, and then is metabolised through the TCA cycle or channelled into the glyoxylate shunt, as previously described in *Euglena* (Nakazawa, 2017) and found in *E. coli* (Zhao and Shimizu, 2003), *Yarrowia lipolytica* (Liu et al., 2016), *S. cerevisiae* (Hayakawa et al., 2018) and *C. reinhardtii* (Boyle et al., 2017). However, considering the same metabolic entry point and carbon composition of the two substrates, the pathway fluxes of *E. gracilis* grown on acetate and ethanol are very different.

On acetate, the flux map of *E. gracilis* shows that pyruvate kinase and PDH are essentially inactive, a result that is similar to that found in *C. reinhardtii* (Boyle et al., 2017) and *E. coli* (Zhao and Shimizu, 2003) when grown on acetate. PDH is also inactive in *E. gracilis* grown on ethanol, but pyruvate kinase is shown to be active similar to observations in *S. cerevisiae* (Hayakawa et al., 2018). Compared to ethanol, the fluxes through the TCA cycle of

acetate grown *E. gracilis* are higher. On the other hand, the flux map of ethanol grown cells shows higher fluxes through the glyoxylate shunt and PEPC, supporting the higher flux through gluconeogenesis to synthesise biomass, corresponding to the higher growth rate observed when grown on ethanol. Fluxes from isocitrate dehydrogenase to succinyl-CoA synthetase of ethanol grown cells are substantially lower, as a higher proportion of isocitrate was drawn to the glyoxylate shunt. In contrast, the reduced fluxes through the decarboxylating arc of the TCA cycle are not observed in *S. cerevisiae* during ethanol assimilation (Hayakawa et al., 2018). This could be the influences from the subcellular location of the glyoxylate shunt which is located in the peroxisome and cytosol of *S. cerevisiae* (Regev-Rudzki et al., 2005). The presences of the glyoxylate shunt and the TCA in mitochondria of *E. gracilis* (Ono et al., 2003) may allow both pathways to operate interchangeably.

With the same carbon intake, ethanol-grown cells release a smaller proportion of CO_2, at about 36% of total carbon intake, than acetate grown cells, at 61%. Although the flux of PEPCK produced substantially more CO_2 on ethanol, CO_2 equivalent to about half of this was used to produce malate by malic enzyme, demonstrating the role of anaplerotic reactions in operating a futile carbon cycle when grown on ethanol, and not on acetate. This futile cycle may be operated in order to dissipate the surplus ATP generated when grown on ethanol. This function of a futile cycle has been suggested in mammals, microorganisms (Portais and Delort, 2002) and plant cells (Schwender et al., 2004).

The results show that the total CO_2 released from anaplerotic reactions of ethanol grown cells (35 mol (100 mol ethanol)$^{-1}$) is higher than that on acetate (23 mol (100 mol

acetate)$^{-1}$). The reduced fluxes of isocitrate dehydrogenase and α-ketoglutarate dehydrogenase contribute much less CO_2 on ethanol (35 mol (100 mol ethanol)$^{-1}$) compared to that on acetate (99 mol (100 mol acetate)$^{-1}$). This demonstrates that the reduced TCA fluxes by partitioning the isocitrate to the glyoxylate shunt lead to the low CO_2 release in *E. gracilis* grown on ethanol.

The glyoxylate shunt is used for metabolism of acetate and ethanol (Nakazawa, 2017) and the flux through the shunt is indicative of C2 substrate assimilation into biomass (Boyle et al., 2017). In this study, the split to the glyoxylate shunt of ethanol grown cells was about 68% of the incoming flux from isocitrate, which is substantially higher than the 32% of acetate. The reactions of isocitrate dehydrogenase and α-ketoglutarate dehydrogenase produce two mol of NADH (or NAD(P)H for isocitrate dehydrogenase) and two mol of CO_2 per mol of fluxes. On ethanol, the flux map shows that two mol of NADH are initially generated via the reaction of ADH and ALDH to produce acetate without losing any carbon atoms. The surplus NADH from this process could be used to compensate for the reduced NADH production of the TCA cycle. This might allow the network to channel more fluxes through the glyoxylate shunt to produce more biomass, which leads to the decrease in carbon loss.

Relative to one mol of substrate taken up, *E. gracilis* grown on ethanol and acetate generate 2.59 and 1.99 mol ATP available for maintenance, respectively. Although the maintenance is estimated to be higher on ethanol, the ATP maintenance required only about 27% of the total ATP produced, which is lower than that of 31% on acetate. This higher proportion of ATP for maintenance may require higher fluxes through isocitrate dehydrogenase and α-ketoglutarate dehydrogenase when grown on acetate. It was reported

that the relatively large flux of isocitrate dehydrogenase observed in *C. reinhardtii* indicates that the cells increase the production of NAD(P)H as a result of a high maintenance energy requirement (Boyle et al., 2017).

5.5.3 Flux through MDH in *E. gracilis* was elevated when grown on acetate and ethanol, responding to the precursors and NADPH demand for biomass production.

^{13}C MFA flux maps estimated that the NADPH production rates are negligible in *E. gracilis* grown on acetate and ethanol and certainly insufficient for the consumption rates. As on glucose, the majority of NAPDH is produced by PDH when grown on pyruvate. However, PDH shows no flux in acetate and ethanol grown cells. This corresponds to a previous study in *Saccharomyces cerevisiae*, showing the downregulation of PDH flux when grown on ethanol (Hayakawa et al., 2018).

With negligible flux through oxPPP and PDH, the requirement for NADPH production in acetate- and ethanol-grown cells could be met using the alternative isoenzymes of Isocitrate dehydrogenase and/or NADP-linked MDH in the TCA cycle. *Euglena* contains NAD- and NADP-specific isozymes of isocitrate dehydrogenase and the NADP-specific isozymes were reported in both mitochondria and the cytosol (Oda et al., 1981, Graves Jr et al., 1972). Two forms of MDH, NAD-linked and NADP-linked isozymes, have been characterised in *E. gracilis* (Isegawa et al., 1984, Miyatake et al., 1985). In contrast to higher plants, where the NADP-linked MDH is present exclusively in the chloroplasts, the activity of NADP-linked MDH was detected in mitochondria and the cytosol of *E. gracilis* (Isegawa et al., 1984, Miyatake et al., 1985).

Flux through MDH is substantially greater than that through other reactions in the TCA cycle in *E. gracilis* grown on acetate and ethanol, but not on pyruvate and glucose. This feature has not been reported in either *E. coli* grown on acetate (Zhao and Shimizu, 2003) or yeast grown on ethanol (Hayakawa et al., 2018). It was reported that the activity of MDH in *E. gracilis* was increased when grown on acetate and ethanol, compared to glucose grown cells (Peak et al., 1972). The flux maps of *E. gracilis* demonstrate that MDH has an important role in channelling the OAA to gluconeogenesis, by increasing the amount of OAA for PEPC. The increased flux of MDH, observed when PDH is inactive and resulting from diverting isocitrate to the glyoxylate shunt, could possibly requite the NADPH demand of *E. gracilis* grown on acetate and ethanol. Flux through isocitrate dehydrogenase is particularly low when grown on ethanol rather than acetate. On the other hand, the flux through MDH is substantially higher, corresponding to the higher estimated demand for biosynthetic precursors and NADPH by the cells. This indicates the likely importance of MDH in provision of NADPH in *E. gracilis* when grown on acetate and ethanol.

5.6 Conclusion

In this chapter, ^{13}C MFA was successfully exploited to reveal the metabolic phenotype of *E. gracilis* grown on acetate, ethanol and pyruvate. The flux maps of *E. gracilis* emphasise the limited contribution of plastid metabolic processes and the oxPPP in heterotrophic metabolism of these substrates. On acetate and ethanol, compounds with the same carbon composition and similar entry points into the metabolic network, the metabolic phenotypes of *E. gracilis* differ in the fluxes through the glyoxylate shunt that respond to the different

cofactor provision from the substrates and demand of the cells. The work demonstrates the importance of the glyoxylate shunt and MDH in anaplerosis when grown on ethanol and acetate. On the other hand, anaplerosis is mainly supported by PC when grown on pyruvate. PEP, which is required for gluconeogenesis to produce the precursors for biomass synthesis, is produced by PEPCK from OAA. The latter is, in turn, generated from pyruvate either directly by PC or indirectly from its decarboxylation product, acetyl-CoA, by the glyoxylate shunt and MDH. The flux maps also indicate the important role for MDH in *E. gracilis* for the production of biomass precursors and NADPH when grown on acetate and ethanol.

Chapter 6 | General discussion

6.1 Summary

In this thesis, I developed metabolic models of *E. gracilis* suitable for MFA and FBA. I constructed the metabolic network model for heterotrophic *E. gracilis* using transcriptomic and proteomic data as described in Chapter 3. A central feature of this model is that pathways necessary for the survival of *Euglena* are present in the cytosol or mitochondria with the duplication of some pathways in chloroplasts, consistent with the unique ability of Euglena to survive when its chloroplasts are eliminated. ^{13}C MFA was subsequently exploited to determine the metabolic phenotypes of *E. gracilis* grown on glucose, acetate, ethanol and pyruvate.

The flux maps presented in Chapter 4 demonstrate that destruction of the chloroplasts increased demand for cofactors by cell maintenance, resulting in higher fluxes through central pathways and decreased CCE during glucose metabolism. The metabolic phenotypes of *E. gracilis* grown on acetate, ethanol and pyruvate were determined in Chapter 5 to investigate the effect of respiratory substrates differing in carbon composition and chemical redox potential in the operation of the metabolic network. The results show that *E. gracilis* exploited different combinations of reactions when grown on substrates with different carbon composition. The flux maps demonstrate the importance of the glyoxylate shunt and MDH in anaplerosis when grown on acetate and ethanol. On the other hand, in *E.*

gracilis grown on pyruvate, anaplerosis was mainly supported by the formation of dicarboxylates via PC instead of the glyoxylate shunt.

6.2 The metabolic flexibility of the central metabolic network of *E. gracilis*

Euglena possesses a diverse metabolic capacity that allows it to grow in a wide range of conditions (Einicker-Lamas et al., 2002). *Euglena* is able to harness energy heterotrophically, mixotrophically and photo-autotrophically, and it can utilise a broad range of carbon substrates. The versatile metabolic capacity of *Euglena* was previously highlighted by previous transcriptomic studies (O'Neill et al., 2015). The analysis of proteins encoded from these transcriptomes demonstrate complex metabolic capacity, obtained from diverse sources (Ebenezer et al., 2019). The flux maps demonstrate the ability of *E. gracilis* to adjust the operation of its metabolic pathways according to the nature of the available substrates. This ability of an organism to respond to the prevailing conditions is described as “metabolic flexibility” (Goodpaster and Sparks, 2017, Köhler, 1985).

Flux maps of *E. gracilis* growing on different substrates establish its ability to selectively use or activate different enzymes/pathways. The flux maps showed that PC is active only when grown on pyruvate, but not other substrates. Consistent with a role restricted to metabolism of a selective group of substrates including pyruvate (and possibly chemically related compounds), the activity of PC was not detected when grown autotrophically (Kitaoka et al., 1989). On the other hand, when grown on acetate and ethanol the glyoxylate shunt is highly active and PDH contributes negligible flux. Additionally, the flux

maps show that PEPC only functions when grown on glucose, whilst PEPCK functions when grown on acetate, ethanol and pyruvate. In comparison to acetate and pyruvate, flux of PEPCK is very high when grown on ethanol to form a futile cycle with the reversed malic enzyme reaction, this futile cycle is not shown in glucose, acetate and pyruvate cultures.

Apart from the variety of enzymes necessary for utilising each of these carbon sources as a growth substrate, several metabolic reactions of *Euglena* are capable of being catalysed by multiple isozymes that are located in different subcellular compartments and which can be selectively used under different conditions. For instance, *E. gracilis* has multiple isozymes of PEPC which are active in response to different light conditions, fixing CO_2 in *Euglena* grown in the dark or in the light under CO_2 limited conditions (Codd and Merrett, 1971, Laval-Martin et al., 1981). *E. gracilis* has the ability to use NAD- and NADP-specific PDH or NAD- and NADP-specific MDH when grown under different conditions (Isegawa et al., 1984, Miyatake et al., 1985). The results from Chapter 5 suggest that these different cofactor-specific isozymes could contribute to satisfying/meeting the NADPH demand of cells grown on acetate and ethanol.

Moreover, the subcellular organisation of the enzymes/pathways in *Euglena* cells could enhance its metabolic efficiency by providing the ability to adapt to different growth conditions and to utilises a variety of substrate efficiently. The glyoxylate shunt, generally localised in peroxisomes or the cytosol in plants and fungi (Kunze et al., 2006), was reported to be confined to mitochondria in *Euglena*, where the activities of isocitrate lyase and malate synthase were exclusively detected (Ono et al., 2003). The flux maps generated in Chapter 5 suggest that the glyoxylate shunt operates in close conjunction with the TCA cycle by

partitioning the isocitrate in response to precursor supply and NAD(P)H requirement. The regulation of the flux between the glyoxylate shunt and the TCA cycle might be facilitated by the presence of both pathways in the same subcellular location, the mitochondria, in *E. gracilis*.

In contrast to most eukaryotes, the enzymes involved in metabolism of C2 substrates are reported to be located in the mitochondria of *Euglena* (Nakazawa, 2017). Having these enzymes located in the same compartment could allow *Euglena* to metabolise C2 substrates with less subcellular transport processes required, and might also accommodate the enzyme activities by concentrating intermediates in the compartment. The ADH and ALDH of *E. gracilis* required for metabolism of ethanol are located in the inner mitochondrial membrane (RodrÍGuez-Zavala et al., 2006, Yoval-Sánchez et al., 2011, Ono et al., 1995). The presence of ADH and ALDH in the inner mitochondrial membrane might facilitate the use of ethanol and acetaldehyde by preventing the toxic effects of these compounds on the respiratory complex and other mitochondrial enzymes (Nakazawa, 2017). Acetyl-CoA synthetase required to convert acetate to acetyl-CoA is also located in *Euglena* mitochondria (Graves Jr and Becker, 1974).

Additionally, as described in Chapter 3, the subcellular organisation of the central metabolic pathways of *E. gracilis* allows the cells to grow in the complete absence of its chloroplasts. In contrast to plants and green algae, in *E. gracilis* the complete PPP is present in the cytosol with a duplication of the nonoxPPP steps in chloroplasts (Inwongwan et al., 2019), giving a functional PPP without chloroplasts. Similarly, all the other key pathways needed for heterotrophic metabolism are present in the cytosol or mitochondria with the

duplications of some pathways in the chloroplasts, enabling the central metabolic pathways of *Euglena* to operate independently of the chloroplasts when grown heterotrophically.

6.3 The implementation of MFA and FBA for analysing the metabolic network of *E. gracilis*

MFA and FBA are both powerful techniques for estimating metabolic fluxes but they are based on different concepts. MFA relies on matching of isotopic rearrangement data of metabolites to the simulated results based on the relationships between carbon atoms within intermediates of the central metabolic pathways (Guo et al., 2015), whilst FBA predicts fluxes based on the stoichiometry of the genome-scale metabolic network model constrained by various parameters, such as biomass composition, energy and redox demand, and respiratory rate (Orth et al., 2010, Fang et al., 2020). In this thesis, I have explored the use of both techniques for studying heterotrophic metabolism in *E. gracilis*.

The comparison of ^{13}C MFA and FBA allow the generalised adjustment to FBA to permit the prediction of several specific fluxes under different sets of conditions. This could lead FBA to generate more accurate prediction of fluxes throughout a comprehensive metabolic network defined by a genome-scale stoichiometric model. By conducting MFA and FBA in parallel under the same conditions we can take advantage of using MFA as a reference to validate the predictive performance of the FBA and to identify alternative metabolic pathways, improving the predictions of FBA in specific conditions. However, performing ^{13}C MFA is relatively more labour-intensive. The well-curated FBA model can be used to predict

the theoretical metabolic phenotype in the conditions of interest, and also to indicate the conditions which might be useful to undertake the MFA.

The FBA results in Chapters 4 and 5 demonstrate the integrity of the model and allow prediction of the flux distribution resulting in the maximal potential efficiency based on metabolic stoichiometry (Chen and Shachar-Hill, 2012). Comparison of the flux values predicted by FBA and MFA can help refine the accuracy of FBA. The results demonstrate that FBA using only the biomass constraint predicted substantially lower fluxes through glycolysis (wild-type and aplastidic cells on glucose) and the TCA cycle (on all selected substrates). This indicates an underestimation of cofactor production by FBA constrained only by the biomass composition. Incorporating maintenance costs, estimated based on the cofactor production deduced from the MFA results, greatly improves the predictive performance of FBA, especially for fluxes of glycolysis and the TCA cycle. Given a set of biomass and maintenance constraints, the flux maps show that FBA was able to accurately predict the majority of fluxes of the central metabolic pathways of heterotrophic metabolism of various carbon substrates using the transcriptome-based metabolic model. This indicates the potential of the network model to perform predictive modelling of the flux distributions in heterotrophic metabolism of *Euglena* grown on other organic carbon sources. However, the maintenance cost specifically analysed for each carbon substrate is needed for the accurate prediction.

Nevertheless, FBA has some limitations for predicting realistic fluxes through alternative pathways. In this study, the FBA of pyruvate grown cells predicted malic enzyme to be highly active to introduce pyruvate to the TCA, which is in contrast to the results of MFA, which showed that PC is responsible for this role. PC is predicted to be completely

inactive by the FBA. This inability to predict the flux of PC could be the result of using both maximising biomass production and minimising the sum of fluxes through the network as the objectives. In ethanol grown cells, FBA cannot predict the futile cycle of PEPC and malic enzyme, as demonstrated by ^{13}C MFA. These considerations exemplify the limitations of FBA for predicting the flux through alternative metabolic routes or futile cycles, previously reported in the FBA of *A. thaliana* (Cheung et al., 2015). In contrast, the MFA flux maps of ethanol, acetate and pyruvate can discriminate between these alternative routes and quantify some futile cycles.

6.4 Future directions

The metabolic model and the protocol for conducting ^{13}C MFA in *Euglena* established in this study can be used to extend studies of metabolism in *Euglena* under a broader range of conditions. The flux maps of acetate and ethanol demonstrate the different metabolic phenotypes when grown on the substrates that have the same carbon composition but differ in redox potential, a futile cycle observed in ethanol-grown cells might be responsible for assimilating the excess reductants generated. This aspect can be further investigated with other substrates with different carbon compositions to gain more insights on how *Euglena* responds to the different energy inputs. As this thesis has already demonstrated the flux map of pyruvate metabolism, exploring another C3 carbon compound like lactate could potentially provide a better understanding on how *E. gracilis* metabolically copes with the additional NADH generated during the reduction of lactate to pyruvate by lactate dehydrogenase (Jasso-Chávez et al., 2005). Previously, ^{13}CMFA has been used to examine how plant cells respond

to stresses or nutrient limitation (Masakapalli et al., 2014, Masakapalli et al., 2013, Masakapalli et al., 2010). Similar approaches could be used to investigate the metabolic response of *E. gracilis* to stress conditions. It has been reported that the osmotic and salt stresses affect the physiology of *Euglena* (Takenaka et al., 1997). These stresses were shown to perturb the distribution of metabolic fluxes and increase the maintenance cost in *Arabidopsis* cells (Cheung et al., 2013) and *Corynebacterium* (Varela et al., 2003). ^{13}C MFA may be used to investigate the metabolic *E. gracilis* under the osmotic stresses, this will provide the insights into the stress responses of *Euglena* and could help extending the understanding on how *Euglena* cells metabolically adapt to different energy demands.

The transcriptomic-based metabolic network model constructed in this thesis provided a substantial basis for FBA to predict fluxes distribution under heterotrophic conditions. However, model development is an iterative process that should be continuously refined to improve the prediction performance and capacity of the model. For example, the energy cost for nutrient uptake and intracellular transport of metabolites for *Euglena* should be incorporated in the model, which has been shown to improve the predictive performance of a plant cell model (Cheung et al., 2013). Furthermore, the model can be further curated to be able to predict the fluxes in autotrophic conditions, which may be done by adding the light reactions of photosynthesis. This will expand the scope of the model to predict the metabolic phenotype of *E. gracilis* under autotrophic and mixotrophic conditions. However, more extensive biochemical analysis and the estimation of maintenance cost for each specific condition of interest would be required to provide the optimal constraint for FBA in the presence of light.

Understanding the cellular metabolic physiology is essential for selecting targets for metabolic engineering. ^{13}C MFA and FBA have been demonstrated to be useful in systems metabolic engineering, as the techniques can reveal the metabolic phenotype and/or predict the metabolic changes by environmental perturbations (Lee et al., 2011, Kim et al., 2008). ^{13}C MFA was used to identify the strategy to enhance lipid production in *E. coli* (He et al., 2014) and *S. cerevisiae* (Ghosh et al., 2016). FBA of *Arabidopsis* based on a genome-scale model was used to identify the optimal metabolic flux modes for improving terpenoid production (Chung et al., 2013). This thesis is only the starting point for the further applications in metabolic engineering of *Euglena*. The tools and models for metabolic pathways analysis of *Euglena* developed in this thesis could be used in pathway engineering to optimise yields of desired products.

The analysis of metabolic network models using constraint-based approach, FBA, can be exploited to assist metabolic engineering. The predicted fluxes from FBA can help suggesting gene targets for knockout and/or overexpression that would enhance the biosynthesis of the valuable compounds in the desirable organisms (Orth et al., 2010, Curran et al., 2012). The metabolic model of *E. gracilis* developed in this thesis can be further curated and used for metabolic engineering to optimise the production of the high-value products, such as vitamin A, vitamin C, vitamin E and paramylon. In order to do that, apart from maximising growth, the use of different strategies for FBA, such as OptKnock , OptGene, and OptStrain (Curran et al., 2012), could be applied with this model to extend the application in metabolic engineering of *E. gracilis*.

E. gracilis has been studied as a source for biofuel production (Mahapatra et al., 2013, Toyama et al., 2019). To optimise the yield of lipid in *Euglena*, several targets can be considered. Flux maps on glucose show that about 37-41% of carbon intake was directly channelled to produced carbon storage (paramylon), and the activity of PDH is a major source for acetyl-CoA, the major precursor for lipid biosynthesis. To increase the flux through glycolysis to produce pyruvate for PDH, down regulating the activity of paramylon synthetase could be a potential. In addition, the flux map of pyruvate-grown cells demonstrates the importance of PC in channelling pyruvate to the TCA cycle and anaplerosis. Silencing PC to prevent these alternative routes could be considered to improve the lipid production in *Euglena*. The flux maps of *E. gracilis* on acetate and ethanol show that the reactions catalysed by malate synthase, isocitrate lyase and PEPCK carry high fluxes needed for channelling carbon into gluconeogenesis. These three enzymes are thus potential targets for down-regulation to limit acetyl-CoA consumption by the glyoxylate shunt and TCA cycle and thereby divert intermediates into the production of lipid in *E. gracilis* when grown on C2 substrates.

^{13}C MFA and FBA of *Euglena* in this study do not only provide insights into heterotrophic metabolism of carbon substrates, but also offer potentially powerful tools for rational metabolic engineering of *Euglena*. I hope that the insights into the operation of the metabolic network of *Euglena* presented in this thesis will stimulate investigation of metabolic flux in other non-model organisms, and be extended in the application of *Euglena* in the industrial sector.

References

ADL, S. M., SIMPSON, A. G. B., LANE, C. E., LUKEŠ, J., BASS, D., BOWSER, S. S., BROWN, M. W., BURKI, F., DUNTHORN, M., HAMPL, V., HEISS, A., HOPPENRATH, M., LARA, E., LE GALL, L., LYNN, D. H., MCMANUS, H., MITCHELL, E. A. D., MOZLEY-STANRIDGE, S. E., PARFREY, L. W., PAWLOWSKI, J., RUECKERT, S., SHADWICK, L., SCHOCH, C. L., SMIRNOV, A. & SPIEGEL, F. W. 2012. The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology,* 59**,** 429-514.

AHMADINEJAD, N., DAGAN, T. & MARTIN, W. 2007. Genome history in the symbiotic hybrid *Euglena gracilis*. *Gene,* 402**,** 35-39.

AHN, S., JUNG, J., JANG, I.-A., MADSEN, E. L. & PARK, W. 2016. Role of glyoxylate shunt in oxidative stress response. *Journal of Biological Chemistry,* 291**,** 11928-11938.

ALONSO, A. P., GOFFMAN, F. D., OHLROGGE, J. B. & SHACHAR-HILL, Y. 2007. Carbon conversion efficiency and central metabolic fluxes in developing sunflower (*Helianthus annuus* L.) embryos. *The Plant Journal,* 52**,** 296-308.

ANTHONY J.F. GRIFFITHS, U. O. B. C., SUSAN R. WESSLER, U. O. C. R., SEAN B. CARROLL, H. H. M. I. U. O. W.-M. & JOHN DOEBLEY, U. O. W.-M. 2015. *Introduction to Genetic Analysis*, Eleventh edition. New York, NY : W.H. Freeman & Company.

ANTONIEWICZ, M. R., KELLEHER, J. K. & STEPHANOPOULOS, G. 2006. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metabolic Engineering,* 8**,** 324-337.

ANTONIEWICZ, M. R., KELLEHER, J. K. & STEPHANOPOULOS, G. 2007. Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. *Analytical Chemistry,* 79**,** 7554-7559.

ARMBRUST, E. V., BERGES, J. A., BOWLER, C., GREEN, B. R., MARTINEZ, D., PUTNAM, N. H., ZHOU, S., ALLEN, A. E., APT, K. E., BECHNER, M., BRZEZINSKI, M. A., CHAAL, B. K., CHIOVITTI, A., DAVIS, A. K., DEMAREST, M. S., DETTER, J. C., GLAVINA, T., GOODSTEIN, D., HADI, M. Z., HELLSTEN, U., HILDEBRAND, M., JENKINS, B. D., JURKA, J., KAPITONOV, V. V., KRÖGER, N., LAU, W. W. Y., LANE, T. W., LARIMER, F. W., LIPPMEIER, J. C., LUCAS, S., MEDINA, M., MONTSANT, A., OBORNIK, M., PARKER, M. S., PALENIK, B., PAZOUR, G. J., RICHARDSON, P. M., RYNEARSON, T. A., SAITO, M. A., SCHWARTZ, D. C., THAMATRAKOLN, K., VALENTIN, K., VARDI, A., WILKERSON, F. P. & ROKHSAR, D. S. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science,* 306**,** 79.

BACHVAROFF, T. R., SANCHEZ PUERTA, M. V. & DELWICHE, C. F. 2005. Chlorophyll c–containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. *Molecular Biology and Evolution,* 22**,** 1772-1782.

BARSANTI, L., BASTIANINI, A., PASSARELLI, V., TREDICI, M. R. & GUALTIERI, P. 2000. Fatty acid content in wild type and WZSL mutant of *Euglena gracilis*. *Journal of Applied Phycology,* 12**,** 515-520.

BEN-SHAUL, Y. & OPHIR, I. 1970. Effects of streptomycin on plastids in dividing *Euglena*. *Planta,* 91**,** 195-203.

BLUM, J. J. & WITTELS, B. 1968. Mannose as a metabolite and an inhibitor of metabolism in *Euglena*. *Journal of Biological Chemistry,* 243**,** 200-210.

BOLTE, K., BULLMANN, L., HEMPEL, F., BOZARTH, A., ZAUNER, S. & MAIER, U.-G. 2009. Protein targeting into secondary plastids1. *Journal of Eukaryotic Microbiology,* 56**,** 9-15.

BORAH, K. 2017. *Metabolic Flux Analysis of the Nitrogen-fixing Bacterium Azorhizobium caulinodans.* University of Oxford.

BOWSHER, C. G., LACEY, A. E., HANKE, G. T., CLARKSON, D. T., SAKER, L. R., STULEN, I. & EMES, M. J. 2007. The effect of Glc6P uptake and its subsequent oxidation within pea root plastids on nitrite reduction and glutamate synthesis. *Journal of Experimental Botany,* 58**,** 1109-18.

BOYLE, N. R., SENGUPTA, N. & MORGAN, J. A. 2017. Metabolic flux analysis of heterotrophic growth in *Chlamydomonas reinhardtii*. *PLOS One,* 12**,** e0177292.

BRAWERMAN, G., HUFNAGEL, D. A. & CHARGAFF, E. 1962. On the nucleic acids of green and colorless *Euglena gracilis*: isolation and composition of deoxyribonucleic acid and of transfer ribonucleic acid. *Biochimica et Biophysica Acta,* 61**,** 340-345.

BUETOW, D. E. 1989. *The biology of Euglena,* New York ; London, Academic Press.

CALVAYRAC, R., LAVAL-MARTIN, D., BRIAND, J. & FARINEAU, J. 1981. Paramylon synthesis by *Euglena gracilis* photoheterotrophically grown under low O_2 pressure. *Planta,* 153**,** 6-13.

CHEN, X. & SHACHAR-HILL, Y. 2012. Insights into metabolic efficiency from flux analysis. *Journal of Experimental Botany,* 63**,** 2343-2351.

CHEUNG, C. Y. M., RATCLIFFE, R. G. & SWEETLOVE, L. J. 2015. A method of accounting for enzyme costs in flux balance analysis reveals alternative pathways and metabolite stores in an illuminated *Arabidopsis* leaf. *Plant Physiology,* 169**,** 1671-1682.

CHEUNG, C. Y. M., WILLIAMS, T. C. R., POOLMAN, M. G., FELL, D. A., RATCLIFFE, R. G. & SWEETLOVE, L. J. 2013. A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. *The Plant Journal,* 75**,** 1050-1061.

CHEW, S. Y., CHEE, W. J. Y. & THAN, L. T. L. 2019. The glyoxylate cycle and alternative carbon metabolism as metabolic adaptation strategies of *Candida glabrata*: perspectives from *Candida albicans* and *Saccharomyces cerevisiae*. *Journal of Biomedical Science,* 26**,** 52.

CHUNG, B. K.-S., LAKSHMANAN, M., KLEMENT, M., MOHANTY, B. & LEE, D.-Y. 2013. Genome-scale in silico modeling and analysis for designing synthetic terpenoid-producing microbial cell factories. *Chemical Engineering Science,* 103**,** 100-108.

CODD, G. A. & MERRETT, M. J. 1971. The regulation of glycolate metabolism in division synchronized cultures of *Euglena*. *Plant Physiology,* 47**,** 640.

COGNE, G., RÜGEN, M., BOCKMAYR, A., TITICA, M., DUSSAP, C.-G., CORNET, J.-F. & LEGRAND, J. 2011. A model-based method for investigating bioenergetic processes in autotrophically growing eukaryotic microalgae: application to the green algae *Chlamydomonas reinhardtii*. *Biotechnology Progress,* 27**,** 631-640.

CONSTANTOPOULOS, G. & BLOCH, K. 1967. Effect of light intensity on the lipid composition of *Euglena gracilis*. *Journal of Biological Chemistry,* 242**,** 3538-3542.

CROWN, S. B., LONG, C. P. & ANTONIEWICZ, M. R. 2016. Optimal tracers for parallel labeling experiments and ^{13}C metabolic flux analysis: a new precision and synergy scoring system. *Metabolic Engineering,* 38**,** 10-18.

CURRAN, K. A., CROOK, N. C. & ALPER, H. S. 2012. Using flux balance analysis to guide microbial metabolic engineering. *Methods in Molecular Biology* 834**,** 197-216.

CUSHMAN, J. C. & PRICE, C. A. 1986. Synthesis and turnover of proteins in proplastids and chloroplasts of *Euglena gracilis*. *Plant Physiology,* 82**,** 972.

DOLAN, S. K. & WELCH, M. 2018. The glyoxylate shunt, 60 years on. *Annual Review of Microbiology,* 72**,** 309-330.

DOUGLAS, S., ZAUNER, S., FRAUNHOLZ, M., BEATON, M., PENNY, S., DENG, L.-T., WU, X., REITH, M., CAVALIER-SMITH, T. & MAIER, U.-G. 2001. The highly reduced genome of an enslaved algal nucleus. *Nature,* 410**,** 1091.

DURNFORD, D. G. & GRAY, M. W. 2006. Analysis of *Euglena gracilis* plastid-targeted proteins reveals different classes of transit sequences. *Eukaryotic cell,* 5**,** 2079-2091.

DWYER, M. R. 1986. The soluble carbohydrates of *Euglena gracilis*—their identity and response to increased external solute concentration. *Plant, Cell & Environment,* 9**,** 485-489.

EBENEZER, T. E., CARRINGTON, M., LEBERT, M., KELLY, S. & FIELD, M. C. 2017. *Euglena gracilis* genome and transcriptome: organelles, nuclear genome assembly strategies and initial features. *In:* SCHWARTZBACH, S. D. & SHIGEOKA, S. (eds.) *Euglena: Biochemistry, Cell and Molecular Biology.* Cham: Springer International Publishing.

EBENEZER, T. E., ZOLTNER, M., BURRELL, A., NENAROKOVA, A., NOVÁK VANCLOVÁ, A. M. G., PRASAD, B., SOUKAL, P., SANTANA-MOLINA, C., O'NEILL, E., NANKISSOOR, N. N., VADAKEDATH, N., DAIKER, V., OBADO, S., SILVA-PEREIRA, S., JACKSON, A. P., DEVOS, D. P., LUKEŠ, J., LEBERT, M., VAUGHAN, S., HAMPL, V., CARRINGTON, M., GINGER, M. L., DACKS, J. B., KELLY, S. & FIELD, M. C. 2019. Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biology,* 17**,** 11.

EBRAHIM, A., LERMAN, J. A., PALSSON, B. O. & HYDUKE, D. R. 2013. COBRApy: Constraints-based reconstruction and analysis for python. *BMC Systems Biology,* 7**,** 74.

EINICKER-LAMAS, M., ANTUNES MEZIAN, G., BENEVIDES FERNANDES, T., SILVA, F. L. S., GUERRA, F., MIRANDA, K., ATTIAS, M. & OLIVEIRA, M. M. 2002. *Euglena gracilis* as a model for the study of Cu^{2+} and Zn^{2+} toxicity and accumulation in eukaryotic cells. *Environmental Pollution,* 120**,** 779-786.

EMANUELSSON, O., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols,* 2**,** 953.

EMANUELSSON, O., NIELSEN, H., BRUNAK, S. & VON HEIJNE, G. 2000. Predicting subcellular localization of proteins based on their n-terminal amino acid sequence. *Journal of Molecular Biology,* 300**,** 1005-1016.

FANG, X., LLOYD, C. J. & PALSSON, B. O. 2020. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology*.

FEIST, A. M. & PALSSON, B. O. 2010. The biomass objective function. *Current Opinion in Microbiology,* 13**,** 344-349.

FELDMAN, J. F. & BRUCE, V. G. 1972. Circadian rhythm changes in autotrophic *Euglena* induced by organic carbon sources. *The Journal of Protozoology,* 19**,** 370-373.

FIBLA, J. & GONZÀLEZ-DUARTE, R. 1993. Colorimetric assay to determine alcohol dehydrogenase activity. *Journal of Biochemical and Biophysical Methods,* 26**,** 87-93.

FOGUET, C., JAYARAMAN, A., MARIN, S., SELIVANOV, V. A., MORENO, P., MESSEGUER, R., DE ATAURI, P. & CASCANTE, M. 2019. p13CMFA: Parsimonious ^{13}C metabolic flux analysis. *PLOS Computational Biology,* 15**,** e1007310.

FOO, C. S., PETHE, K. & LUPIEN, A. 2020. Oxidative phosphorylation—an update on a new, essential target space for drug discovery in *Mycobacterium tuberculosis*. *Applied Sciences,* 10.

GEIMER, S., BELICOVÁ, A., LEGEN, J., SLÁVIKOVÁ, S., HERRMANN, R. G. & KRAJČOVIČ, J. 2009. Transcriptome analysis of the *Euglena gracilis* plastid chromosome. *Current Genetics,* 55**,** 425-438.

GHOSH, A., ANDO, D., GIN, J., RUNGUPHAN, W., DENBY, C., WANG, G., BAIDOO, E. E. K., SHYMANSKY, C., KEASLING, J. D. & GARCÍA MARTÍN, H. 2016. ^{13}C metabolic flux analysis for systematic metabolic engineering of *S. cerevisiae* for overproduction of fatty acids. *Frontiers in Bioengineering and Biotechnology,* 4.

GIBBS, S. P. 1978. The chloroplasts of *Euglena* may have evolved from symbiotic green algae. *Canadian Journal of Botany,* 56**,** 2883-2889.

GLEASON, F. K. & HOGENKAMP, H. P. C. 1970. Ribonucleotide reductase from *Euglena gracilis*, a deoxyadenosylcobalamin-dependent enzyme. *Journal of Biological Chemistry,* 245**,** 4894-4899.

GONZALEZ, J. E., LONG, C. P. & ANTONIEWICZ, M. R. 2017. Comprehensive analysis of glucose and xylose metabolism in *Escherichia coli* under aerobic and anaerobic conditions by ^{13}C metabolic flux analysis. *Metabolic Engineering,* 39**,** 9-18.

GOODPASTER, B. H. & SPARKS, L. M. 2017. Metabolic flexibility in health and disease. *Cell Metabolism,* 25**,** 1027-1036.

GÖRGENS, J. F., VAN ZYL, W. H. & KNOETZE, J. H. 2005. Reliability of methods for the determination of specific substrate consumption rates in batch culture. *Biochemical Engineering Journal,* 25**,** 109-112.

GRAVES JR, L. B. & BECKER, W. M. 1974. Beta-oxidation in glyoxysomes from *Euglena*. *The Journal of Protozoology,* 21**,** 771-774.

GRAVES JR, L. B., TRELEASE, R. N., GRILL, A. & BECKER, W. M. 1972. Localization of glyoxylate cycle enzymes in glyoxysomes in *Euglena*. *The Journal of Protozoology,* 19**,** 527-532.

GRAY, L. R., TOMPKINS, S. C. & TAYLOR, E. B. 2014. Regulation of pyruvate metabolism and human disease. *Cellular and Molecular Life Sciences* 71**,** 2577-2604.

GUO, W., SHENG, J. & FENG, X. 2015. ^{13}C-Metabolic flux analysis: an accurate approach to demystify microbial metabolism for biochemical production. *Bioengineering,* 3**,** 3.

HADARIOVÁ, L., VESTEG, M., BIRČÁK, E., SCHWARTZBACH, S. D. & KRAJČOVIČ, J. 2017. An intact plastid genome is essential for the survival of colorless *Euglena longa* but not *Euglena gracilis*. *Current Genetics,* 63**,** 331-341.

HALLICK, R. B., HONG, L., DRAGER, R. G., FAVREAU, M. R., MONFORT, A., ORSAT, B., SPIELMANN, A. & STUTZ, E. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Research,* 21**,** 3537-3544.

HALTER, D., ANDRES, J., PLEWNIAK, F., POULAIN, J., DA SILVA, C., ARSÈNE-PLOETZE, F. & BERTIN, P. N. 2015. Arsenic hypertolerance in the protist *Euglena mutabilis* is mediated by specific transporters and functional integrity maintenance mechanisms. *Environmental Microbiology,* 17**,** 1941-1949.

HARRISON, P. W. & KRUGER, N. J. 2008. Validation of the design of feeding experiments involving [^{14}C]substrates used to monitor metabolic flux in higher plants. *Phytochemistry,* 69**,** 2920-2927.

HASAN, M. T., SUN, A., MIRZAEI, M., TE'O, J., HOBBA, G., SUNNA, A. & NEVALAINEN, H. 2017. A comprehensive assessment of the biosynthetic pathways of ascorbate, α-tocopherol and free amino acids in *Euglena gracilis* var. saccharophila. *Algal Research,* 27**,** 140-151.

HAYAKAWA, K., MATSUDA, F. & SHIMIZU, H. 2018. ^{13}C-metabolic flux analysis of ethanol-assimilating *Saccharomyces cerevisiae* for S-adenosyl-l-methionine production. *Microbial Cell Factories,* 17**,** 82.

HE, L., XIAO, Y., GEBRESELASSIE, N., ZHANG, F., ANTONIEWIEZ, M. R., TANG, Y. J. & PENG, L. 2014. Central metabolic responses to the overproduction of fatty acids in *Escherichia coli* based on ^{13}C-metabolic flux analysis. *Biotechnology and Bioengineering,* 111**,** 575-85.

HORTON, P., PARK, K.-J., OBAYASHI, T., FUJITA, N., HARADA, H., ADAMS-COLLIER, C. J. & NAKAI, K. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Research,* 35**,** W585-W587.

HURLBERT, R. E. & RITTENBERG, S. C. 1962. Glucose metabolism of *Euglena gracilis* var. Bacillaris; growth and enzymatic studies. *The Journal of Protozoology,* 9**,** 170-182.

INUI, H., MIYATAKE, K., NAKANO, Y. & KITAOKA, S. 1982. Wax ester fermentation in *Euglena gracilis*. *Federation of European Biochemical Societies Letters,* 150**,** 89-93.

INWONGWAN, S., KRUGER, N. J., RATCLIFFE, R. G. & O'NEILL, E. C. 2019. *Euglena* central metabolic pathways and their subcellular locations. *Metabolites,* 9**,** 115.

ISEGAWA, Y., NAKANO, Y. & KITAOKA, S. 1984. Submitochondrial location and some properties of NAD^+- and $NADP^+$-linked malate dehydrogenase in *Euglena*. *Agricultural and Biological Chemistry,* 48**,** 549-552.

IVUŠIĆ, F. & ŠANTEK, B. 2015. Optimization of complex medium composition for heterotrophic cultivation of *Euglena gracilis* and paramylon production. *Bioprocess and Biosystems Engineering,* 38**,** 1103-1112.

IYER, V. V., SRIRAM, G., FULTON, D. B., ZHOU, R., WESTGATE, M. E. & SHANKS, J. V. 2008. Metabolic flux maps comparing the effect of temperature on protein and oil biosynthesis in developing soybean cotyledons. *Plant, Cell & Environment,* 31**,** 506-517.

JACKSON, C., KNOLL, A. H., CHAN, C. X. & VERBRUGGEN, H. 2018. Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of secondary green plastids. *Scientific Reports,* 8**,** 1523.

JASSO-CHÁVEZ, R., GARCÍA-CANO, I., MARÍN-HERNÁNDEZ, Á., MENDOZA-CÓZATL, D., RENDÓN, J. L. & MORENO-SÁNCHEZ, R. 2005. The bacterial-like lactate shuttle

components from heterotrophic *Euglena gracilis*. *Biochimica et Biophysica Acta,* 1709**,** 181-190.

JITRAPAKDEE, S., ST MAURICE, M., RAYMENT, I., CLELAND, W. W., WALLACE, JOHN C. & ATTWOOD, PAUL V. 2008. Structure, mechanism and regulation of pyruvate carboxylase. *Biochemical Journal,* 413**,** 369-387.

KATO, S. 1982. Laboratory culture and morphology of colacium vesiculosum euglenophyceae. *Japanese Journal of Phycology,* 30**,** 63-67.

KAUFFMAN, K. J., PRAKASH, P. & EDWARDS, J. S. 2003. Advances in flux balance analysis. *Current Opinion in Biotechnology,* 14**,** 491-496.

KEMPNER, E. S. & MILLER, J. H. 1965a. The molecular biology of *Euglena gracilis* I. Growth conditions and cellular composition. *Biochimica et Biophysica Acta,* 104**,** 11-17.

KEMPNER, E. S. & MILLER, J. H. 1965b. The molecular biology of *Euglena gracilis*. III. general carbon metabolism. *Biochemistry,* 4**,** 2735-2739.

KIM, H. U., KIM, T. Y. & LEE, S. Y. 2008. Metabolic flux analysis and metabolic engineering of microorganisms. *Molecular Omics,* 4**,** 113-20.

KITAOKA, S., NAKANO, Y., MIYATAKE, K. & YOKOTA, A. 1989. Enzymes and their functional location *In:* BUETOW, D. E. (ed.) *The Biology of Euglena.* London: Academic Press Limited.

KÖHLER, P. 1985. The strategies of energy conservation in helminths. *Molecular and Biochemical Parasitology,* 17**,** 1-18.

KOIZUMI, N., SAKAGAMI, H., UTSUMI, A., FUJINAGA, S., TAKEDA, M., ASANO, K., SUGAWARA, I., ICHIKAWA, S., KONDO, H., MORI, S., MIYATAKE, K., NAKANO, Y., NAKASHIMA, H.,

MURAKAMI, T., MIYANO, N. & YAMAMOTO, N. 1993. Anti-HIV (human immunodeficiency virus) activity of sulfated paramylon. *Antiviral Research,* 21**,** 1-14.

KONDO, Y., KATO, A., HOJO, H., NOZOE, S., TAKEUCHI, M. & OCHI, K. 1992. Cytokine-related immunopotentiating activities of paramylon, a β-(1-3)-D-glucan from *Euglena gracilis*. *Journal of Pharmacobio-Dynamics,* 15**,** 617-621.

KORN, E. D. 1964. The fatty acids of *Euglena gracilis*. *Journal of Lipid Research,* 5**,** 352-362.

KRISTOFFERSEN, L., SKUTERUD, B., LARSSEN, B. R., SKURTVEIT, S. & SMITH-KIELLAND, A. 2005. Fast quantification of ethanol in whole blood specimens by the enzymatic alcohol dehydrogenase method. Optimization by experimental design. *Journal of Analytical Toxicology,* 29**,** 66-70.

KRUGER, N. J., HUDDLESTON, J. E., LE LAY, P., BROWN, N. D. & RATCLIFFE, R. G. 2007. Network flux analysis: Impact of ^{13}C-substrates on metabolism in *Arabidopsis thaliana* cell suspension cultures. *Phytochemistry,* 68**,** 2176-2188.

KRUGER, N. J., MASAKAPALLI, S. K. & RATCLIFFE, R. G. 2012. Strategies for investigating the plant metabolic network with steady-state metabolic flux analysis: lessons from an *Arabidopsis* cell culture and other systems. *Journal of Experimental Botany,* 63**,** 2309-23.

KRUGER, NICHOLAS J. & RATCLIFFE, R. G. 2015. Fluxes through plant metabolic networks: measurements, predictions, insights and challenges. *Biochemical Journal,* 465**,** 27-38.

KUNZE, M., PRACHAROENWATTANA, I., SMITH, S. M. & HARTIG, A. 2006. A central role for the peroxisomal membrane in glyoxylate cycle function. *Biochimica et Biophysica Acta,* 1763**,** 1441-1452.

KUSMIC, C., BARSACCHI, R., BARSANTI, L., GUALTIERI, P. & PASSARELLI, V. 1998. *Euglena gracilis* as source of the antioxidant vitamin E. effects of culture conditions in the wild strain and in the natural mutant WZSL. *Journal of Applied Phycology,* 10**,** 555-559.

LAURENS, L. M. L., DEMPSTER, T. A., JONES, H. D. T., WOLFRUM, E. J., VAN WYCHEN, S., MCALLISTER, J. S. P., RENCENBERGER, M., PARCHERT, K. J. & GLOE, L. M. 2012. Algal biomass constituent analysis: method uncertainties and investigation of the underlying measuring chemistries. *Analytical Chemistry,* 84**,** 1879-1887.

LAVAL-MARTIN, D., FARINEAU, J., PINEAU, B. & CALVAYRAC, R. 1981. Evolution of enzymes involved in carbon metabolism (phosphoenolpyruvate and ribulose-bisphosphate carboxylases, phosphoenolpyruvate carboxykinase) during the light-induced greening of *Euglena gracilis* strains Z and ZR. *Planta,* 151**,** 157-167.

LEE, S. Y., PARK, J. M. & KIM, T. Y. 2011. Application of metabolic flux analysis in metabolic engineering. *Methods in Enzymology,* 498**,** 67-93.

LEVASSEUR, M., THOMPSON, P. A. & HARRISON, P. J. 1993. Physiological acclimation of marine phytoplankton to different nitrogen sources. *Journal of Phycology,* 29**,** 587-595.

LEWIS, N. E., HIXSON, K. K., CONRAD, T. M., LERMAN, J. A., CHARUSANTI, P., POLPITIYA, A. D., ADKINS, J. N., SCHRAMM, G., PURVINE, S. O., LOPEZ-FERRER, D., WEITZ, K. K., EILS, R., KÖNIG, R., SMITH, R. D. & PALSSON, B. Ø. 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology,* 6**,** 390.

LI, M., GAO, X., WU, B., QIAN, X., GIESY, J. P. & CUI, Y. 2014. Microalga *Euglena* as a bioindicator for testing genotoxic potentials of organic pollutants in Taihu Lake, China. *Ecotoxicology,* 23**,** 633-640.

LINTON, E. W., KARNKOWSKA-ISHIKAWA, A., KIM, J. I., SHIN, W., BENNETT, M. S., KWIATOWSKI, J., ZAKRYŚ, B. & TRIEMER, R. E. 2010. Reconstructing euglenoid evolutionary relationships using three genes: nuclear SSU and LSU, and chloroplast SSU rDNA sequences and the description of *Euglenaria* gen. nov. (Euglenophyta). *Protist,* 161**,** 603-619.

LIU, N., QIAO, K. & STEPHANOPOULOS, G. 2016. ^{13}C metabolic flux analysis of acetate conversion to lipids by *Yarrowia lipolytica*. *Metabolic Engineering,* 38**,** 86-97.

LOWRY, O. H., ROSEBROUGH, N. J., FARR, A. L. & RANDALL, R. J. 1951. Protein measurement with the folin phenol reagent. *Journal of Biological Chemistry,* 193**,** 265-75.

MAHAPATRA, D. M., CHANAKYA, H. N. & RAMACHANDRA, T. V. 2013. *Euglena* sp. as a suitable source of lipids for potential use as biofuel and sustainable wastewater treatment. *Journal of Applied Phycology,* 25**,** 855-865.

MAIER, U. G., ZAUNER, S. & HEMPEL, F. 2015. Protein import into complex plastids: cellular organization of higher complexity. *European Journal of Cell Biology,* 94**,** 340-348.

MALEK, L., BOGORAD, L., AYERS, A. R. & GOLDBERG, A. L. 1984. Newly synthesized proteins are degraded by an ATP-stimulated proteolytic process in isolated pea chloroplasts. *FEBS Letters,* 166**,** 253-257.

MARÉCHAL, L. R. & BELOCOPITOW, E. 1972. Metabolism of trehalose in *Euglena gracilis* : I. partial purification and some properties of trehalose phosphorylase. *Journal of Biological Chemistry,* 247**,** 3223-3228.

MARTIN, W. & HERRMANN, R. G. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiology,* 118**,** 9.

MARTIN, W., SOMERVILLE, C. C. & LOISEAUX-DE GOËR, S. 1992. Molecular phylogenies of plastid origins and algal evolution. *Journal of Molecular Evolution,* 35**,** 385-404.

MARUYAMA, S., SUZAKI, T., WEBER, A. P., ARCHIBALD, J. M. & NOZAKI, H. 2011a. Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evolutionary Biology,* 11**,** 105.

MARUYAMA, S., SUZAKI, T., WEBER, A. P. M., ARCHIBALD, J. M. & NOZAKI, H. 2011b. Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evolutionary Biology,* 11**,** 105.

MASAKAPALLI, S. K. 2011. *Network Flux Analysis of Central Metabolism in Plants.* University of Oxford.

MASAKAPALLI, S. K., BRYANT, F. M., KRUGER, N. J. & RATCLIFFE, R. G. 2014. The metabolic flux phenotype of heterotrophic *Arabidopsis* cells reveals a flexible balance between the cytosolic and plastidic contributions to carbohydrate oxidation in response to phosphate limitation. *The Plant Journal,* 78**,** 964-977.

MASAKAPALLI, S. K., KRUGER, N. J. & RATCLIFFE, R. G. 2013. The metabolic flux phenotype of heterotrophic *Arabidopsis* cells reveals a complex response to changes in nitrogen supply. *The Plant Journal,* 74**,** 569-582.

MASAKAPALLI, S. K., LE LAY, P., HUDDLESTON, J. E., POLLOCK, N. L., KRUGER, N. J. & RATCLIFFE, R. G. 2010. Subcellular flux analysis of central metabolism in a heterotrophic *Arabidopsis* cell suspension using steady-state stable isotope labeling. *Plant Physiology,* 152**,** 602-619.

MCATEE, A. G., JAZMIN, L. J. & YOUNG, J. D. 2015. Application of isotope labeling experiments and ^{13}C flux analysis to enable rational pathway engineering. *Current Opinion in Biotechnology,* 36**,** 50-56.

MILLER, J. H. & KEMPNER, E. S. 1976. The molecular biology of *Euglena gracilis*. X. amino acid composition of protein. *The Journal of Protozoology,* 23**,** 444-446.

MILLER, M. E., JURGENSON, J. E., REARDON, E. M. & PRICE, C. A. 1983. Plastid translation in organello and in vitro during light-induced development in *Euglena*. *Journal of Biological Chemistry,* 258**,** 14478-14484.

MISHRA, S. K., SUH, W. I., FAROOQ, W., MOON, M., SHRIVASTAV, A., PARK, M. S. & YANG, J.-W. 2014. Rapid quantification of microalgal lipids in aqueous medium by a simple colorimetric method. *Bioresource Technology,* 155**,** 330-333.

MIYATAKE, K., ENOMOTO, T. & KITAOKA, S. 1984a. Detection and subcellular distribution of pyrophosphate: D-fructose 6- phosphate phosphotransferase (PFP) in *Euglena gracilis*. *Agricultural and Biological Chemistry,* 48**,** 2857-2859.

MIYATAKE, K., ITO, T. & KITAOKA, S. 1984b. Subcellular location and some properties of phosphoenolpyruvate carboxykinase (PEPCK) in *Euglena gracilis*. *Agricultural and Biological Chemistry,* 48**,** 2139-2141.

MIYATAKE, K., WASHIO, K., YOKOTA, A., NAKANO, Y. & KITAOKA, S. 1985. Occurrence of two NAD+-linked malate dehydrogenase isozymes in *Euglena gracilis*. *Agricultural and Biological Chemistry,* 49**,** 859-860.

MORIBER, L. G., HERSHENOV, B., AARONSON, S. & BENSKY, B. 1963. Teratological chloroplast structures in *Euglena gracilis* permanently bleached by exogenous physical and chemical agents. *The Journal of Protozoology,* 10**,** 80-86.

MURIMA, P., ZIMMERMANN, M., CHOPRA, T., POJER, F., FONTI, G., DAL PERARO, M., ALONSO, S., SAUER, U., PETHE, K. & MCKINNEY, J. D. 2016. A rheostat mechanism governs the bifurcation of carbon flux in mycobacteria. *Nature Communications,* 7**,** 12527.

NAKAZAWA, M. 2017. C2 metabolism in *Euglena*. *Advances in Experimental Medicine and Biology,* 979**,** 39-45.

NAKAZAWA, M., HAYASHI, R., TAKENAKA, S., INUI, H., ISHIKAWA, T., UEDA, M., SAKAMOTO, T., NAKANO, Y. & MIYATAKE, K. 2017. Physiological functions of pyruvate: $NADP^+$ oxidoreductase and 2-oxoglutarate decarboxylase in *Euglena gracilis* under aerobic and anaerobic conditions. *Bioscience, Biotechnology, and Biochemistry,* 81**,** 1386-1393.

NAVARRO, L., TORRES-MÁRQUEZ, M. A. E., GONZÁLEZ-MORENO, S., DEVARS, S., HERNÁNDEZ, R. & MORENO-SÁNCHEZ, R. 1997. Comparison of physiological changes in *Euglena gracilis* during exposure to heavy metals of heterotrophic and autotrophic cells. *Comparative Biochemistry and Physiology* 116**,** 265-272.

NIELSEN, J. 2003. It is all about metabolic fluxes. *Journal of Bacteriology,* 185**,** 7031-7035.

NOVÁK VANCLOVÁ, A. M. G., ZOLTNER, M., KELLY, S., SOUKAL, P., ZÁHONOVÁ, K., FÜSSY, Z., EBENEZER, T. E., LACOVÁ DOBÁKOVÁ, E., ELIÁŠ, M., LUKEŠ, J., FIELD, M. C. & HAMPL, V. 2019. Proteome of the secondary plastid of *Euglena gracilis* reveals metabolic quirks and colourful history. *bioRxiv***,** 573709.

NOVÁK VANCLOVÁ, A. M. G., ZOLTNER, M., KELLY, S., SOUKAL, P., ZÁHONOVÁ, K., FÜSSY, Z., EBENEZER, T. E., LACOVÁ DOBÁKOVÁ, E., ELIÁŠ, M., LUKEŠ, J., FIELD, M. C. & HAMPL, V. 2020. Metabolic quirks and the colourful history of the *Euglena gracilis* secondary plastid. *New Phytologist,* 225**,** 1578-1592.

O'NEILL, E. C., TRICK, M., HILL, L., REJZEK, M., DUSI, R. G., HAMILTON, C. J., ZIMBA, P. V., HENRISSAT, B. & FIELD, R. A. 2015. The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Molecular BioSystems,* 11**,** 2808-2820.

O'NEILL, E. C., TRICK, M., HENRISSAT, B. & FIELD, R. A. 2015. *Euglena* in time: evolution, control of central metabolic processes and multi-domain proteins in carbohydrate and natural product biochemistry. *Perspectives in Science,* 6**,** 84-93.

ODA, Y., MIYATAKE, K., NAKANO, Y. & KITAOKA, S. 1981. Subcellular location and some properties of isocitrate dehydrogenase isozymes in *Euglena gracilis*. *Agricultural and Biological Chemistry,* 45**,** 2619-2621.

ODA, Y., NAKANO, Y. & KITAOKA, S. 1982. Utilization and toxicity of exogenous amino acids in *Euglena gracilis*. *Microbiology,* 128**,** 853-858.

OGAWA, T., KIMURA, A., SAKUYAMA, H., TAMOI, M., ISHIKAWA, T. & SHIGEOKA, S. 2015. Identification and characterization of cytosolic fructose-1,6-bisphosphatase in *Euglena gracilis*. *Bioscience, Biotechnology, and Biochemistry,* 79**,** 1957-1964.

OGBONNA, J. C., TOMIYAMA, S. & TANAKA, H. 1998. Heterotrophic cultivation of *Euglena gracilis* Z for efficient production of alpha-tocopherol. *Journal of Applied Phycology,* 10**,** 67-74.

ONO, K., KAWANAKA, Y., IZUMI, Y., INUI, H., MIYATAKE, K., KITAOKA, S. & NAKANO, Y. 1995. Mitochondrial alcohol dehydrogenase from ethanol-grown *Euglena gracilis*. *The Journal of Biochemistry,* 117**,** 1178-1182.

ONO, K., KONDO, M., OSAFUNE, T., MIYATAKE, K., INUI, H., KITAOKA, S., NISHIMURA, M. & NAKANO, Y. 2003. Presence of glyoxylate cycle enzymes in the mitochondria of *Euglena gracilis*. *Journal of Eukaryotic Microbiology,* 50**,** 92-96.

ORATA, F. 2012. Derivatization reactions and reagents for gas chromatography analysis. *Advanced Gas Chromatography - Progress in Agricultural, Biomedical and Industrial Applications.*

ORTH, J. D., THIELE, I. & PALSSON, B. Ø. 2010. What is flux balance analysis? *Nature Biotechnology,* 28**,** 245-248.

PARFREY, L. W., LAHR, D. J. G., KNOLL, A. H. & KATZ, L. A. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences,* 108**,** 13624-13629.

PATRA, K. C. & HAY, N. 2014. The pentose phosphate pathway and cancer. *Trends in Biochemical Sciences,* 39**,** 347-354.

PEAK, M. J., PEAK, J. G. & TING, I. P. 1972. Isoenzymes of malate dehydrogenase and their regulation in *Euglena gracilis* Z. *Biochimica et Biophysica Acta,* 284**,** 1-15.

PONTIUS, R. G., THONTTEH, O. & CHEN, H. 2008. Components of information for multiple resolution comparison between maps that share a real variable. *Environmental and Ecological Statistics,* 15**,** 111-142.

PORTAIS, J.-C. & DELORT, A.-M. 2002. Carbohydrate cycling in micro-organisms: what can ^{13}C-NMR tell us? *FEMS Microbiology Reviews,* 26**,** 375-402.

POSKAR, C. H., HUEGE, J., KRACH, C., FRANKE, M., SHACHAR-HILL, Y. & JUNKER, B. H. 2012. iMS2Flux--a high-throughput processing tool for stable isotope labeled mass spectrometric data used for metabolic flux analysis. *BMC Bioinformatics,* 13**,** 295-295.

PRIGENT, S., FRIOUX, C., DITTAMI, S. M., THIELE, S., LARHLIMI, A., COLLET, G., GUTKNECHT, F., GOT, J., EVEILLARD, D., BOURDON, J., PLEWNIAK, F., TONON, T. & SIEGEL, A. 2017. Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. *PLOS Computational Biology,* 13**,** e1005276.

PROVASOLI, L., HUTNER, S. H. & SCHATZ, A. 1948. Streptomycin-induced chlorophyll-less races of *Euglena*. *Proceedings of the Society for Experimental Biology and Medicine,* 69**,** 279-282.

QIN, H., GUO, Q., LIU, C., LI, F., ZHANG, H., CHU, Z., WANG, J. & LEI, A. 2020. Occurrence and light response of residual plastid genes in a *Euglena gracilis* bleached mutant strain OflB2. *Journal of Oceanology and Limnology*.

RATCLIFFE, R. G. & SHACHAR-HILL, Y. 2005. Revealing metabolic phenotypes in plants: inputs from NMR analysis. *Biological Reviews,* 80**,** 27-43.

REGEV-RUDZKI, N., KARNIELY, S., BEN-HAIM, N. N. & PINES, O. 2005. Yeast aconitase in two locations and two metabolic pathways: seeing small amounts is believing. *Molecular Biology of the Cell,* 16**,** 4163-71.

REGNAULT, A., CHERVIN, D., CHAMMAI, A., PITON, F., CALVAYRAC, R. & MAZLIAK, P. 1995. Lipid composition of *Euglena gracilis* in relation to carbon-nitrogen balance. *Phytochemistry,* 40**,** 725-733.

REINBOTHE, C., ORTEL, B., PARTHIER, B. & REINBOTHE, S. 1994. Cytosolic and plastid forms of 5-enolpyruvylshikimate-3-phosphate synthase in *Euglena gracilis* are differentially expressed during light-induced chloroplast development. *Molecular and General Genetics,* 245**,** 616-622.

RODRÍGUEZ-ZAVALA, J. S., ORTIZ-CRUZ, M. A. & MORENO-SANCHEZ, R. 2006. Characterization of an aldehyde dehydrogenase from *Euglena gracilis*. *Journal of Eukaryotic Microbiology,* 53**,** 36-42.

ROSENBERG, A. 1963. A comparison of lipid patterns in photosynthesizing and nonphotosynthesizing cells of *Euglena gracilis*. *Biochemistry,* 2**,** 1148-1154.

SANDBERG, T. E., LONG, C. P., GONZALEZ, J. E., FEIST, A. M., ANTONIEWICZ, M. R. & PALSSON, B. O. 2016. Evolution of *E. coli* on [U-^{13}C]glucose reveals a negligible isotopic influence on metabolism and physiology. *PLOS One,* 11**,** e0151130.

SATO, S. 2011. The apicomplexan plastid and its evolution. *Cellular and Molecular Life Sciences,* 68**,** 1285-1296.

SCHMIDT, K., NIELSEN, J. & VILLADSEN, J. 1999. Quantitative analysis of metabolic fluxes in *Escherichia coli*, using two-dimensional NMR spectroscopy and complete isotopomer models. *Journal of Biotechnology,* 71**,** 175-189.

SCHNARRENBERGER, C. & MARTIN, W. 2002. Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. *European Journal of Biochemistry,* 269**,** 868-883.

SCHWENDER, J., OHLROGGE, J. & SHACHAR-HILL, Y. 2004. Understanding flux in plant metabolic networks. *Current Opinion in Plant Biology,* 7**,** 309-317.

SCOTT, M. S., CALAFELL, S. J., THOMAS, D. Y. & HALLETT, M. T. 2005. Refining protein subcellular localization. *PLOS Computational Biology,* 1**,** e66-e66.

SEAVER, S. M. D., HENRY, C. S. & HANSON, A. D. 2012. Frontiers in metabolic reconstruction and modeling of plant genomes. *Journal of Experimental Botany,* 63**,** 2247-2258.

SHAMEER, S., RATCLIFFE, R. G. & SWEETLOVE, L. J. 2019. Leaf energy balance requires mitochondrial respiration and export of chloroplast NADPH in the light. *Plant Physiology,* 180**,** 1947.

SHAMEER, S., VALLARINO, J. G., FERNIE, A. R., RATCLIFFE, R. G. & SWEETLOVE, L. J. 2020. Flux balance analysis of metabolism during growth by osmotic cell expansion and its application to tomato fruits. *The Plant Journal,* 103**,** 68-82.

SHAO, Q., HU, L., QIN, H., LIU, Y., TANG, X., LEI, A. & WANG, J. 2019. Metabolomic response of *Euglena gracilis* and its bleached mutant strain to light. *PLOS One,* 14**,** e0224926.

SHIBATA, S., ARIMURA, S.-I., ISHIKAWA, T. & AWAI, K. 2018. Alterations of membrane lipid content correlated with chloroplast and mitochondria development in *Euglena gracilis*. *Frontiers In Plant Science,* 9**,** 370-370.

SLÁVIKOVÁ, S., VACULA, R., FANG, Z., EHARA, T., OSAFUNE, T. & SCHWARTZBACH, S. D. 2005. Homologous and heterologous reconstitution of golgi to chloroplast transport and protein import into the complex chloroplasts of *Euglena*. *Journal of Cell Science,* 118**,** 1651.

SOKOL, S., MILLARD, P. & PORTAIS, J.-C. 2011. influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments. *Bioinformatics,* 28**,** 687-693.

TAKENAKA, S., KONDO, T., NAZERI, S., TAMURA, Y., TOKUNAGA, M., TSUYAMA, S., MIYATAKE, K. & NAKANO, Y. 1997. Accumulation of trehalose as a compatible solute under osmotic stress in *Euglena gracilis* Z. *Journal of Eukaryotic Microbiology,* 44**,** 609-613.

TAKEYAMA, H., KANAMARU, A., YOSHINO, Y., KAKUTA, H., KAWAMURA, Y. & MATSUNAGA, T. 1997. Production of antioxidant vitamins, beta-carotene, vitamin C, and vitamin E, by two-step culture of *Euglena gracilis* Z. *Biotechnology and Bioengineering,* 53**,** 185-90.

THIELE, I. & PALSSON, B. Ø. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols,* 5**,** 93-121.

TORRUELLA, G., SUGA, H., RIUTORT, M., PERETÓ, J. & RUIZ-TRILLO, I. 2009. The evolutionary history of lysine biosynthesis pathways within eukaryotes. *Journal of Molecular Evolution,* 69**,** 240-248.

TOYAMA, T., HANAOKA, T., YAMADA, K., SUZUKI, K., TANAKA, Y., MORIKAWA, M. & MORI, K. 2019. Enhanced production of biomass and lipids by *Euglena gracilis* via co-culturing with a microalga growth-promoting bacterium, *Emticicia* sp. EG3. *Biotechnology for Biofuels,* 12**,** 205.

TUMANOV, S., BULUSU, V., GOTTLIEB, E. & KAMPHORST, J. J. 2016. A rapid method for quantifying free and bound acetate based on alkylation and GC-MS analysis. *Cancer & Metabolism,* 4**,** 17.

TURMEL, M., GAGNON, M.-C., O'KELLY, C. J., OTIS, C. & LEMIEUX, C. 2009. The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Molecular Biology and Evolution,* 26**,** 631-648.

UTTER, M. F. & KEECH, D. B. 1960. Formation of oxaloacetate from pyruvate and carbon dioxide. *Journal of Biological Chemistry,* 235**,** Pc17-8.

VARELA, C., AGOSIN, E., BAEZ, M., KLAPA, M. & STEPHANOPOULOS, G. 2003. Metabolic flux redistribution in *Corynebacterium glutamicum* in response to osmotic stress. *Applied Microbiology and Biotechnology,* 60**,** 547-555.

WAHL, S. A., DAUNER, M. & WIECHERT, W. 2004. New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. *Biotechnology and Bioengineering,* 85**,** 259-268.

WALKER, J. R. L. 1992. Spectrophotometric determination of enzyme activity: Alcohol dehydrogenase (ADH). *Biochemical Education,* 20**,** 42-43.

WANG, J., SHI, Z. & XU, X. 2004. Residual plastids of bleached mutants of *Euglena gracilis* and their effects on the expression of nucleus-encoded genes. *Progress in Natural Science,* 14**,** 213-217.

WANG, L., BEHR, S. R., NEWMAN, R. K. & NEWMAN, C. W. 1997. Comparative cholesterol-lowering effects of barley β-glucan and barley oil in golden syrian hamsters. *Nutrition Research,* 17**,** 77-88.

WANG, Y., SEPPÄNEN-LAAKSO, T., RISCHER, H. & WIEBE, M. G. 2018. *Euglena gracilis* growth and cell composition under different temperature, light and trophic conditions. *PLOS One,* 13**,** e0195329.

WEXLER, I. D., DU, Y., LISGARIS, M. V., MANDAL, S. K., FREYTAG, S. O., YANG, B.-S., LIU, T.-C., KWON, M., PATEL, M. S. & KERR, D. S. 1994. Primary amino acid sequence and structure of human pyruvate carboxylase. *Biochimica et Biophysica Acta,* 1227**,** 46-52.

WIECHERT, W. 2001. ^{13}C metabolic flux analysis. *Metabolic Engineering,* 3**,** 195-206.

WILLIAMS, T. C. R., POOLMAN, M. G., HOWDEN, A. J. M., SCHWARZLANDER, M., FELL, D. A., RATCLIFFE, R. G. & SWEETLOVE, L. J. 2010. A genome-scale metabolic model accurately predicts fluxes in central carbon metabolism under stress conditions. *Plant Physiology,* 154**,** 311-323.

WILSON, B. W., BUETOW, D. E., JAHN, T. L. & LEVEDAHL, B. H. 1959. A differential effect of pH on cell growth and respiration. *Experimental Cell Research,* 18**,** 454-465.

WOLKEN, J. J. 1961. *Euglena; an Experimental Organism for Biochemical and Biophysical Studies,* New Brunswick, Institute of Microbiology, Rutgers.

WOODS, C. A. 2014. *Respiratory Carbon Loss in Plant Tissues Under Environmental Stress.* University of Oxford.

XIONG, W., LIU, L., WU, C., YANG, C. & WU, Q. 2010. ^{13}C-tracer and gas chromatography-mass spectrometry analyses reveal metabolic flux distribution in the oleaginous microalga *Chlorella protothecoides*. *Plant Physiology,* 154**,** 1001.

XU, H., ANDI, B., QIAN, J., WEST, A. H. & COOK, P. F. 2006. The α-aminoadipate pathway for lysine biosynthesis in fungi. *Cell Biochemistry and Biophysics,* 46**,** 43-64.

YAMADA, T., LETUNIC, I., OKUDA, S., KANEHISA, M. & BORK, P. 2011. iPath2.0: interactive pathway explorer. *Nucleic Acids Research,* 39**,** W412-W415.

YAMANE, Y.-I., UTSUNOMIYA, T., WATANABE, M. & SASAKI, K. 2001. Biomass production in mixotrophic culture of *Euglena gracilis* under acidic condition and its growth energetics. *Biotechnology Letters,* 23**,** 1223-1228.

YANG, S. T., ZHANG, K., ZHANG, B. & HUANG, H. 2011. Fumaric acid. *In:* MOO-YOUNG, M. (ed.) *Comprehensive Biotechnology (Second Edition).* Burlington: Academic Press.

YOKOTA, A., HOSOTANI, K. & KITAOKA, S. 1982. Mechanism of metabolic regulation in photoassimilation of propionate in *Euglena gracilis* Z. *Archives of Biochemistry and Biophysics,* 213**,** 530-537.

YOON, H. S., HACKETT, J. D. & BHATTACHARYA, D. 2002. A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proceedings of the National Academy of Sciences,* 99**,** 11724.

YOSHIDA, Y., TOMIYAMA, T., MARUTA, T., TOMITA, M., ISHIKAWA, T. & ARAKAWA, K. 2016. De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics,* 17**,** 182.

YOUNG, J. D. 2014. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics,* 30**,** 1333-1335.

YOVAL-SÁNCHEZ, B., JASSO-CHÁVEZ, R., LIRA-SILVA, E., MORENO-SÁNCHEZ, R. & RODRÍGUEZ-ZAVALA, J. S. 2011. Novel mitochondrial alcohol metabolizing enzymes of *Euglena gracilis*. *Journal of Bioenergetics and Biomembranes,* 43**,** 519.

ZÁHONOVÁ, K., FÜSSY, Z., BIRČÁK, E., NOVÁK VANCLOVÁ, A. M. G., KLIMEŠ, V., VESTEG, M., KRAJČOVIČ, J., OBORNÍK, M. & ELIÁŠ, M. 2018. Peculiar features of the plastids of the colourless alga *Euglena longa* and photosynthetic euglenophytes unveiled by transcriptome analyses. *Scientific Reports,* 8**,** 17012.

ZAMBONI, N., FENDT, S.-M., RUHL, M. & SAUER, U. 2009. ^{13}C-based metabolic flux analysis. *Nature Protocols,* 4**,** 878-892.

ZAMBONI, N., FISCHER, E. & SAUER, U. 2005. FiatFlux – a software for metabolic flux analysis from ^{13}C-glucose experiments. *BMC Bioinformatics,* 6**,** 209.

ZHAO, J. & SHIMIZU, K. 2003. Metabolic flux analysis of *Escherichia coli* K12 grown on ^{13}C-labeled acetate and glucose using GC-MS and powerful flux calculation method. *Journal of Biotechnology,* 101**,** 101-117.

Appendices

Appendix A. Amino acid fragments used in ^{13}C MFA

Amino acid	Fragmet	m/z	Carbon atom
Alanine-2TBDMS	ALA M-57	260	1,2,3
Alanine-2TBDMS	ALA M-85	232	2,3
Aspartate-3TBDMS	ASP M-57	417	1,2,3,4
Aspartate-3TBDMS	ASP M-85	389	2,3,4
Glutamate-3TBDMS	GLU M-15	474	1,2,3,4,5
Glutamate-3TBDMS	GLU M-57	432	1,2,3,4,5
Glutamate-3TBDMS	GLU M-85	404	2,3,4,5
Glutamate-3TBDMS	GLU M-159	330	2,3,4,5
Glycine-2TBDMS	GLY M-57	246	1,2
Glycine-2TBDMS	GLY M-85	218	2
Leucine-2TBDMS	LEU M-57	302	1,2,3,4,5,6
Leucine-2TBDMS	LEU M-85	274	2,3,4,5,6
Phenylalanine-2TBDMS	PHE M-57	336	1,2,3,4,5,6,7,8,9
Phenylalanine-2TBDMS	PHE M-85	308	2,3,4,5,6,7,8,9
Phenylalanine-2TBDMS	PHE M-159	234	2,3,4,5,6,7,8,9
Phenylalanine-2TBDMS	PHE f302	302	1,2
Serine-3TBDMS	SER M-57	390	1,2,3
Serine-3TBDMS	SER M-85	362	2,3
Serine-3TBDMS	SER M-159	288	2,3
Tyrosine-3TBDMS	TYR f302	302	1,2
Valine-2TBDMS	VAL M-57	288	1,2,3,4,5
Valine-2TBDMS	VAL M-85	260	2,3,4,5

Appendix B. Biomass synthesis reactions of *E. gracilis* grown heterotrophically on glucose, acetate, ethanol or pyruvate.

Glucose

Protein	0.157*ALA + 0.089*ARG + 0.062*ASN + 0.062*ASP + 0.006*CYS + 0.127*GLY + 0.092*GLU + 0.092*GLN + 0.032*HIS + 0.067*ILE + 0.151*LEU + 0.101*LYS + 0.017*MET + 0.055*PHE + 0.106*PRO + 0.088*SER + 0.087*THR + 0.007*TRP + 0.043*TYR + 0.108*VAL -> biomass_P
Carbohydrate	3.047*G6P.b + 3.047*ATPb + 1.387*F6P.b + 1.387*NADHb -> biomass_C
Lipid	0.98*AcCoA.b + 0.02324*Gly3P.b + 0.98*ATPb + 1.003*NADHb + 0.98*NADPHb + 0.008*PPNCoA + 0.004*G6P.b + 0.012*3PGA.b -> biomass_L
Nucleic acid	0.0571*IMP + 0.051*UMP + 0.004*TMP -> biomass_NA

Acetate

Protein	0.048*ALA + 0.027*ARG + 0.019*ASN + 0.019*ASP + 0.002*CYS + 0.038*GLY + 0.028*GLU + 0.028*GLN + 0.01*HIS + 0.02*ILE + 0.045*LEU + 0.031*LYS + 0.005*MET + 0.017*PHE + 0.032*PRO + 0.026*SER + 0.026*THR + 0.002*TRP + 0.013*TYR + 0.033*VAL -> biomass_P
Carbohydrate	4.524*G6P.b + 4.524*ATPb + 0.067*F6P.b + 0.067*NADHb -> biomass_C
Lipid	1.535*AcCoA.b + 1.535*ATPb + 1.535*NADPHb + 0.0364*Gly3P.b + 1.571*NADHb + 0.011*PPNCoA + 0.006*G6P.b + 0.0195*3PGA.b -> biomass_L
Nucleic acid	0.0571*IMP + 0.051*UMP + 0.004*TMP -> biomass_NA

Ethanol

Protein	0.085*ALA + 0.048*ARG + 0.033*ASN + 0.033*ASP + 0.003*CYS + 0.068*GLY + 0.049*GLU + 0.049*GLN + 0.017*HIS + 0.036*ILE + 0.081*LEU + 0.054*LYS + 0.009*MET + 0.03*PHE + 0.057*PRO + 0.047*SER + 0.047*THR + 0.004*TRP + 0.023*TYR + 0.058*VAL -> biomass_P
Carbohydrate	4.698*G6P.b + 4.698*ATPb + 0.119*F6P.b + 0.119*NADHb -> biomass_C
Lipid	1.498*AcCoA.b + 1.498*ATPb + 1.498*NADPHb + 0.03639*Gly3P.b + 1.534*NADHb + 0.011*PPNCoA + 0.0061*G6P.b + 0.0195*3PGA.b -> biomass_L
NA	0.0571*IMP + 0.051*UMP + 0.004*TMP -> biomass_NA

Pyruvate

Protein	0.1*ALA + 0.056*ARG + 0.039*ASN + 0.039*ASP + 0.004*CYS + 0.08*GLY + 0.058*GLU + 0.058*GLN + 0.021*HIS + 0.042*ILE + 0.095*LEU + 0.064*LYS + 0.011*MET + 0.035*PHE + 0.067*PRO + 0.055*SER + 0.055*THR + 0.005*TRP + 0.027*TYR + 0.069*VAL -> biomass_P
Carb	4.546*G6P.b + 4.546*ATPb + 0.063*F6P.b + 0.063*NADHb -> biomass_C
Lipid	1.447*AcCoA.b + 1.447*ATPb + 1.447*NADPHb + 0.0354*Gly3P.b + 1.482*NADHb + 0.011*PPNCoA + 0.006*G6P.b + 0.0188*3PGA.b -> biomass_L
NA	0.0571*IMP + 0.051*UMP + 0.004*TMP -> biomass_NA

Appendix C. ^{13}C MFA fluxes of central metabolic pathways of wild-type and aplastidic *E. gracilis* grown on glucose

ID	Pathways	Reactions	Wild-type			Aplastidic		
			Flux	LB	UB	Flux	LB	UB
CO2 out	CO2 release	CO2 -> CO2out	1.0268	0.6513	1.3069	2.1557	1.7905	2.4184
Gly0	Glucose input	GLC + ATPc -> G6P	1	1	1	1	1	1
Gly1 net	Cytosolic glycolysis	G6P <-> F6P	0.5214	0.4537	0.5845	0.5213	0.4813	0.5532
Gly2 net	Cytosolic glycolysis	F6P + ATPc <-> F1_6P	0.3656	0.3178	0.3922	0.5067	0.4629	0.5384
Gly3 net	Cytosolic glycolysis	F1_6P <-> DHAP + G3P	0.3656	0.3178	0.3922	0.5067	0.4629	0.5384
Gly4 net	Cytosolic glycolysis	DHAP <-> G3P	0.3625	0.3155	0.3945	0.5031	0.4589	0.5351
Gly5 net	Cytosolic glycolysis	G3P <-> 1_3BPGA	0.7399	0.6511	0.8009	1.0344	0.9459	1.0981
Gly6 net	Cytosolic glycolysis	1_3BPGA <-> 3PGA + ATPc + NADHc	0.7399	0.6511	0.8009	1.0344	0.9459	1.0981
Gly8 net	Cytosolic glycolysis	3PGA <-> 2PGA	0.667	0.5118	0.7223	0.9743	0.8765	1.0492
Gly9 net	Cytosolic glycolysis	2PGA <-> PEP	0.667	0.5118	0.7223	0.9743	0.8765	1.0492
Gly10	Cytosolic glycolysis	PEP -> PYRc + ATPc	0.4348	0.3531	0.6875	0.7541	0.6493	0.9944
GlyP1 net	Plastidic glycolysis	G6Pp <-> F6Pp	0.0045	NaN	0.01	-	-	-
GlyP2 net	Plastidic glycolysis	F6Pp + ATPc <-> F1_6Pp	0.0023	NaN	0.0091	-	-	-
GlyP3 net	Plastidic glycolysis	F1_6Pp <-> DHAPp + G3Pp	0.0023	NaN	0.0091	-	-	-
GlyP4 net	Plastidic glycolysis	DHAPp <-> G3Pp	0.0023	-0.0075	0.0087	-	-	-
GlyP5 net	Plastidic glycolysis	G3Pp <-> 1_3BPGAp	1.73E-04	1.71E-04	1.89E-04	-	-	-
GlyP6 net	Plastidic glycolysis	1_3BPGAp <-> 3PGAp + ATPc + NADHc	1.73E-04	1.71E-04	1.89E-04	-	-	-
GlyP8 net	Plastidic glycolysis	3PGAp <-> 2PGAp	1.91E-04	NaN	1.95E-04	-	-	-
GlyP9 net	Plastidic glycolysis	2PGAp <-> PEPp	1.91E-04	NaN	1.95E-04	-	-	-
GlyP10	Plastidic glycolysis	PEPp -> PYRp + ATPc	0.0083	0	0.0681	-	-	-
PPP1	oxPPP	G6P -> 6PGL + NADPHc	0.0675	0.0152	0.1337	0.1088	0.088	0.1266
PPP2	oxPPP	6PGL -> 6PG	0.0675	0.0152	0.1337	0.1088	0.088	0.1266
PPP3	oxPPP	6PG -> Ru5P + CO2 + NADPHc	0.0675	0.0152	0.1337	0.1088	0.088	0.1266
PPP4 net	Cytosolic nonoxPPP	Ru5P <-> X5P	0.0291	-0.0099	0.0718	0.0584	0.0445	0.0708
PPP5 net	Cytosolic nonoxPPP	Ru5P <-> R5P	0.0384	0.0168	0.0624	0.0504	0.0434	0.0573
PPP6 net	Cytosolic nonoxPPP	R5P + TKC2 <-> S7P	0.0183	-4.05E-04	0.0438	0.0347	0.0277	0.0407
PPP7 net	Cytosolic nonoxPPP	E4P + TKC2 <-> F6P	0.0108	-0.0119	0.03	0.0237	0.0167	0.0303
PPP8 net	Cytosolic nonoxPPP	X5P <-> G3P + TKC2	0.0291	-0.0097	0.0725	0.0584	0.0445	0.0708
PPP9 net	Cytosolic nonoxPPP	S7P <-> E4P + TAC3	0.0183	-4.14E-04	0.0424	0.0347	0.0277	0.0407

PPP10 net	Cytosolic nonoxPPP	G3P + TAC3 <-> F6P	0.0183	-4.14E-04	0.0424	0.0347	0.0277	0.0407
PPPp4 net	Plastidic nonoxPPP	Ru5Pp <-> X5Pp	-0.0022	-0.0058	-6.22E-06	-	-	-
PPPp5 net	Plastidic nonoxPPP	Ru5Pp <-> R5Pp	0.0022	6.22E-06	0.0058	-	-	-
PPPp6 net	Plastidic nonoxPPP	R5Pp + TKC2p <-> S7Pp	0.0022	NaN	0.0028	-	-	-
PPPp7 net	Plastidic nonoxPPP	E4Pp + TKC2p <-> F6Pp	-0.0044	-0.0072	-2.44E-04	-	-	-
PPPp8 net	Plastidic nonoxPPP	X5Pp <-> G3Pp + TKC2p	-0.0022	-0.0063	-5.63E-05	-	-	-
PPPp9 net	Plastidic nonoxPPP	S7Pp <-> E4Pp + TAC3p	0.0022	NaN	0.0028	-	-	-
PPPp10 net	Plastidic nonoxPPP	G3Pp + TAC3p <-> F6Pp	0.0022	NaN	0.0028	-	-	-
AcCoA	Pyruvate dehydrogenase	PYRm -> CO2 + AcCoA + NADPHc	0.4724	0.3472	0.5397	0.8281	0.7359	0.9054
TCA1	TCA cycle	AcCoA + OAA -> CIT	0.3021	0.1726	0.386	0.6543	0.5479	0.7434
TCA2 net	TCA cycle	CIT <-> IsC	0.3021	0.1726	0.386	0.6543	0.5479	0.7434
TCA3	TCA cycle	IsC -> CO2 + AKG + NADHc	0.2885	0.1302	0.3984	0.6543	0.5042	0.7434
TCA4	TCA cycle	AKG -> CO2 + SUCC + NADHc	0.2246	0.0759	0.339	0.6047	0.4537	0.6977
TCA5 net	TCA cycle	SUCC <-> SUC + ATPc	0.2212	0.0576	0.3353	0.6017	0.4439	0.695
TCA6 net	TCA cycle	SUC <-> FUM + FADH2c	0.237	0.1173	0.3338	0.6036	0.5153	0.6963
TCA7 net	TCA cycle	FUM <-> MAL	0.2532	0.135	0.3436	0.6162	0.5292	0.7084
TCA8 net	TCA cycle	MAL <-> OAA + NADHc	0.1733	0.1104	0.427	0.505	0.3646	0.7487
PEPcarl	PEP carboxylase	PEP + CO2 + ATPc <-> OAA	0.1961	0	0.2263	0.1983	0	0.2671
MAL1 net	Malate dehydrogenase	MAL <-> PYRc + CO2 + NADPHc	0.0934	-0.1396	0.1296	0.1112	-0.0922	0.2297
Glyox1 net	Glyoxylate shunt	IsC <-> Glyox + SUC	0.0135	-1.94E-16	0.0543	1.00E-07	-5.55E-17	0.0344
Glyox2	Glyoxylate shunt	Glyox + AcCoA -> MAL	0.0135	0	0.0543	1.00E-07	0	0.0344
PC	Pyruvate carboxylase	PYRc + CO2 + ATPc -> OAA	4.55E-06	0	0.2717	1.19E-05	0	0.3102
Biomass	Biomass sybthesis	biomass_P + biomass_C + biomass_L + biomass_NA -> Biomass	0.1333	0.1263	0.1439	0.1023	0.0953	0.1119

Appendix D. The stichometry of cofactors provision in ^{13}C MFA model

ID	Reaction	Cofactors provision			
		ATP	NADH	$FADH_2$	NADPH
AC1	AC + ATPc -> AcCoA	-1			
ADH	ACE -> AC + NADHc		1		
ALDH	EtOH -> ACE + NADHc		1		
Gly0	GLC + ATPc -> G6P	-1			
Gly2	F6P + ATPc -> F1_6P	-1			
Gly6 net	1_3BPGA <-> 3PGA + ATPc + NADHc	1	1		
Gly10	PEP -> PYRc + ATPc	1			
GlyP2 net	F6Pp + ATPc <-> F1_6Pp	-1			
GlyP6 net	1_3BPGAp <-> 3PGAp + ATPc + NADHc	1	1		
GlyP10	PEPp -> PYRp + ATPc	1			
PPP1	G6P -> 6PGL + NADPHc				1
PPP3	6PG -> Ru5P + CO2 + NADPHc				1
AcCoA	PYRm -> CO2 + AcCoA + NADPHc				1
TCA3	IsC -> CO2 + AKG + NADHc		1		
TCA4	AKG -> CO2 + SUCC + NADHc		1		
TCA5 net	SUCC <-> SUC + ATPc	1			
TCA6 net	SUC <-> FUM + FADH2c			1	
TCA8 net	MAL <-> OAA + NADHc		1		
PEPcarl	PEP + CO2 + ATPc -> OAA	-1			
MAL1 net	MAL <-> PYRc + CO2 + NADPHc				1
AC1	AC + ATPc -> AcCoA	-1			
AA1 net	THR <-> GLY + ACE + NADHb		1		
AA4 net	GLY <-> CO2 + MEETHF + NADHb + NH3		1		
AA9	AKG.b + NADHb + NH3 + ATPb -> GLU	-1	-1		
AA11	GLU + ATPb + NH3 -> GLN	-1			
AA12	GLU + ATPb + 2*NADPHb -> PRO	-1			-2
AA13	ASP + GLN + ATPb -> ASN + GLU	-1			
AA14	GLU + CO2 + GLN + ASP + AcCoA + 5*ATPb + NADPHb -> ARG + AKG + FUM + AC	-5			-1
AA15	R5Pp.b + FTHF + GLN + ASP + 5*ATPb -> HIS + AKG.p + FUM + 2*NADHb	-5	2		
AA16	R5P.b + FTHF + GLN + ASP + 5*ATPb -> HIS + AKG.p + FUM + 2*NADHb	-5	2		
AA17	3PGAp.b + GLU -> SER + AKG.p + NADHb		1		
AA18	3PGA.b + GLU -> SER + AKG.p + NADHb		1		
AA19	SER + AcCoA + 3*ATPb + 4*NADPHb + SO4 -> CYS + AC	-3			-4
AA20	PYRc.b + GLU -> ALA + AKG.p				
AA21	E4Pp.b + PEPp.b + PEPp.b + GLU + ATPb + NADPHb -> PHE + CO2 + AKG.p	-1			-1
AA22	E4Pp.b + PEPp.b + PEPp.b + GLU + ATPb + NADPHb -> TYR + CO2 + AKG.p	-1			-1
AA23	E4P.b + PEP.b + PEP.b + GLU + ATPb + NADPHb -> PHE + CO2 + AKG.p	-1			-1

AA24	E4P.b + PEP.b + PEP.b + GLU + ATPb + NADPHb -> TYR + CO2 + AKG.p	-1			-1
AA25	PYRp.b + PYRp.b + GLU + NADPHb -> VAL + CO2 + AKG.p				-1
AA26	PYRp.b + PYRp.b + AcCoAp.b + GLU + NADPHb -> LEU + CO2 + CO2 + AKG.p + NADHb		1		-1
AA27	PYRc.b + PYRc.b + GLU + NADPHb -> VAL + CO2 + AKG.p				-1
AA28	PYRc.b + PYRc.b + AcCoA.b + GLU + NADPHb -> LEU + CO2 + CO2 + AKG.p + NADHb		1		-1
AA29	ASP + METHF + CYS + SUCC + ATPb + 2*NADPHb -> MET + PYRc.p + SUC + NH3	-1			-2
AA30	ASP + 2*ATPb + 2*NADPHb -> THR	-2			-2
AA31	PYRp.b + THR + GLU + NADPHb -> ILE + CO2 + AKG.p + NH3				-1
AA32	PYRc.b + THR + GLU + NADPHb -> ILE + CO2 + AKG.p + NH3				-1
AA33	ASP + PYRp.b + GLU + SUCC + ATPb + 2*NADPHb -> LL_DAP + AKG.p + SUC	-1			-2
AA34	ASP + PYRc.b + GLU + SUCC + ATPb + 2*NADPHb -> LL_DAP + AKG.p + SUC	-1			-2
AA35	LL_DAP -> LYS + CO2				
AA36	SER + R5P.b + PEP.b + E4P.b + PEP + GLN + 3*ATPb + NADPHb -> TRP + CO2 + G3P + PYRc + GLU	-3			-1
AA37	SER + R5Pp.b + PEPp.b + E4Pp.b + PEPp + GLN + 3*ATPb + NADPHb -> TRP + CO2 + G3Pp + PYRp + GLU	-3			-1
IMP	R5P.b + GLY + FTHF + CO2 + FTHF + 5*ATPb -> IMP	-5			
UMP	ASP + CO2 + R5P.b + ATPb -> UMP + CO2 + NADHb	-1	1		
TMP	ASP + CO2 + R5P.b + MEETHF + ATPb -> TMP + CO2 + NADHb	-1	1		
Carb*	3.047*G6P.b + 3.047*ATPb + 1.387*F6P.b + 1.387*NADHb -> biomass_C	-3.047	-1.387		
Lipid*	0.98*AcCoA.b + 0.02324*Gly3P.b + 0.98*ATPb + 1.003*NADHb + 0.98*NADPHb + 0.008*PPNCoA + 0.004*G6P.b + 0.012*3PGA.b -> biomass_L	-0.98	-1.003		-0.98
CX2	CX + ATPb -> FTHF	-1			
CX3	MEETHF + NADHb -> METHF		-1		
CX4	MEETHF -> FTHF + NADPHb				1
PYR car	PYRc + CO2 + ATPc -> OAA	-1			
AAA LYS	AKG.b + AcCoA + NADHb + 2*GLU + ATPb + NADPHb -> LYS + CO2 + 2*AKG.p	-1	-1		-1

* For carbohydrate (carb) and lipid biomass sythesis reactions, the cofactors provision depends on the mol contribution of precursors in the reactions

Appendix E. ^{13}C MFA fluxes of central metabolic pathways of *E. gracilis* grown on acetate, ethanol and pyruvate

ID	Pathways	Reactions	Acetate			Ethanol			Pyruvate		
			Flux	LB	UB	Flux	LB	UB	Flux	LB	UB
CO2 out	CO2 release	CO2 -> CO2out	1.2185	1.2041	1.2565	0.7309	0.6994	0.7711	2.0161	1.9655	2.0587
Input	Substrate input	substrateout -> substrate	1	1	1	1	1	1	1	1	1
Gly1 net	Cytosolic glycolysis	G6P <-> F6P	-0.1035	-0.1474	0.0478	-0.1606	-0.2091	-0.1556	-0.1301	-0.1375	-0.1255
Gly2 net	Cytosolic glycolysis	F6P + ATPc <-> F1_6P	-0.1135	-0.1532	-0.107	-0.1685	-0.1745	-0.1661	-0.1306	NaN	-0.125
Gly3 net	Cytosolic glycolysis	F1_6P <-> DHAP + G3P	-0.1135	-0.1532	-0.107	-0.1685	-0.1745	-0.1661	-0.1306	NaN	-0.125
Gly4 net	Cytosolic glycolysis	DHAP <-> G3P	-0.0998	NaN	NaN	-0.1698	-0.1764	-0.1673	-0.1305	-0.2317	NaN
Gly5 net	Cytosolic glycolysis	G3P <-> 1_3BPGA	-0.2179	NaN	NaN	-0.3401	-0.3502	-0.3353	-0.2603	NaN	-0.2498
Gly6 net	Cytosolic glycolysis	1_3BPGA <-> 3PGA + ATPc + NADHc	-0.2179	NaN	NaN	-0.3401	-0.3502	-0.3353	-0.2603	NaN	-0.2498
Gly8 net	Cytosolic glycolysis	3PGA <-> 2PGA	-0.2212	-0.3262	NaN	-0.3765	-0.5132	-0.3584	-0.2733	NaN	-0.2589
Gly9 net	Cytosolic glycolysis	2PGA <-> PEP	-0.2212	-0.3262	NaN	-0.3765	-0.5132	-0.3584	-0.2733	NaN	-0.2589
Gly10	Cytosolic glycolysis	PEP -> PYRc + ATPc	0.0041	0	0.3866	0.3467	0.0348	0.5196	1.00E-07	0	0.0336
GlyP1 net	Plastidic glycolysis	G6Pp <-> F6Pp	1.10E-05	-0.1467	0.0266	3.88E-13	-3.32E-05	2.21E-04	6.07E-04	1.14E-04	NaN
GlyP2 net	Plastidic glycolysis	F6Pp + ATPc <-> F1_6Pp	0.0063	1.40E-04	0.0444	1.30E-04	-2.07E-05	2.41E-04	0.0011	5.97E-04	0.0248
GlyP3 net	Plastidic glycolysis	F1_6Pp <-> DHAPp + G3Pp	0.0063	1.40E-04	0.0444	1.30E-04	-2.07E-05	2.41E-04	0.0011	5.97E-04	0.0248
GlyP4 net	Plastidic glycolysis	DHAPp <-> G3Pp	-0.0082	NaN	NaN	2.05E-04	-3.76E-05	8.88E-04	3.43E-05	NaN	NaN
GlyP5 net	Plastidic glycolysis	G3Pp <-> 1_3BPGAp	0.0013	NaN	NaN	-4.63E-04	-8.87E-04	NaN	4.45E-04	-0.0428	NaN
GlyP6 net	Plastidic glycolysis	1_3BPGAp <-> 3PGAp + ATPc + NADHc	0.0013	NaN	NaN	-4.63E-04	-8.87E-04	NaN	4.45E-04	-0.0428	NaN
GlyP8 net	Plastidic glycolysis	3PGAp <-> 2PGAp	4.75E-04	NaN	NaN	-4.64E-04	-8.89E-04	4.96E-05	-0.0024	-0.0285	NaN
GlyP9 net	Plastidic glycolysis	2PGAp <-> PEPp	4.75E-04	NaN	NaN	-4.64E-04	-8.89E-04	4.96E-05	-0.0024	-0.0285	NaN
GlyP10	Plastidic glycolysis	PEPp -> PYRp + ATPc	1.00E-07	0	0.0023	1.00E-07	0	0.0019	1.00E-07	0	4.64E-04
PPP1	oxPPP	G6P -> 6PGL + NADPHc	2.71E-05	-2.64E-16	0.0431	1.00E-07	-1.33E-15	0.0223	0.0072	0.0049	0.0074
PPP2	oxPPP	6PGL -> 6PG	2.71E-05	0	0.0431	1.00E-07	-1.11E-16	0.0223	0.0072	0.0049	0.0074
PPP3	oxPPP	6PG -> Ru5P + CO2 + NADPHc	2.71E-05	0	0.0431	1.00E-07	0	0.0223	0.0072	0.0049	0.0074
PPP4 net	Cytosolic nonoxPPP	Ru5P <-> X5P	0.0012	-0.0706	0.0409	-0.0048	-0.005	0.0099	0.002	0.0018	0.0022
PPP5 net	Cytosolic nonoxPPP	Ru5P <-> R5P	-0.0011	-0.0068	0.0145	0.0048	0.0031	0.0119	0.0052	0.0049	NaN

PPP6 net	Cytosolic nonoxPPP	R5P + TKC2 <-> S7P	-0.0038	-0.0092	0.0103	-9.05E-04	-0.001	0.0015	6.80E-04	1.29E-04	7.93E-04
PPP7 net	Cytosolic nonoxPPP	E4P + TKC2 <-> F6P	-0.0046	-0.0109	0.0054	-0.0029	-0.003	0.0039	5.40E-04	NaN	5.83E-04
PPP8 net	Cytosolic nonoxPPP	X5P <-> G3P + TKC2	-0.0084	-0.0216	0.0231	-0.0038	-0.004	-6.48E-04	0.0012	NaN	0.0013
PPP9 net	Cytosolic nonoxPPP	S7P <-> E4P + TAC3	-0.0038	-0.0092	0.0103	-9.05E-04	-0.001	0.0015	6.80E-04	1.29E-04	7.93E-04
PPP10 net	Cytosolic nonoxPPP	G3P + TAC3 <-> F6P	-0.0038	-0.0092	0.0103	-9.05E-04	-0.001	0.0015	6.80E-04	1.29E-04	7.93E-04
PPPp4 net	Plastidic nonoxPPP	Ru5Pp <-> X5Pp	-0.0033	-0.0137	-4.91E-05	0.0011	-5.57E-04	0.0012	-3.01E-04	NaN	-2.23E-04
PPPp5 net	Plastidic nonoxPPP	Ru5Pp <-> R5Pp	0.0033	4.91E-05	0.0137	-0.0011	-0.0012	5.57E-04	3.01E-04	2.23E-04	NaN
PPPp6 net	Plastidic nonoxPPP	R5Pp + TKC2p <-> S7Pp	0.0031	5.07E-05	0.0136	3.87E-05	-3.48E-05	1.41E-04	0.0011	0.001	NaN
PPPp7 net	Plastidic nonoxPPP	E4Pp + TKC2p <-> F6Pp	0.0032	-1.36E-04	0.0178	9.15E-05	-1.56E-06	1.93E-04	-5.70E-04	-6.48E-04	-2.54E-04
PPPp8 net	Plastidic nonoxPPP	X5Pp <-> G3Pp + TKC2p	0.0063	1.69E-04	0.0273	1.30E-04	-1.67E-05	3.36E-04	5.20E-04	3.82E-04	6.95E-04
PPPp9 net	Plastidic nonoxPPP	S7Pp <-> E4Pp + TAC3p	0.0031	5.07E-05	0.0136	3.87E-05	-3.48E-05	1.41E-04	0.0011	0.001	NaN
PPPp10 net	Plastidic nonoxPPP	G3Pp + TAC3p <-> F6Pp	0.0031	5.07E-05	0.0136	3.87E-05	-3.48E-05	1.41E-04	0.0011	0.001	NaN
AcCoA	Pyruvate dehydrogenase	PYRm -> CO2 + AcCoA + NADPHc	1.63E-04	0	0.2585	1.00E-07	0	0.0455	0.8338	0.806	0.8419
TCA1	TCA cycle	AcCoA + OAA -> CIT	0.7291	0.7239	0.7366	0.5859	0.5707	0.6193	0.6748	0.6567	0.6898
TCA2 net	TCA cycle	CIT <-> IsC	0.7291	0.7239	0.7366	0.5859	0.5707	0.6193	0.6748	0.6567	0.6898
TCA3	TCA cycle	IsC -> CO2 + AKG + NADHc	0.4951	0.2352	0.5037	0.1814	0.1559	0.2099	0.527	0.4955	0.5565
TCA4	TCA cycle	AKG -> CO2 + SUCC + NADHc	0.4924	0.2328	0.5012	0.1747	0.1446	0.2038	0.5177	0.4457	0.5406
TCA5 net	TCA cycle	SUCC <-> SUC + ATPc	0.4914	0.2317	0.5002	0.1719	0.1478	0.2049	0.5183	0.4864	NaN
TCA6 net	TCA cycle	SUC <-> FUM + FADH2c	0.7262	0.721	0.7332	0.5788	0.5624	0.61	0.6652	0.6491	0.6817
TCA7 net	TCA cycle	FUM <-> MAL	0.7271	0.7218	0.7346	0.581	0.5646	0.6114	0.6672	0.6495	0.6833
TCA8 net	TCA cycle	MAL <-> OAA + NADHc	0.9595	0.9569	1.3433	1.3688	NaN	1.5519	0.7034	0.6146	0.751
PEPcarl	PEP carboxylase	PEP + CO2 + ATPc <-> OAA	-0.2263	-0.609	-0.2197	0.7276	0.5664	NaN	-0.2793	NaN	NaN
MAL1 net	Malate dehydrogenase	MAL <-> PYRc + CO2 + NADPHc	0.0016	-0.1324	0.1873	-0.3833	-0.5483	0.6493	0.1115	0.0837	0.1229
Glyox1 net	Glyoxylate shunt	IsC <-> Glyox + SUC	0.234	0.2301	0.4925	0.4045	0.3951	0.4476	0.1477	0.0383	0.1674
Glyox2	Glyoxylate shunt	Glyox + AcCoA -> MAL	0.234	0.2301	0.4925	0.4045	0.3951	0.4476	0.1477	0.0383	0.1674
PC	Pyruvate caboxlyase	PYRc + CO2 + ATPc -> OAA	2.11E-04	0	0.1449	1.00E-07	0	0.8292	0.2878	0.2686	0.3051
ADH	Alcohol dehydrogenase	ACE -> AC + NADHc	-	-	-	1.0438	1.0285	1.0819	-	-	-

ALDH	Aldehyde dehydrogenase	EtOH -> ACE + NADHc	-	-	-	1	1	1	-	-	-
Biomass	Total	biomass_P + biomass_C + biomass_L + biomass_NA -> Biomass	0.0228	0.0218	0.0233	0.0341	0.0334	0.035	0.0269	0.0257	0.0284