Application of label-free shotgun proteomics for the analysis of metabolic pathways in *Euglena gracilis* and its potential as a source of food supplements

Mafruha Tasnin Hasan

A thesis submitted for the degree of

Doctor of Philosophy

Supervisors:

Prof. Helena Nevalainen

A/Prof. Anwar Sunna

Dr. Angela Sun

MACQUARIE University
SYDNEY·AUSTRALIA

Department of Molecular Sciences

Macquarie University, Sydney, Australia

January 2018

Statement of declaration

I certify that the work presented in this thesis has not been submitted as a part of the requirement for a degree or course to any institute or university other than Macquarie University.

I also certify that the thesis is an original piece of research conducted by me and it contains no materials previously published or written by any other person except where due reference is made in the text. I consent that a copy of this thesis is made available at Macquarie University Library for loan and photocopying forthwith.

Mafruha Tasnin Hasan

January 2018

Abstract

Proteomic studies provide insight into the global expression patterns of proteins within cells. Proteomic profiling of the freshwater protist *Euglena gracilis* has not been previously investigated, although its transcriptome has recently been published. This work aimed at profiling and identifying proteins involved in carbon metabolism, the biosynthetic and degradation pathways of paramylon, and the biosynthetic pathways of α-tocopherol, ascorbate and the twenty protein-building amino acids under photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivations.

Paramylon content was the highest under HT condition, with the streptomycin-bleached mutant *E. gracilis* ZSB synthesising the most paramylon and the wild-type *E. gracilis* var. *saccharophila* synthesising paramylon the fastest. However, antioxidant content was the highest under PT condition, with the wild-type *E. gracilis* Z strain accumulating the most antioxidants. The abundance of some free amino acids varied between the mid-exponential phase and the beginning of the stationary phase, and between growth conditions, but the total amount remained about the same with arginine as the most abundant amino acid.

Label-free shotgun proteomics enabled identification of over 4000 translated proteins, about 30% of which could not be annotated by sequence similarity alone. Many enzymes exhibited several isoforms that were influenced by growth condition. Not all isozymes identified in the transcriptome were detected in the proteome, suggesting post-transcriptional regulation. The results indicated diversity of pathways similar to different organisms, such as lysine biosynthesis to fungi, and serine and proline biosyntheses to plants. Some pathways were unique to *Euglena*, such as the TCA cycle, and paramylon, ascorbate and arginine biosyntheses.

The proteomic studies revealed that instead of hexokinase *E. gracilis* uses a high-specificity glucokinase for the EMP pathway. Two paramylon synthase candidates (EgGSL1 and EgGSL2) were identified of which EgGSL2 was predominant and expressed under all

growth conditions. EgGSL1 was not expressed under HT condition, but the gene transcript was detected by qRT-PCR across all growth conditions indicating light induction and post-transcriptional regulation. Some enzymes of the Calvin pathway were expressed under HT cultivation suggesting post-translational regulation. HT cells may also carry out CO_2 fixation in the dark even in the presence of sufficient glucose in the medium. Two pathways for serine biosynthesis were identified, one of which was prevalent under PT and MT cultivation, and the other under HT cultivation.

The MT cultivation of *E. gracilis* var. *saccharophila* was further chosen for evaluation of this strain as a potential source of food supplements in a laboratory-scale bioreactor, as this strain produced higher paramylon than *E. gracilis* Z, and under MT cultivation its antioxidant levels were higher than those of *E. gracilis* ZSB.

This work expands on the existing knowledge of metabolic pathways in *E. gracilis* and provides insight into how these pathways are influenced by growth condition, thus providing a foundation for future strain engineering.

Acknowledgements

I would like to express my utmost gratefulness to several individuals who have contributed to the completion of this project and made its writing worthwhile.

Foremost, I am indebted to my supervisor, Helena Nevalainen, for her relentless support, adept guidance and academic assistance throughout the course of this project. Without her sheer knowledge and expertise this project would not have been completed. Her confidence in me has encouraged me to grow as a researcher and provided a strong foundation for my future research endeavours. I would also like to take this opportunity to thank Angela Sun for training me and helping me to hone my laboratory and writing skills, and for her earnest encouragement and support as my co-supervisor. I am also grateful to my associate supervisor Anwar Sunna, and my adjunct supervisors Junior Te'o and Graham Hobba for taking the time to review all of my work and guide me through the initial planning of this project.

I want to express my gratitude to all the postdocs, research fellows and technical officers in the Department of Molecular Sciences, especially Nicole Vella, Dylan Xavier, Chihung Lin, Leon McQuade, Karthik Kamath, Mehdi Mirzaei, Dana Pascovici, Heinrich Kroukamp and Hugh Goold for helping and guiding me through microscopy, mass spectrometry, metabolite determination, proteomic analysis, biostastical analysis and qRT-PCR. I would also like to thank Liisa Kautto, who always took time out of her schedule to guide me through various lab techniques. My gratitude extends to my friends and laboratory mates Zhiping, Edward, Thi, Aneesh, Anna, Alex, Bishal, Patrick, Yunqi, Vineet, Daniel, Hasinika, Wei, Ian, Iniga, Wisam, Sameera and Shaz for the countless discussions that helped to broaden my scientific knowledge, and to keep lab and writing work fun and entertaining.

Lastly, I am thankful to my family and my beloved partner Tanzil, for their unconditional love, patience and support, especially in times of despondency.

Table of contents

List of publications and manuscripts

This thesis includes one published article (Chapter 3) and a manuscript presented as Chapter 4.

Published article from this thesis:

Hasan MT, Sun A, Mirzaei M, Te'o J, Hobba G, Sunna A and Nevalainen H. 2017. A comprehensive assessment of the biosynthetic pathways of ascorbate, α-tocopherol and free amino acids in *Euglena gracilis* var. *saccharophila*. *Algal Res*. 27: 140-151.

Manuscript:

Hasan MT, Sun A, Mirzaei M, Te'o J, Hobba G, Sunna A and Nevalainen H. Comparative proteomics investigation of central carbon metabolism in *Euglena gracilis* grown under photoautotrophic, mixotrophic and heterotrophic cultivations. Manuscript prepared for submission.

Additional publication:

Sun A, Hasan MT, Hobba G, Nevalainen H and Te'o J. 2018. Comparative assessment of the *Euglena gracilis* var. *saccharophila* variant strain as a producer of the β-1,3-glucan paramylon under varying light conditions. *J. Phycol*. In Press.

Conference and workshop attendance, and awards

List of conferences and workshops attended:

1. MT Hasan, A Sun, A Sunna, J Te'o, G Hobba and H Nevalainen. A comparative study of paramylon production in the wild type Z strain and var. *saccharophila* strain of *Euglena gracilis* under photoauto- and heterotrophic conditions. 5th International Conference on Algal Biomass, Biofuels and Bioproducts. 7-10th June 2015, San Diego, California, USA. **Poster**.

2. MT Hasan, A Sun, G Hobba, J Te'o, A Sunna and H Nevalainen. A proteomic study of paramylon metabolism in *Euglena gracilis* under photoauto-, mixo- and heterotrophic cultivations. Networks and Pathways Workshop. 2-6th November 2015, Wellcome Genome Campus, Cambridge, UK. **Poster**.

3. MT Hasan, A Sun, G Hobba, J Te'o, A Sunna and H Nevalainen. A comparative study of paramylon and other metabolites in *Euglena gracilis* Z and *Euglena gracilis* var. *saccharophila* under photoauto-, mixo- and heterotrophic cultivations. Advances in Biotechnology for Food and Medical Applications by ARC Food Processing Training Centre and the Institute of Biomedical Engineering and IT. 5-7th October 2016, University of Sydney, Sydney, Australia. **Poster**.

List of awards:

1. Macquarie University Postgraduate Fund 2015.

2. International Macquarie University Postgraduate Research Excellence (iMQRES) Scholarship 2014.

3. Australian Research Council's Industrial Transformation Training Centre (ARC ITTC) Scholarship 2014.

List of abbreviations

2-D	two-dimensional
2-DE	two-dimensional gel electrophoresis
acetyl-CoA	acetyl-coenzyme A
ACN	acetonitrile
ALDO	fructose-bisphosphate aldolase
ANOVA	analysis of variance
APAF	The Australian Proteome Analysis Facility
AQUA	absolute quantification methods
ATP	adenosine triphosphate
BCA	bicinchoninic acid
BHT	butylated hydroxytoluene
BLASTP	basic local alignment search tool for proteins
BSA	bovine serum albumin
C. reinhardtii	*Chlamydomonas reinhardtii*
CAM	Crassulacean acid metabolism
CAZys	carbohydrate-active enzymes
cDNA	complementary deoxyribonucleic acid
C_T	threshold cycle
DIGE	difference gel electrophoresis
DNA	deoxyribonucleic acid
DO	dissolved oxygen
DM	dry mass
DTT	dithiothreitol

E. gracilis	*Euglena gracilis*
E. gracilis ZSB	*Euglena gracilis* Z, streptomycin-bleached mutant
EgGSL	transcript of *Euglena* glucan synthase-like protein (paramylon synthase)
EgGSL	*Euglena* glucan synthase-like protein (paramylon synthase)
EC	enzyme class number
EDTA	ethylenediaminetetraacetic acid
EM	*Euglena* maintenance medium
EMP	Embden-Meyerhof-Parnas
ENO	enolase
FA	formic acid
FDR	false discovery rate
g	gravitational force
G1P	glucose-1-phosphate
G6P	glucose-6-phosphate
GABA	4-aminobutyrate
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GH	glycoside hydrolase
GK	glucokinase
GNY	glucose, ammonium chloride and yeast extract medium
GO	gene ontology
GPI	glucose-6-phosphate isomerase
GPM	global proteome machine
GRAS	generally regarded as safe
GRN	GRAS notice

GT	glycosyltransferase
h	hours
HIV	human immunodeficiency virus
HK	hexokinase
HT	heterotrophic
IA	indoleacetate
IAA	iodoacetamide
ICAT	isotope-coded affinity tag
ICP	inductively coupled plasma
IDBEST	isotope-differentiated binding energy shift tag
IMP	intramembranous particles
IPTL	isobaric peptide termini labelling
iTRAQ	isobaric tags for absolute and relative quantification
K/R-P	lysine/arginine-proline
LDS	lithium dodecyl sulphate
malyl-CoA	malyl-coenzyme A
min	minutes
MOPS	3-N-morpholinopropansulphonic acid
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MT	mixotrophic
N-starvation	nitrogen-starvation
NADPH	nicotinamide adenine dinucleotide phosphate
nanoLC-MS/MS	nanflow liquid chromatography tandem mass spectrometry

NCBI	National Center for Biotechnology Information
NCRIS	The National Collaborative Research Infrastructure Strategy
NSAF	normalised spectral abundance factor
NY	ammonium chloride and yeast extract medium
oxPP	oxidative pentose phosphate
PBS	phosphate buffered saline
PEP	phosphoenolpyruvate
PEPC	phosphoenolpyruvate carboxylase
PEPCK	phosphoenolpyruvate carboxykinase
PET	petroleum ether
PFK-1	6-phosphofructokinase
PGK	phosphoglycerate kinase
psi	pounds per square inch
PS II	photosystem II
PT	photoautotrophic
qPCR	quantitative polymerase chain reaction
qRT-PCR	real-time quantitative reverse transcription polymerase chain reaction
rC_p	redundant count of peptides
RNA	ribonucleic acid
ROS	reactive oxygen species
RP-HPLC	reverse phase high-performance liquid chromatography
RP-UPLC	reverse phase ultra-performance liquid chromatography
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
RT	room temperature

rTCA	reductive tricarboxylic acid
RuBisCO	ribulose-1,5-bisphosphate carboxylase/oxygenase
SAF	spectral abundance factor
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	seconds
SEM	scanning electron microscope
SILAC	stable isotope labelling by amino acids in cell culture
SSDH	succinate-semialdehyde dehydrogenase
succinyl-CoA	succinyl-coenzyme A
TCA	tricarboxylic acid
TMT	tandem mass tags
TPI	triosephosphate isomerase
UDP	uridine diphosphate
UTEX	University of Texas
UV	ultraviolet
v/v	volume per volume
vvm	vessel volumes per minute
w/v	weight per volume

CHAPTER 1

Introduction

1.1 Importance of food supplements

Owing to the rapid growth in human population, which is expected to reach almost 11 billion by the end of this century [1], higher yielding crops are grown at a faster rate to meet human and farm animal consumption. However, this has led to a decline in biodiversity and nutritional value of food because the high-yielding cultivars often spend most of their energy to grow grains and fruits [2]. There is a trade-off between yield and nutrient concentration, owing to a "genetic dilution effect" caused by selective breeding of high-yielding cultivars, which leads to cultivars having 80-90% carbohydrate content per dry weight, while lacking in vitamins, minerals and proteins [2]. Thus, a vast majority of the human population and farm animals are deficient in proteins and nutrients, and nutritional supplements are required to compensate for this lack. Consumption of food supplemented with antioxidants, vitamins, minerals, β-glucans (fibres), omega-3 fatty acids and other metabolites is also a prophylactic for the prevention of several chronic diseases including cancer, cardiovascular diseases, diabetes and obesity [3-6]. These supplements have to be produced in an environmentally friendly and commercially sustainable way, which has led to the introduction of a microbial platform for nutraceuticals and food supplements, including microalgae like *Euglena gracilis*.

1.2 Classification, evolution, diversity and characteristics of *Euglena gracilis*

Euglena is a unicellular flagellate member of the Excavata supergroup of organisms and phylum Euglenozoa. The phylogeny of *Euglena* is different from other algae that have been described to date as its closest neighbours in the phylogenetic tree are *Trypanosoma* and *Leishmania*, which are both non-photosynthetic human pathogens (Figure 1.1). The evolutionary history of *Euglena* is very complex with genes acquired from several organisms (Figure 1.2). The common ancestor of all eukaryotes formed the mitochondrion by engulfing a

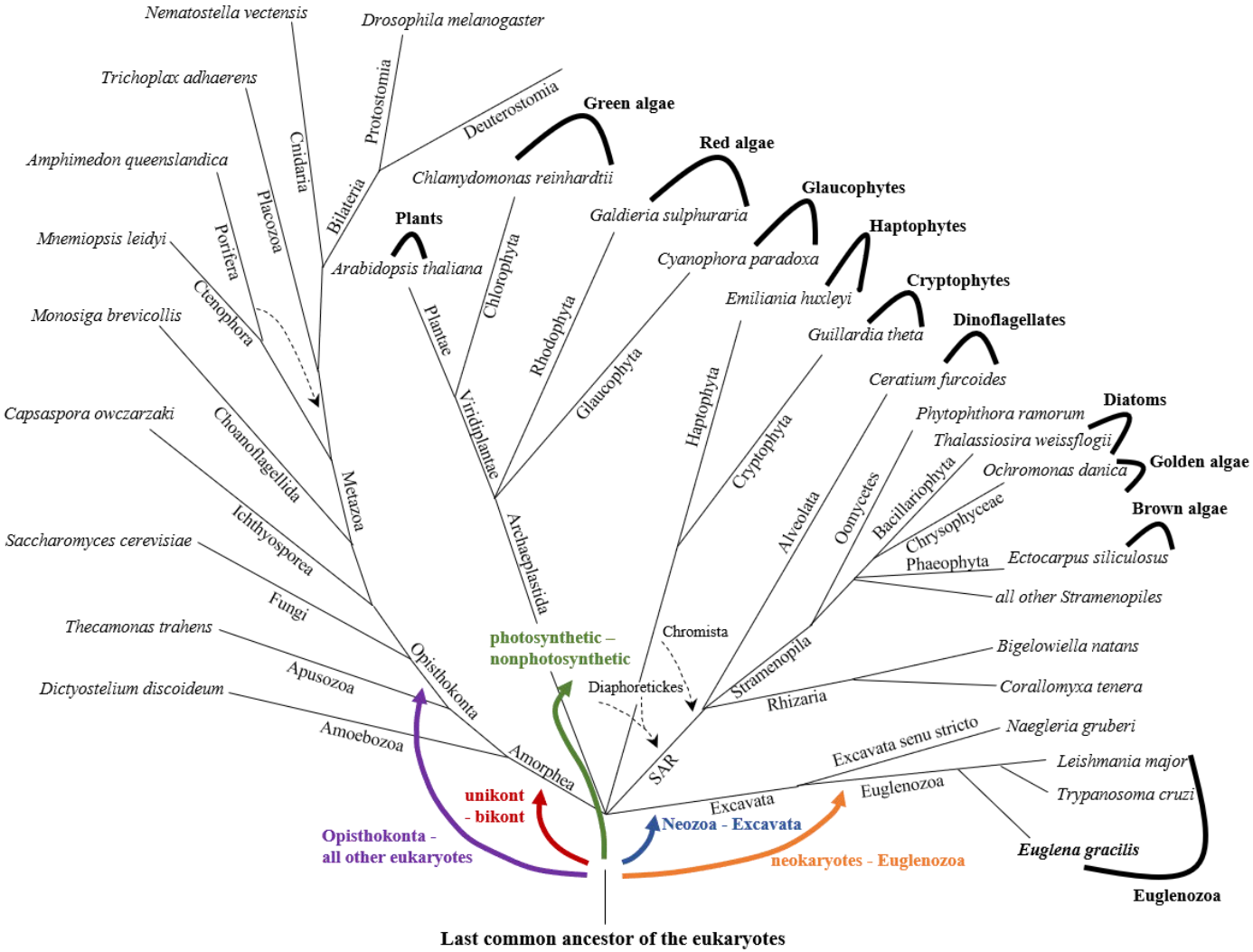

Figure 1.1: The phylogenetic tree shows the collocation of *Euglena* with respect to other eukaryotes. The dotted arrows indicate alternative collocation of some branches and the coloured arrows illustrate alternative roots of the eukaryotes. Other photosynthesising organisms have been highlighted including plants, algae, plankton and diatoms. *Euglena* are unlike any other model algae and the closest neighbours of *Euglena* are the human pathogenic protozoa *Trypanosoma spp.* and *Leishmania spp*. Modified from Kollmar (2015) [7].

prokaryote and developing an endosymbiotic relationship with the latter in which most of the genetic material was transferred to the nucleus [8]. In addition to this, the common ancestor for all photosynthetic eukaryotes subsequently formed a similar endosymbiotic relationship with a cyanobacterial cell, where most of the genetic material was transferred to the nucleus [8]. This resulting ancestral cell diversified to form plants and algae (green, red and golden). At some point in history, the non-photosynthetic ancestor of Euglenoids engulfed one of these red algal cells and transferred some of its genetic material to the nucleus, after which the red algal cell was lost [8]. This ancestral cell subsequently took up a green algal cell and formed an endosymbiotic relationship with it, transferring some nuclear and chloroplast genetic material to the Euglenoid ancestral cell [8]. The nucleus and mitochondria of this endosymbiont green algal cell were lost, leaving only the final chloroplast [8].

More than 250 species of the *Euglena* genus have been described, although this may be an overestimate as the size, colour and morphology of cells from the same species vary widely with growth phase, nutrient availability and environmental conditions [9]. *Euglena* cells are typically characterised by their elongated cell length of 12 to over 500 µm, although the shape varies from spherical to rod-shaped, changing at any time by Euglenoid movement [9]. Most cells are green in appearance, due to the presence of chlorophylls *a* and *b* [9]. All species of *Euglena* reproduce asexually by longitudinal cell division [9].

Among *Euglena* species, *E. gracilis* has been studied most extensively in the laboratory as a model algal organism for sustainable production of biomolecules. Figure 1.3 depicts a typical *E. gracilis* cell showing its main structures and organelles. The surface of *E. gracilis* cells is usually covered with a mucilage layer and the beating action of a long flagellum allows for locomotion. The eyespot, which mainly comprises β-carotene (Section 1.6.2.3), is used for phototaxis. Other organelles present in *E. gracilis* are very similar to eukaryotes, although the chloroplast is surrounded by three membranes rather than the two membranes found around chloroplasts of other green algae and higher plants [9].

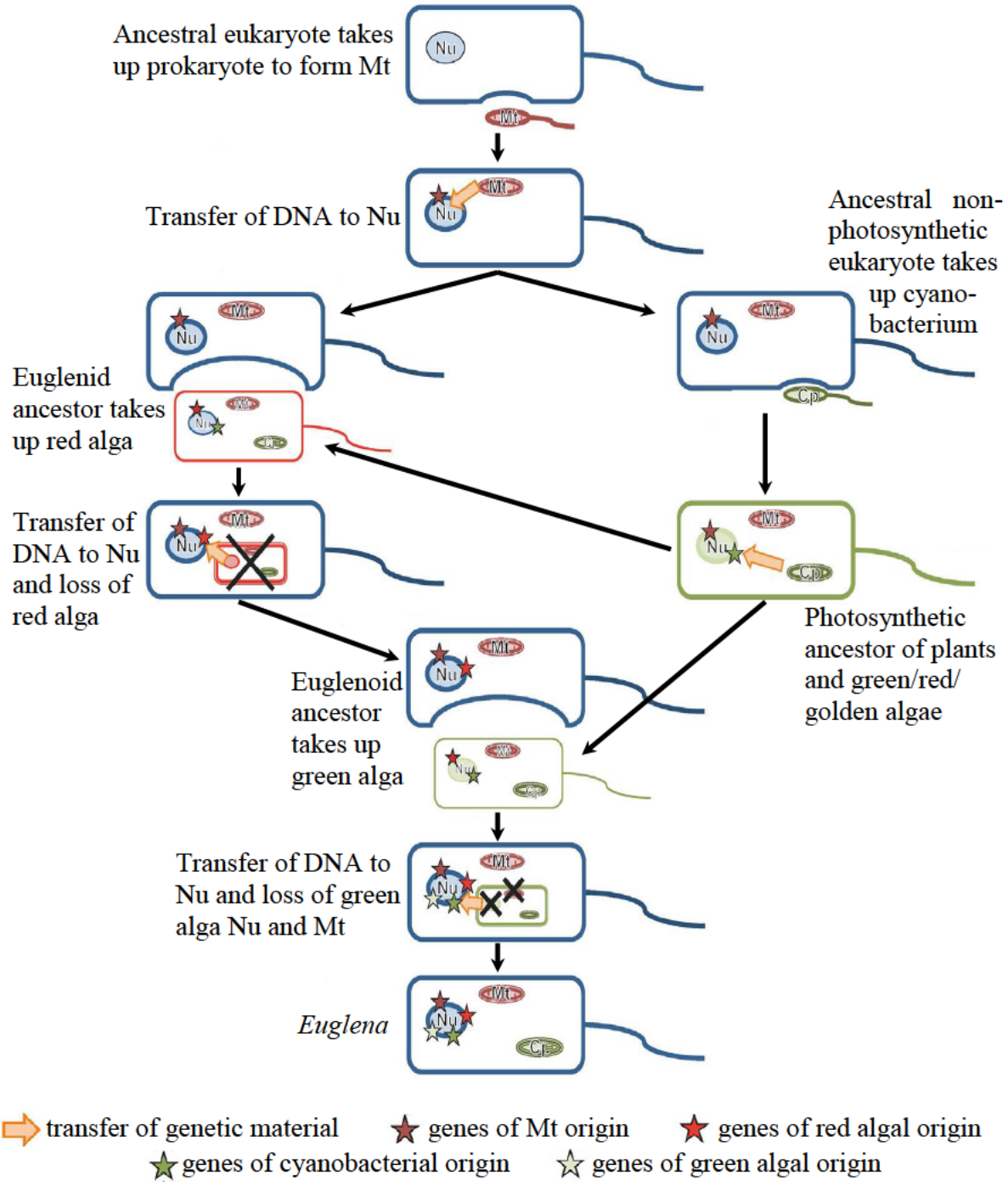

Figure 1.2: The evolutionary history of *Euglena* showing the sources of the *Euglena* genome. The genome sources include α-proteobacterium (mitochondrion), a red alga endosymbiont, and a green alga endosymbiont containing cyanobacterial cells (chloroplasts). Abbreviations: Nu – nucleus; Mt – mitochondria; Cp – chlorplast. Modified from O'Neill *et al*. (2015) [8].

E. gracilis lives in fresh and brackish water, especially those rich in organic matter, forming algal blooms. This microalga has been found growing in the acidic and toxic environments of polluted water, although no marinewater strain has been discovered to date [9]. *E. gracilis* is a mixotrophic organism, which means that it is capable of interchangeably feeding by one of two mechanisms: photoautotrophic nutrition and heterotrophic nutrition (Section 1.5). Most strains of *E. gracilis* are able to photosynthesise due to the presence of chloroplasts. However, the chloroplasts are sensitive to high temperature and antibiotics, and the cells can easily be bleached [9], losing their chloroplasts temporarily during the abiotic stress exposure or permanently after prolonged exposure to the abiotic stress. The chloroplasts of *E. gracilis* contain micro-compartments called pyrenoids that are analogous in function to carboxysomes in cyanobacteria, which sequester the main enzyme for photosynthesis, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) [10]. Since the diffusion of carbon dioxide (CO_2) in water is ten thousand times slower than in air [11], pyrenoids concentrate the CO_2 that has entered the cell to create a CO_2-rich environment for RuBisCO to fix CO_2 into glucose [12-13]. This glucose is then polymerised and stored as a complex carbohydrate called paramylon [14]. *E. gracilis* is also known to take up and metabolise a wide variety of carbon sources from the environment to produce paramylon, including glucose, glutamate, malate, pyruvate, lactate and ethanol [14].

1.3 Applications of *Euglena gracilis*

Like other microalgae, *Euglena gracilis* is a high protein-containing organism that can be consumed as a whole cell meal. Based on the current information available, the FDA has deemed *E. gracilis* biomass containing paramylon as GRAS (GRN Number 513) [15]. *E. gracilis* produces antioxidants, vitamins and pigments, such as polyphenolics, vitamin E, vitamin C and carotenoids in high yields. *E. gracilis* also contains all of the twenty protein-

building amino acids [14]. It produces high amounts of polyunsaturated fatty acids and its total nutritional value alone is adequate to support some small animal life, such as that of mice [16]. These high-value metabolites synthesised by *E. gracilis* render it a promising host for the production of nutraceuticals.

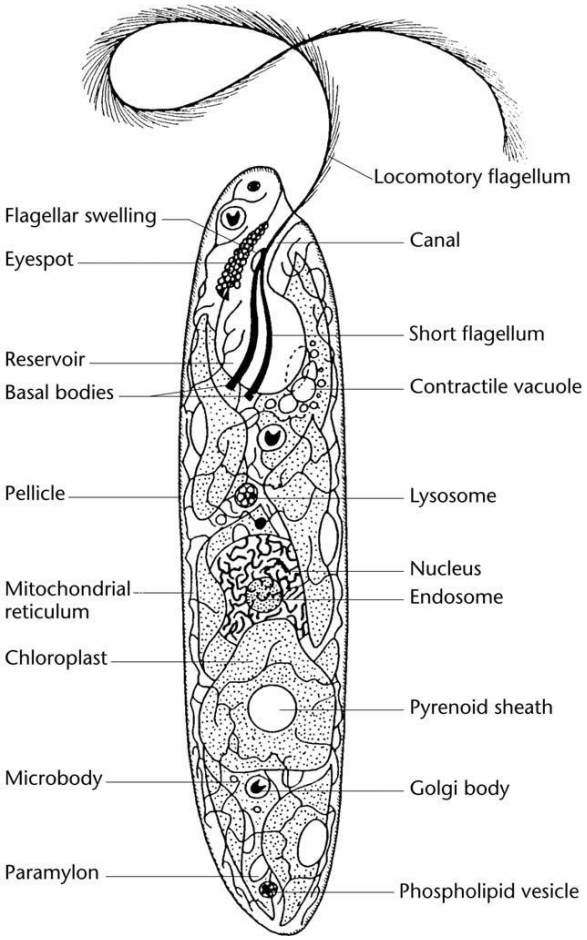

Figure 1.3: The structure of a *Euglena gracilis* cell showing its main organelles. Reproduced with permission from Buetow, 2011 [9]. The roles of some organelles are further discussed in the text.

In addition to the food and health benefits, *E. gracilis* also provides several environmental benefits. Although its natural feeding environment is that of low-light freshwaters, this organism has been found growing in numerous stressful environmental conditions including wastewater streams from industrial and mining areas containing heavy

metals [14, 17-18] and domestic sewage wastewater containing toxic substances [19]. It survives these stressful conditions by accumulating the heavy metals and toxic substances respectively, thus cleaning up the water. *E. gracilis* can outcompete microorganisms that cannot tolerate the acidic pH or the pollutants present in such wastewater [20], and its ability to photosynthesise enables it to successfully contend against other microorganisms that can resist such a low pH. Most of the mechanisms of *E. gracilis* for withstanding and accumulating heavy metals have already been explained [21-29]. *E. gracilis* is also capable of tolerating environments with high CO_2 levels and strong light [30-31]. It has a sixty-fold higher photosynthetic capacity than rice plants and a two-fold higher carbon dioxide to oxygen conversion efficiency than *Chlorella* [31], sequestering about 74 g of CO_2 m^{-3} per day [30]. This capacity of *E. gracilis* can be attributed to its ability to increase PSII reaction centers and decrease the antenna size of PSII as well as increase the size and number of PSI units, resulting in an increased chlorophyll *a*:*b* ratio during low light conditions [32]. These characteristics of *E. gracilis* render it a more efficient photosynthetic host for bioremediation of both wastewaters and CO_2 emissions than green plants and other commercially grown algae [17, 30, 31].

The physiological flexibility of *E. gracilis* also allows it to be used for other industrial applications including toxicity testing, such as ecotoxicological bioassays of nickel, cadmium and copper to monitor water toxicity [33-34], and as a bioindicator for genotoxic potentials of organic pollutants [35]. Moreover, *E. gracilis* can be used for biofuel production and when grown on wastewater this organism has been reported to accumulate high amounts of palmitic, linolenic and linoleic acids, which increase the quality of biodiesel produced from biomass [19]. Anaerobic digestion of the *E. gracilis* biomass also shows promise in biogas production [36]. *E. gracilis* is widely used for vitamin B_{12} bioassays from sera [37], and photo-, gravi- and polaro-taxis studies [38-39]. Recently, this microalga has also been used for intracellular biosynthesis of functional colloids such as superparamagnetic ferrihydrite nanoparticles [40]. Moreover, structurally modified paramylon from *E. gracilis* has been demonstrated to be useful

for the synthesis of several compounds, such as self-assembled polysaccharide nanofiber, succinylated and carboxymethylated paramylon nanofibers, self-standing optical films using acetylparamylon, and thermoplastics by adding acyl groups to paramylon [41-45]. *E. gracilis* extract is also used in the cosmetic industry as a cell energiser for maintaining skin firmness and tone [46]. The industrial applications of *E. gracilis* are summarised in Figure 1.4.

Despite the numerous industrial applications of *E. gracilis*, genetic information and molecular tools for genetic manipulation are limited, with genetic transformation described for only the chloroplasts of *E. gracilis* [47, 48]. A method for nuclear genome transformation has not been described yet and improvement of strains for industrial applications is thus largely dependent on adaptive evolutionary strain engineering and optimisation of cultivation conditions.

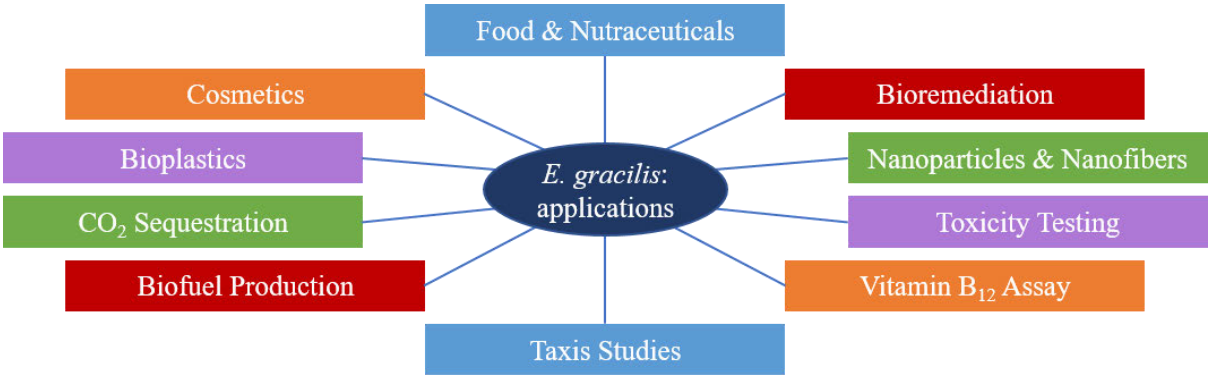

Figure 1.4: Potential industrial applications of *E. gracilis*. Some of these applications have already been commercialised, such as the use of cell extracts in skin cosmetics by Sederma SAS, France and paramylon ARX by Naturally Plus, Japan.

1.4 *Euglena gracilis* strains

Euglena gracilis strains can be divided into three groups depending on their ability to photosynthesise or grow heterotrophically [49, 50] as shown in Table 1.1. Some strains of *E.*

gracilis, such as *E. gracilis* L, are obligate phototrophs incapable of growing on any source of organic carbon in the dark, but can readily grow in the presence of light [51]. On the other extreme are obligate heterotrophs that cannot photosynthesise. This group includes all bleached mutants and naturally occurring non-photosynthetic variants of *E. gracilis*, such as the WZSL spontaneous non-photosynthetic mutant described by Barsanti *et al.* (2001) [52]. Some obligate heterotrophs can utilise various organic carbon sources [17, 52], while others are incapable of growing on glucose, such as the Mainx and Vischer strains [49, 53]. A third intermediate group exists (mixotrophs), comprising of *E. gracilis* Z [52], *E. gracilis* var. *bacillaris* [54-55] and *E. gracilis* var. *saccharophila* [56] strains among others, which can grow on organic carbon sources, as well as photosynthesise. This group can be further divided into strains that have a preference towards photosynthesis even when organic carbon sources are present, although they can use both (*E. gracilis* Z and var. *bacillaris*), and strains that readily utilise organic carbon sources before resorting to photosynthesis (*E. gracilis* var. *saccharophila*) [49]. The genomic, transcriptomic, proteomic and metabolomic profiles of the *E. gracilis* strains have never been compared, and thus the underlying molecular reasons for these differences are unclear.

Table 1.1: Classification of *E. gracilis* strains based on their method of nutrient assimilation.

Method of Nutrient Assimilation	**Example of Strain**
Obligate phototroph	*E. gracilis* L
Obligate heterotroph	*E. gracilis* WZSL, bleached mutants
Mixotrophs: - preference for photosynthesis - preference forheterotrophy	 *E. gracilis* Z, *E. gracilis* var. *bacillaris* *E. gracilis* var. *saccharophila*

1.5 Modes of cultivation

Euglena gracilis can grow under photoautotrophic (PT) conditions using inorganic CO_2 from the atmosphere in the presence of light, heterotrophic (HT) conditions using various organic carbon sources from its surroundings, and mixotrophic (MT) conditions using both inorganic CO_2 and organic carbon sources in the presence of light. This metabolic versatility is owed to the complex evolutionary history of *E. gracilis* (Section 1.2), during which it acquired genes from an α-proteobacterium (mitochondrion), a red alga endosymbiont, and a green alga endosymbiont containing cyanobacterial cells (chloroplasts) that were retained [8]. Figure 1.5 shows the difference in appearance of cultures grown under the three conditions. The cultures appear bright green during PT cultivation as they mainly carry out carbon fixation via photosynthesis (Calvin cycle). During HT cultivation, the cultures appear yellowish-white as they mainly assimilate exogenous organic carbon in the absence of light. MT cultures, on the other hand, carry out photosynthesis and heterotrophic nutrient assimilation simultaneously, thus appearing yellowish-green.

During PT cultivation, accumulation of biomass and paramylon is restricted by the slower rate of growth by photosynthesis. It is difficult to control cultivation conditions and maintain sterile conditions in large-scale photobioreactors that support PT growth, such as open cultivation ponds, artificial raceway ponds and tubular photobioreactors [57]. Even at a low pH (2.8), and periodic addition of 2% ethanol and the antifungal nystatin to the medium, the contamination by other organisms in such open systems can be as high as 10% of the total cell density [58]. Moreover, the light intensity from the sun can surpass 2000 µmol photons m^{-2} s^{-1} during midday, and consequently, for a significant part of the day the cells do not grow at maximum efficiency, as the maximum photosynthetic efficiency of most species of microalgae is at a light intensity between 200 and 400 µmol photons m^{-2} s^{-1} [59]. Above this light intensity range, the growth rate is dramatically reduced owing to photoinhibition caused by the light

damage to photosystem II (PSII) [59]. One way to overcome this is by reducing the number of light-harvesting complexes per chloroplast to lower sunlight absorption of individual chloroplasts [60]. At higher light intensities this will allow more efficient light dispersion in high-density cultures and thus reduce photoinhibition [60]. However, this requires intensive photosynthesis studies of *E. gracilis* and significant reengineering of its chloroplast genome, which has not been accomplished yet. On the other hand, closed sterilisable photobioreactors for PT cultivation using artificial lights can have very high operating costs, and are thus limited to high-value products [61]. Nevertheless, many metabolites of interest in *E. gracilis* are produced in higher concentrations during light cultivation, including most of the antioxidants, such as ascorbate, α-tocopherol and carotenoids, and several of the important free amino acids.

HT cultivations do not require light and thus offer the scope of acquiring very high cell densities by using optimum culture conditions throughout the cultivation. It is more economical to run bioreactors under HT conditions than photobioreactors that require artificial light and are limited by technical problems of development such as self-shading of light as algal density increases [57]. In addition, the ease of sterilisation, maintenance and control of culture conditions outweigh the operating costs of open pond systems. However, it has been reported for other microalgal systems that the efficacy of HT cultivation in producing valuable metabolites is limited compared to PT cultures, especially those related to photosynthesis such as α-tocopherol, carotenoids and other antioxidants [57]. This is largely owing to lower intracellular metabolite concentrations, especially those associated with the chloroplast [57]. Although heterotrophically grown *E. gracilis* can accumulate much higher biomass yields and synthesise higher amounts of paramylon than its PT counterpart [62], these cultures contain lower amounts of the other beneficial metabolites such as ascorbate, α-tocopherol, carotenoids and unsaturated fatty acids. However, Ogbonna *et al.* (1999) showed that this problem can be overcome by using a two-step sequential HT-PT cultivation method, where the biomass is allowed to increase by first growing the cells heterotrophically [63]. The cells are then exposed

to light and metabolites related to light cultivation, such as antioxidants, are allowed to accumulate during PT growth [63]. The limitation of this system is that exposure to light quickly breaks down the paramylon accumulated during HT growth, as discussed in Section 1.6.1.2.

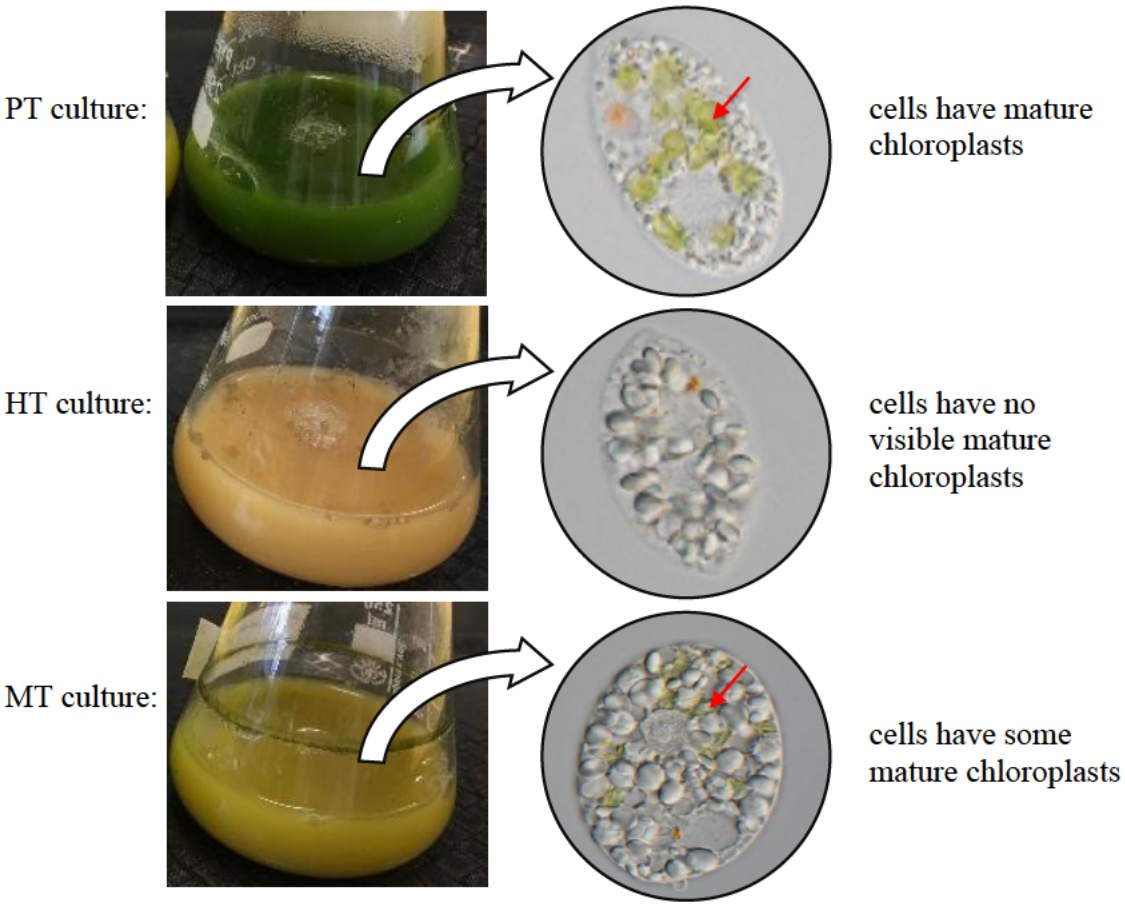

Figure 1.5: Appearance of *E. gracilis* var. *saccharophila* cultures and cells within cultures grown under photoautotrophic (PT), heterotrophic (HT) and mixotrophic (MT) cultivations. PT cultures appear green with cells having mature chloroplasts (red arrow), while HT cultures appear yellowish-white with no mature chloroplasts visible. MT cultures, on the other hand, appear yellowish-green with some mature chloroplasts (red arrow).

A greater understanding of metabolic pathways is required to reengineer *E. gracilis* to optimise production of paramylon and other metabolites under either growth condition. An alternative is to grow *E. gracilis* under MT conditions, during which it can utilise organic carbon sources from its surroundings to grow rapidly and increase in biomass. MT cultivation of cells can also allow paramylon production comparable to HT cultures. Moreover, simultaneous exposure to light can enable cells to accumulate other metabolites that are predominantly accumulated under light cultivation, such as antioxidants. It is thus important to compare metabolite production between MT and the other two cultivation conditions. A study of the metabolic pathways under the different cultivation conditions will give insight into which pathways need to be induced or reengineered for optimum production of all desired metabolites.

1.6 Metabolites produced by *Euglena gracilis*

1.6.1 Paramylon

Paramylon is the primary carbon reserve of *Euglena gracilis*. It is a robust carbohydrate storage with linear unbranched glucose chains linked by β-1,3 glycosidic bonds, which are difficult for most other organisms to hydrolyse [14]. Paramylon has a fibrillar structure of triple helices of microfibrils measuring between 4 to 10 nm in thickness [62]. The molecular mass of paramylon is estimated to be higher than 500 kDa, and the degree of polymerisation is difficult to determine due to its high crystallinity of about 90% [62, 64]. Paramylon is a unique β-1,3-glucan, as such glucans are not known to be deposited as granules in any other organism but Euglenoids [62]. Paramylon is a versatile storage molecule in that it can be dispersed throughout the cytoplasm, form caps that cover pyrenoids, be amassed together or be formed as large solitary granules at constant locations (Figure 1.6) [52]. Paramylon granules can be membrane-bound bodies or be located in membrane-deficient cavities [52]. Electron

microscopy of the granule membrane has revealed it to be a single membrane of normal lipid bilayer, containing varying numbers and sizes of intramembranous particles (IMP) [65].

There are two routes reported to be involved in the biosynthesis of paramylon. The photosynthetic route involves CO_2 fixation in the pyrenoids, in which most of the paramylon granules are closely located just outside the chloroplasts [62]. Paramylon is also synthesised heterotrophically in the presence of surplus organic carbon in the culture medium [14]. This route involves the swelling and vesiculation of mitochondria to form paramylon granules that are eventually released and distributed throughout the cytoplasm [66]. Heterotrophically grown cells can accumulate up to a six-fold higher paramylon content than photoautotrophically grown cells, which are thought to activate a regulator for paramylon synthesis [62, 67]. *E. gracilis* can survive prolonged periods of light deprivation and environmental carbon depletion by catabolising this storage carbohydrate for growth and maintenance [14].

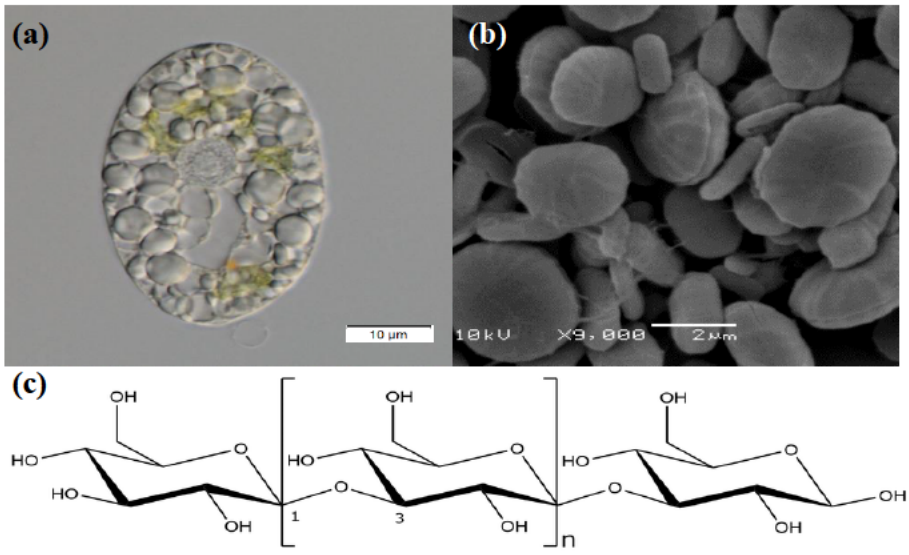

Figure 1.6: Paramylon granules are distributed throughout *E. gracilis* cells (a) and can have a wide range of sizes. However, images taken with a scanning electron microscope (b) reveal that the shapes of native paramylon granules are quite similar [68]. This granular structure is formed from a linear unbranched chain of glucose monomers (c) that form a triple helix, which in turn forms microfibrils.

1.6.1.1 Applications of paramylon

Several studies have shown the effectiveness of β-glucans to treat various diseases. Like curdlan, another linear β-glucan, paramylon has the potential to be used in medicine and veterinary sciences as an immune-stimulant of macrophages and as an anti-tumour drug [69]. Koizumi *et al.* (1993) observed that sulphated derivatives of paramylon showed anti-HIV properties by inhibiting the cytopathic effect of HIV and the expression of HIV antigen [70]. Cytokine-related immune-potentiating activities of paramylon have also been determined [71]. Additionally, other studies on rats and mice have shown protective effects of paramylon through an antioxidative mechanism on acute liver injury caused by carbon tetrachloride [72], and inhibition of atopic dermatitis-like skin lesions [73]. Studies have also shown that β-glucans in general have activity in both humans and animals as anti-tumour and anti-infection drugs [74]. Wang *et al.* (1997) suggested that β-glucans lower cholesterol levels when incorporated into the diet of hamsters [75], and Wood (1994) showed that they can aid in lowering the postprandial blood glucose levels and promote insulin response [76]. Currently, the sources of non-cellulosic β-glucans, including both the 1,3- and 1,6-polymer chains, are yeast, and other fungal and plant cell walls [14]. Among these, *Saccharomyces cerevisiae* (baker's yeast) is industrially used to extract β-glucans [52]. However, *E. gracilis* can accumulate a much larger quantity of β-glucan, especially during HT cultivation, during which it is reported to contribute up to 90% of the dry mass [52].

1.6.1.2 Paramylon biosynthetic and degradation routes

Paramylon is synthesised as the primary carbon storage of *E. gracilis*, analogous to starch in plants, in response to surplus carbon in the surroundings via photosynthesis as well as up-take of organic carbon. A route for paramylon biosynthesis has been predicted from the

transcriptome (Figure 1.7), and enzymes for the different steps have been predicted from the recently published transcriptomes of two strains of *E. gracilis* [77-78]. Paramylon is synthesised from uridine diphosphate (UDP)-glucose by paramylon synthase, which is a β-1,3-glucan synthase. Such enzymes are classified as carbohydrate-active enzymes (CAZys) belonging to the glycosyl transferase family 48 (GT48) [79]. This family of enzymes usually consists of very large transmembranous proteins that transfer a glucose residue from UDP-glucose to a growing β-1,3-glucan chain [79] and includes enzymes such as callose and curdlan synthase. Yeast and other fungal glucan synthases are reported to have UDP-glucose binding consensus sequences of RXTG and RVSG facing the cytoplasmic side [80-81], while bacterial and plant glucan synthases have the conserved D,D,D and QXXRW motifs [82]. Between two [77] to four [78] candidates belonging to the GT48 family have been identified from the *E. gracilis* transcriptome. Silencing of the paramylon synthase gene expression with double-stranded RNA showed that one synthase candidate was predominant and was indispensible for paramylon biosynthesis [79]. The role of the other synthase identified in the same report is unclear until now, as silencing the gene had no effect on paramylon accumulation [79]. The proteomic expression levels of the GT48 family paramylon synthase candidates in the *E. gracilis* has not been studied, and the effect of growth condition on the relative abundance of these proteins has not been reported until now.

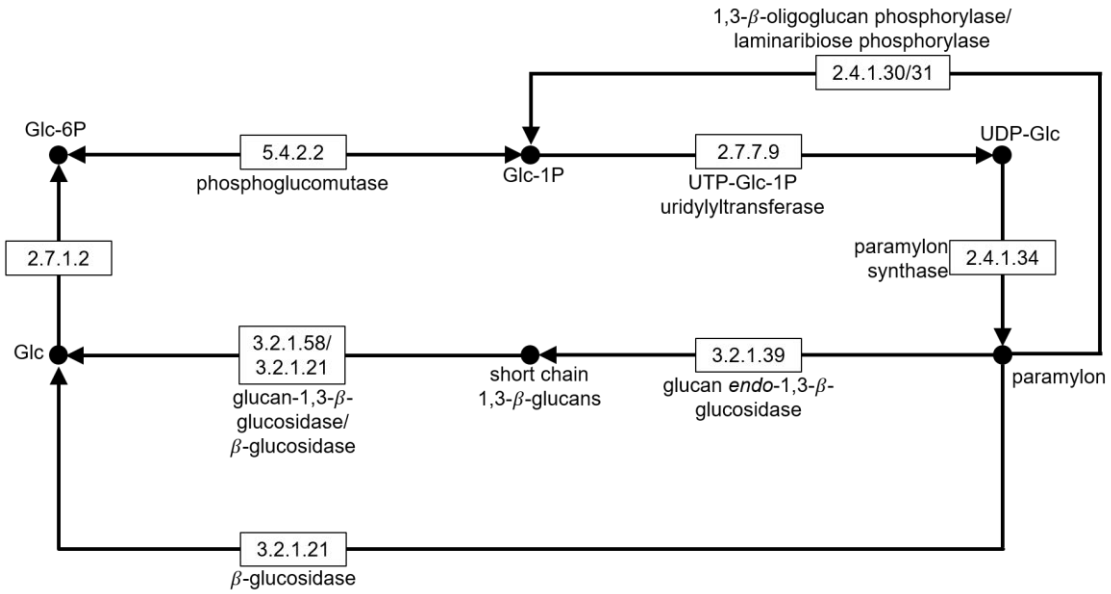

Figure 1.7: The anabolic and catabolic pathways of paramylon, showing biosynthesis from UDP-glucose using β-1,3-glucan synthase (paramylon synthase) enzymes, and degradation to glucose using *endo*-β-1,3-glucan hydrolases and *exo*-β-1,3-glucan hydrolases. Abbreviations: Glc – glucose; P – phosphate; UDP – uridine diphosphate; UTP – uridine triphosphate. This pathway has been created using information from the published transcriptome [78].

Paramylon is broken down (Figure 1.7) when cells require a high amount of energy, such as during carbon starvation, oxygen enrichment and other environmental stresses [79]. Paramylon is also converted to wax esters during anaerobic conditions by a process called wax ester fermentation, which enables the generation of ATP from paramylon without any energy loss during anaerobiosis [77]. The most rapid catabolism of paramylon occurs during organelle development, especially chloroplast development [83]. When dark-adapted *E. gracilis* cells are exposed to light, they rapidly break down paramylon for the development of chloroplasts from proplastids [83]. However, when bleached mutant *E. gracilis* cells that do not have mature chloroplasts are transferred from dark growth condition to light condition, the degradation of paramylon is very slow [83]. Investigation into paramylon catabolism showed that the carbon released was utilised mainly for chloroplast lipid synthesis, followed by chloroplast protein

synthesis and as the CO_2 by-product of respiration [84]. Very little of the carbon from paramylon catabolism was redistributed to the cytosolic reactions that required carbon [84].

The recently published transcriptomes of *E. gracilis* revealed the presence of several glucan hydrolases, including both *endo*- and *exo*-β-1,3-glucan hydrolases [77-78]. The transcriptome contains candidate glucan hydrolases belonging to the GH81, GH17 and GH64 families of *endo*-glucan hydrolases, which cleave along the middle of the β-1,3-glucan chains to form shorter glucan chains [78]. Beta-glucosidases and *exo*-glucan hydrolases belonging to GH1, GH2, GH3, GH5, GH30 and GH55 families, which further hydrolyse the short-chain glucans to glucose were also identified in the transcriptome [78]. Several laminaribiose phosphorylases have also been characterised at the enzymatic level in *E. gracilis* [85-86]. However, the transcriptomic sequences of these phosphorylases are very different to their bacterial counterparts that belong to the GH94 family, and thus may represent a new family of glucan hydrolases [78]. Recently, the first functional *endo*-β-1,3-glucan hydrolase from *E. gracilis* has been identified [87]. However, the presence of most of the remaining glucan hydrolases from the transcriptome has not been reported at the proteomic level until now, and the effect of cultivation on their expression levels have yet to be determined.

1.6.2 Antioxidants

Reactive oxygen species (ROS) are highly reactive radical and non-radical oxygen-containing chemicals including peroxides, superoxides hydroxyl radicals and singlet oxygen [88-89]. They are formed as a result of partial reduction of oxygen and are a natural byproduct of oxygen metabolism [88]. They are produced by the electron transport system in the mitochondria and plasma membrane [90], as well as metabolic processes localised in other cellular compartments, such as the cytosol, nucleus, mitochondrial matrix, endoplasmic reticulum, peroxisomes, Golgi apparatus, lysosomes, cell wall and apoplast [91-93]. The

primary photochemical reactions of photosynthesis also produce ample O_2, which leads to the generation of a large amount of ROS [90, 92]. Although ROS play an important role in cell signalling during cell proliferation, survival and homeostasis, the presence of large amounts of ROS causes oxidative stress within cells that can damage nucleic acids, proteins and lipids [88]. Environmental stresses including hypoxia, drought, low temperature, salinity, radiation, high metal toxicity and pathogenic attacks can further elevate cellular ROS leading to oxidative stress [91-93]. In humans, cellular oxidative stress has been shown to have a strong correlation with diseases such as cardiovascular diseases [94], cancer [95] and Alzheimer's disease [96].

Oxidative stress can be remediated by antioxidants, which can be enzymatic and non-enzymatic. The enzymatic antioxidants include superoxide dismutase, catalase, glutathione peroxidase, guaiacol peroxidase, ascorbate peroxidase, monodehydroascorbate reductase and glutathione reductase [92, 95]. On the other hand, ascorbate, tocopherols and tocotrienols, carotenoids, glutathione, thioredoxin, lipoic acid, phenolics, flavonoids, polyphenols and selenium, among others, can serve as non-enzymatic antioxidants [92, 95]. *E. gracilis* has the potential to produce large amounts of both non-enzymatic and enzymatic antioxidants, as has been demonstrated from its transcriptome [78], and it shows great promise as a source of antioxidant supplement. The transcriptome suggests that *E. gracilis* can utilise many alternative pathways to produce antioxidants and this may indicate why it is an efficient producer of antioxidants. Ascorbate, α-tocopherol and carotenoids are among the predominant non-enzymatic antioxidants found in *E. gracilis*.

1.6.2.1 Ascorbate

Humans cannot synthesise ascorbate naturally and are thus completely dependent upon food sources to meet dietary requirements [97]. Ascorbate is the most abundant water-soluble non-enzymatic antioxidant in plants and algae [90]. Currently, the major source of ascorbate

and other nutraceutical antioxidants as supplements is plants [89]. As a source of ascorbate, *Euglena gracilis* has been reported to produce up to 4.25 mg g^{-1} of dry mass [98]. This is much lower than the common citrus fruit juices [99-100] and spinach leaves [101], but higher than other algae including both green and brown edible seaweeds [102-103].

Ascorbate is ubiquitous to all subcellular compartments and is particularly prevalent in the chloroplast [89]. Ascorbate takes part in the ascorbate/glutathione cycle, where the enzyme ascorbate peroxidase uses it as an electron donor to reduce H_2O_2 [90]. There are three major biosynthetic pathways for ascorbate (Figure 1.8) designated as the D-glucuronate/L-gulonate (animal pathway) [104], the D-mannose/L-galactose pathway (plant pathway) [105], and the D-galacturonate/L-galactonate pathway (*Euglena* pathway) [106]. The lack of clear evidence of an ascorbate biosynthetic pathway in prokaryotes suggests that the ability to produce ascorbate either evolved convergently in eukaryotes or through the modifications of an ancestral pathway [107]. All three major routes lead to the synthesis of an aldonolactone precursor (Figure 1.8), which is then converted to ascorbate [107]. The transcriptome of *E. gracilis* revealed that this organism contains candidates for every enzyme of all three major pathways, although some of the transcripts could only be weakly annotated owing to sequence similarity with other enzymes [78]. Nevertheless, the main pathway used by *E. gracilis* for ascorbate biosynthesis appears to be via D-galacturonate/L-galactonate [78, 106]. The enzyme ascorbate peroxidase has also been characterised in *E. gracilis* [108]. This enzyme scavenges H_2O_2, as well as superoxide radicals and alkyl hydroperoxides, preventing oxidative stress and lipid hydroperoxidation that can lead to membrane damage [108].

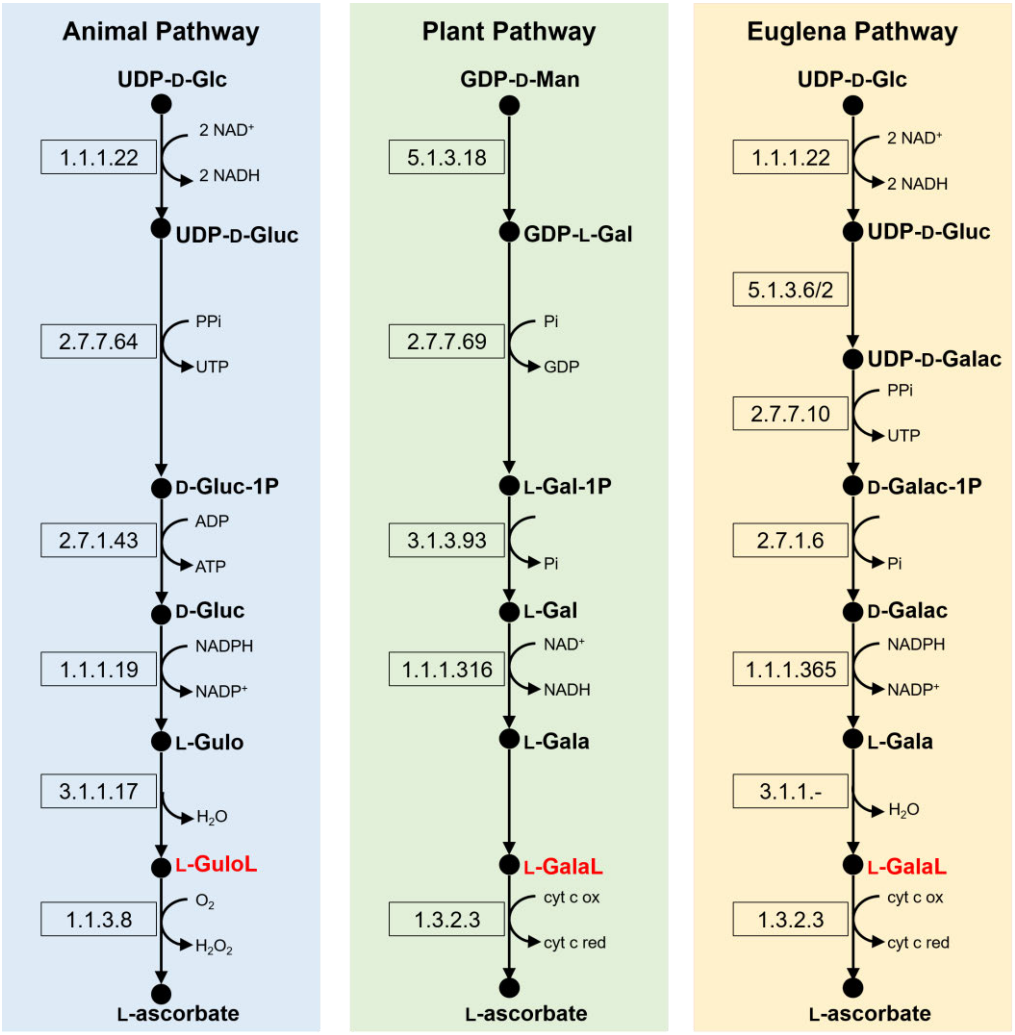

Figure 1.8: The three major ascorbate biosynthetic pathways for animals (blue box), plants (green box) and *Euglena* (yellow box). The aldonolactone precursors are highlighted in red font, and the enzyme class numbers are in black rectangles. Abbreviations: Glc – glucose; Man – mannose; Gal – galactose; Gluc – glucuronate; Galac – galacturonate; Gulo – gulonate; Gala – galactonate; GuloL – gulono-lactone; GalaL – galactono-lactone; P – phosphate. Modified from Wheeler *et al.* (2015) [107].

1.6.2.2 Alpha-tocopherol

The antioxidant vitamin E is a class of fat-soluble methylated phenols that are important food preservatives in the food processing industry [57, 63]. Vitamin E is used to remediate oxidative stress and is becoming popular in recent times as a nutraceutical to help neutralise the effects of pollution and stress on urban residents [14]. Vitamin E exists in many isomeric forms including alpha, beta, gamma and delta tocopherols and tocotrienols, of which the α-isomer of tocopherol has the highest antioxidant activity [14, 57].

The main source of vitamin E in the human diet is vegetable oils, such as sunflower, soybean, olive and corn, in which α-tocopherol is accompanied by a large quantity of the other isomers with lower antioxidant activity [14, 63]. It is also extracted from vegetable oils for pharmaceutical applications that require pure α-tocopherol [63], but this is a difficult task due to the similarity in structure and molecular weight between the isomers [14]. A synthetic version of α-tocopherol also exists, but it has a much lower biopotency than its natural counterpart [14]. Many other microalgae, yeast, moulds and bacteria also synthesise tocopherols [109-110]. However, *Euglena gracilis* has the potential to be a better source of α-tocopherol than vegetable oils, microorganisms and synthetic systems. *E. gracilis* was reported to have the highest cellular content of tocopherols among 56 genera of organisms tested, with a concentration of up to 7.35 mg g^{-1} cells [109, 111]. Additionally, the percentage distribution of tocopherols in light-adapted green cells (Figure 1.9) reveals that more than 97% of the total cellular tocopherols in *E. gracilis* is in the highly bioactive α-form [112]. Development of *E. gracilis* as an efficient host production system of α-tocopherol is therefore expected to provide a more stable, higher yielding and cheaper natural source of this antioxidant [57].

There are contradictory reports in the literature regarding the distribution of tocopherols within cells [111], with Shigeoka *et al.* (1986) reporting mitochondria had the highest tocopherols under light cultivation [112], and Threlfall and Goodwin (1967) reporting

chloroplasts had the highest tocopherols under light cultivation [113]. Other conflicting reports include the notion of higher tocopherol amount in bleached mutants (without chloroplasts) than photosynthetic strains [114], no correlation between the amount of chlorophyll *a* and α-tocopherol [114-115], and concurrent increase of chlorophyll and α-tocopherol [116]. Tocopherol content has been reported to be associated with light intensity through its effects on photosynthesis, with PT cultures having a much higher content than HT cultures grown in the dark and MT cultures grown in the light [109, 113]. However, Kusmic *et al.* (1999) showed that increasing light intensity stimulated α-tocopherol synthesis regardless of the presence of chloroplasts, even in bleached strains of *E. gracilis* that lacked chloroplasts [114]. The variations in these reports suggest that both mitochondria and chloroplasts may play important roles in α-tocopherol synthesis [111].

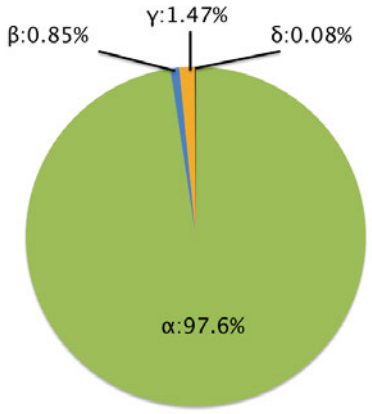

Figure 1.9: Percentage distribution of tocopherol isomers in light-adapted green *E. gracilis* cells. Chart created from data provided in Shigeoka *et al.* (1986) [112]. The Greek letters represent the tocopherol isomers. The percentage indicates the ratio of the given isomer to the total tocopherols present.

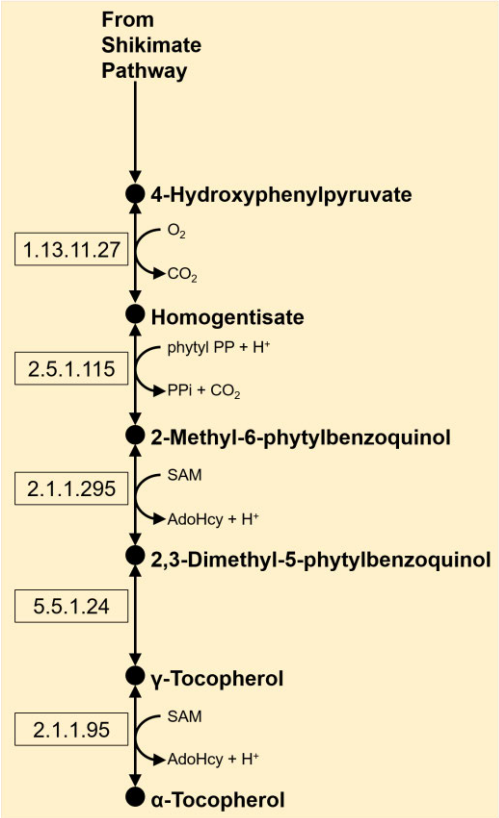

Figure 1.10: The proposed biosynthetic pathway of α-tocopherol in *E. gracilis*. Abbreviations: phytyl PP – phytyl pyrophosphate; PPi – pyrophosphate; SAM – S-adenosyl-L-methionine; AdoHcy – S-adenosyl-L-homocyteine. Modified from Shigeoka *et al.* (1992) [117].

A hypothetical biosynthetic pathway for α-tocopherol connected to the shikimate pathway has been proposed in *E. gracilis* [117]. In this proposed pathway, the α-tocopherol is synthesised from 4-hydroxyphenylpyruvate and homogentisate via 2,3-dimethyl-5-phytylbenzoquinol and γ-tocopherol intermediates, rather than via β- and δ-tocopherol intermediates as shown in Figure 1.10 [117]. This pathway is also connected to the phenylalanine and tyrosine biosynthetic pathways for which 4-hydroxyphenylpyruvate is an intermediate, and thus the synthesis of α-tocopherol may be influenced by the metabolism of these two amino acids [117]. The enzymes of the biosynthetic pathway of α-tocopherol have been predicted from the transcriptome [78]. However, their expression has not been confirmed in the proteome, and the effect of growth condition on enzyme expression has not been

revealed. Extensive research on metabolic engineering in plants for higher accumulation of tocopherols has been carried out [111]. Genetic engineering experiments have also been carried out on the rate-limiting reaction in the α-tocopherol biosynthesis pathway to increase tocopherol content by five-folds in the cyanobacterium *Synechocystis* sp. [118]. However, the actual proteomic knowledge of the biosynthetic pathways and molecular tools, as well as identification and characterisation of biosynthetic enzymes under different cultivation conditions for increasing α-tocopherol synthesis in *E. gracilis* remains quite poor.

1.6.2.3 Beta-carotene

Another source of antioxidants in *Euglena gracilis* is carotenoids. These fat-soluble antioxidants that are derivatives of tetraterpenes are found in chloroplasts and in the eyespot of the cell [119]. In the chloroplast, carotenoids absorb light in the wavelengths where absorption by chlorophyll is low, and afterward transfer this absorbed energy to chlorophyll, increasing the cell's photosynthetic capacity [120]. Carotenoids also help to protect *E. gracilis* against photodynamic destruction that may be caused by chlorophyll by preventing free radical damage to the cell [120]. In the eyespot, the high density of carotenoids helps in phototaxis [120] guiding the cell towards light and orienting it in a way that allows optimum light absorption. Beta-carotene makes up the second highest carotenoid in *E. gracilis*, accounting for about 11% of the total carotenoids [120].

Beta-carotene is consumed by humans to prevent oxidative damage, thus rendering it important in clinical and nutritional fields. It occurs as the provitamin A, which means it is metabolised into retinol after absorption from the gut [121]. It plays an important role in cell growth and differentiation, and is also required for immunological functions and vision [121]. Products of carotenoid degradation are important fragrant components of perfumes [122].

Carotenoid synthesis has thus far mainly been described in cyanobacteria and plants [121], and only a few reports on carotenoid determination exist in *E. gracilis* [123-124].

1.6.3 Free amino acids

Euglena gracilis can synthesise all of the 20 protein-building amino acids, which are very important for the agriculture industry, where they are used as food additives in animal feeds [14]. Eight of these amino acids are also considered essential in the human diet (isoleucine, valine, leucine, phenylalanine, tryptophan, threonine, lysine and methionine) [125], and four more are considered essential in the diet of infants and growing children (arginine, cysteine, histidine and tyrosine) [126]. *E. gracilis* is especially abundant in arginine, which acts as a nitrogen reserve, equivalent to the paramylon carbon reserve [127]. Although some of the enzymes of the various biosynthetic pathways of the amino acids are known in *E. gracilis*, the majority of them have not been identified in this organism. Moreover, the expression levels of these enzymes and their relation to different strains and cultivation conditions have not been studied comprehensively.

1.6.4 Other metabolites

Euglena gracilis also produces other metabolites of interest such as wax esters, lipids, polyunsaturated fatty acids, pigments, and other enzymatic and non-enzymatic antioxidants. Under anaerobic conditions, *E. gracilis* converts paramylon to wax ester by a process called wax ester fermentation [77]. Mahapatra *et al.* (2013) have reported that Euglenoids can achieve a lipid content of 24.6% (w/w) of the dry weight, a large portion of which comprises polyunsaturated fatty acids, including linolenic acid (23%) and linoleic acid (22%) [19]. About 46% of the lipid content also consists of palmitic acids [19]. These constituents are not only

beneficial as nutraceutical and cosmetic supplements, but are also important in biofuel production [19]. Besides ascorbate, α-tocopherol and carotenoids, *E. gracilis* produces other non-enzymatic and enzymatic antioxidants including glutathione and glutathione reductase [128], glutathione peroxidase [129], ascorbate peroxidase [108], monodehydroascorbate reductase and dehydroascorbate reductase [130] and superoxide dismutases [131].

1.7 Proteomics as a tool for understanding gene expression

Proteomics is a large-scale study of the global proteome of an organism. The proteome of an organism is not constant, varying from cell to cell and changing over time, and is reflective of the underlying transcriptome to some degree. Proteomics can be used to investigate proteins involved in metabolic pathways and biological processes and the expression profiles of proteins varying under a given set of conditions [132]. Proteomics can also be used to determine the rates of synthesis, degradation and steady-state abundance of proteins, and post-translational modifications of proteins [132]. Localisation of proteins to subcellular compartments and their movements between compartments, and how proteins interact with one another can also be studied using proteomics data [132].

The improvement of *Euglena gracilis* as a nutraceutical, synthesising a desirable ratio of secondary metabolites under a single optimum growth condition requires the identification of biosynthetic and degradation enzymes and an understanding of the metabolic pathways. Thus, studying gene expression at the transcript or protein level is essential to comprehend the molecular mechanisms involved in metabolite production. However, transcriptomic changes do not always correlate with proteomic changes and proteins are the final products of gene expression. Moreover, *E. gracilis* tends to extensively alter mRNA sequences prior to translation and sometimes regulates protein expression at the post-transcriptional level [78, 133-141]. Thus, molecular responses in *E. gracilis* cannot be completely understood without

proteomic analysis. The genome of *E. gracilis* has recently been sequenced [142], but it has not been fully annotated or made publicly available yet. The transcriptomes of two strains of *E. gracilis* have been published recently as well, which suggest an exhaustive metabolic capability with over 30,000 predicted proteins [77, 78]. However, there have been only two proteomic studies reported in *E. gracilis* as described in Section 1.8 [134, 143].

1.7.1 Key challenges associated with proteomics studies of *Euglena gracilis*

There remain several challenges to carrying out proteomic studies in organisms with a high abundance of metabolites, especially in regard to sample preparation. The common protein extraction methods remove these compounds either before or after extraction. Unlike plant cells and several other microalgae, *E. gracilis* does not contain any cell walls [7]. The cells are thus easier to lyse by sonication or French press using stable buffers and protease inhibitors [14, 62]. The common plant protein extraction methods involve protein precipitation using trichloroacetic acid:acetone and phenol:ammonium acetate-methanol. A relatively faster method involves the use of aqueous methanol:chloroform precipitation to concentrate proteins at the aqueous-organic interphase in under five min. This is also useful for protein extraction in *E. gracilis* as centrifugation can then be used to sediment any remaining paramylon to the bottom of the tube, while the protein precipitates at the interphase.

Another challenge in proteomic analysis is the identification of less abundant proteins, as their mass spectrometry (MS) signals can be masked by more abundant proteins. In *E. gracilis*, highly abundant proteins such as light-harvesting complexes [144], malate synthase-isocitrate lyase [145], tubulin [138] and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) [146] during PT cultivation, and phosphoenolpyruvate carboxykinase [147] and elongation factors [148] during HT and MT cultivations can mask lesser abundant proteins. This can partly be overcome by using more sensitive MS analysers, such as quadrupole Orbitrap mass

spectrometer (QExactive), with higher field strength and robustness, and double the resolution [149]. Such powerful MS analysers have been reported to easily detect about 5000 proteins in a standard 90 min gradient of mammalian cell lysate, digested with trypsin [149]. Another strategy for overcoming the masking is to remove the highly abundant proteins by chromatography or fractionation, during sample preparation, prior to tryptic digestion.

1.7.2 Common approaches used in proteomics studies

Proteomic analysis is traditionally carried out using comparative two-dimensional gel electrophoresis (2-DE), followed by analysis of differential protein spots using MS to identify up- and down-regulated proteins. This technique thus allows for a visual interpretation of protein profiling and has been used in numerous plant-based proteomic studies [150]. Another type of 2-DE is the difference gel electrophoresis (DIGE), in which samples are pre-labelled with fluorescent dyes prior to electrophoresis on 2-D gels [151]. The gels are then scanned at different wavelengths corresponding to the dye fluorescence and analysed for changes in volume ratios between spots [151]. This enables quantification of protein abundance through spot intensity determination. DIGE thus circumvents the usual 2-DE problems of variation between gels and difficulty of spot matching between gels [151]. A more popular approach involves shotgun proteomics, which has the ability to identify a larger number of proteins concurrently and from a complex mixture of proteins [152]. This high-throughput approach has been facilitated by a shift in MS instrumentation to enable analysis of more proteins simultaneously [152].

Label-free shotgun proteomics is a spectral counting-based approach, in which tandem mass spectrometry (MS/MS) is used to identify the number of MS/MS spectra of a given protein in a sample, to calculate its relative abundance to other proteins in the sample. There are several methods for quantitative proteomic analyses using spectral counting [153]. One

such method involves calculating the normalised spectral abundance factor (NSAF) values for each protein. This data analysis method accounts for MS/MS spectra, as well as protein length and variability between runs to normalise the relative abundance of proteins among samples [154]. This overcomes the problem with most other spectral counting methods, where larger proteins are expected to be over-represented through the generation of more peptides [154]. This is achieved by dividing the number of MS/MS spectra by the protein length to obtain a spectral abundance factor (SAF) value [154]. The variability between runs is further overcome by dividing the SAF of individual proteins by the sum of SAFs for all proteins in the sample complex to obtain the NSAF value [154]. The NSAF values are thus standardised across each sample preparation, and can directly be compared with other individual runs. This approach is also simpler and quicker to conduct, and reproducible.

Other quantitative shotgun proteomics approaches include stable isotope-based labelling, such as isotope-coded affinity tag (ICAT) reagents, isotope-differentiated binding energy shift tag (IDBEST), isobaric tags for absolute and relative quantification (iTRAQ), tandem mass tags (TMT), isobaric peptide termini labelling (IPTL) and stable isotope labelling by amino acids in cell culture (SILAC), and absolute quantification methods (AQUA) [155]. Such strategies eliminate the error arising from variations between runs using different samples, and allow direct assessment of peaks within the same MS/MS scan. Up to 54-plex multiplexing has been reported in the literature, but there appears to be a lack of chromatographic unity across such large-sized labelled samples, often requiring multiple reagent subsets with distinct isobaric masses [156]. These strategies are sometimes accompanied by inductively coupled plasma MS (ICP-MS), which allows absolute protein quantification through suitable elemental-based tags [155].

1.8 Proteomic analysis in microalgae

Apart from the comprehensive proteomics publication of *Euglena gracilis* described in Chapter 3 [143], there has been only one other proteomic publication of *E. gracilis,* involving proteomics of the mitochondria under aerobic and anaerobic conditions by Hoffmeister *et al.* (2004) [134]. This latter study was, however, performed using 2-DE. No proteins specific to mitochondria were identified during anaerobic condition [134]. It was also suggested that under aerobic condition, the *Euglena* mitochondria are always preparing for anaerobic function [134].

A rising number of proteomics studies has been carried out in other microalgae over the last decade. Proteomic response studies have been used to investigate the highly versatile metabolic capability of microalgae, including the effect of nitrogen starvation on central carbon metabolism, and lipid and starch accumulation, the effect of wastewater cultivation on nitrogen sensing and carbon fixation, the examination of specific biosynthetic pathways, the analysis of the proteome during micronutrient and phosphorus deficiency, the determination of proteomic changes at different time points in the circadian cycle, and the effect of salt, high light and cold stresses on metabolic pathways. Some of the proteomic response studies reported in other microalgae are discussed here, related to depletion of certain nutrients, the circadian cycle, and abiotic stress.

1.8.1 Proteomic response to nitrogen starvation and depletion of other minerals

Nitrogen starvation is known to increase lipid accumulation and reduce carbohydrate, protein and chlorophyll biosynthesis in microalgae [157]. Rai *et al.* (2017) used DIGE- and iTRAQ-based proteome profiling to compare between N-sufficient and N-starvation stages in *Chlorella* sp. FC2 IITG. They found that iTRAQ provided a larger coverage of the proteome compared to DIGE, and identified 200 differentially expressed proteins using the former [157].

Of these proteins, porphobilinogen deaminase (involved in chlorophyll biosynthesis), and sedoheptulose-1,7 bisphosphate and phosphoribulokinase (involved in carbon fixation) were down-regulated, while hydroxyacyl-ACP dehydrogenase and enoyl-ACP reductase (involved in lipid accumulation) were up-regulated [157]. Label-free shotgun proteomic analyses in *Chlorella vulgaris*, the genome of which was not sequenced at the time of publication, have also enabled identification of specific pathway components of the triacylglycerol biosynthetic pathway [158]. By using *de novo* assembled transcriptome as a database against which the proteomic data were searched, the up-regulated proteins from the fatty acid and triacylglycerol biosynthetic machinery were identified under N-starvation [158]. In another proteomic study of N-starvation in the cyanobacterium *Arthrospira* sp. PCC 8005, it was also determined that photosynthetic energy production and carbon fixation were down-regulated [159]. It was thus concluded that photosynthetic energy was redirected from protein synthesis to glycogen synthesis without affecting biomass productivity [159]. Comparative label-free and labelled shotgun proteomics has also been used to identify differentially expressed proteins related to light absorption, and lipid and carbohydrate assimilation in response to nitrogen source and availability in *Chlamydomonas reinhardtii* [160-163]. In another label-free shotgun proteomics study carried out under N-limited heterotrophic condition, the proteomic profiles of the microalga *Neochloris oleoabundans* were investigated [164]. Under preferential lipid and carbohydrate accumulation it was observed that the lipid biosynthetic pathway and central carbon metabolism were up-regulated during N-limitation [164].

Label-free shotgun proteomics has also been used to examine the effect of copper, iron, zinc and manganese micronutrient deficiencies on the proteome of *C. reinhardtii*, using the transcriptome as the search database [165]. About 200 differentially expressed proteins were identified in the study for each metal-deficient growth condition, of which iron- and manganese-deficient cells showed losses in photosynthesis-related function [165]. In another study, the effect of phosphorus deficiency revealed 132 differentially expressed proteins in

Thalassiosira weissflogii, especially those involved in macromolecular biosynthetic pathways [166]. Phosphorus uptake has also been investigated via proteomic profiling of *C. reinhardtii* grown in artificial wastewater, in which 92 proteins were found to be differentially expressed, including photosynthetic antennae proteins, carbon fixation enzymes, biosynthesis of plant hormones and enzymes related to the biosynthesis of secondary metabolites [167].

1.8.2 Proteomic changes during the circadian cycle

A study was carried out by Le Bihan *et al.* (2011) on the marine green alga *Ostreococcus tauri* to analyse its global proteome at different times during the circadian cycle [168]. This included growth under 24 h of illumination, 24 h of darkness and 12 h daylight/12 h darkness photoperiod [168]. The proteomic analyses enabled determination of several proteins that were differentially expressed during normal and aberrant photoperiods, and pathways that were regulated by the circadian cycle [168]. Transition from daylight to dark conditions also led to the up-regulation of several cell cycle-related proteins such as dynamin and kinesin and triggering of pathways such as starch synthesis [168].

1.8.3 Proteomic response to abiotic stress

Proteomic analyses in microalgae have also been used to determine the effect of various abiotic stresses on metabolism. The effect of salt concentration on growth of the halotolerant cyanobacterium *Euhalothece* sp. BAA001 was examined by Pandhal *et al.* (2008) using ^{15}N metabolic labelling [169]. The genome and transcriptome of the organism were not available at the time of publication, and 55 differentially expressed proteins were identified between low and moderate salt concentrations, 18 between moderate and high salt concentrations, and 35 between moderate and extremely high salt concentrations, using the NCBI non-redundant

database as the search database [169]. In another study, SILAC-based proteomic analyses were used to verify the regulatory properties of pathways during salt stress in *C. reinhardtii*, as metabolomic analyses could only be used to describe missing reactions and stoichiometric metabolic networks [170].

The effect of high light stress in the marine diatom *Thalassiosira pseudonana* has also been investigated using iTRAQ-based proteomic analyses and biochemical analyses [171]. A total of 143 differentially expressed proteins were identified under high light stress belonging to the light-harvesting complex proteins, ROS scavenging pathways, photorespiration, lipid metabolism and light protection mechanisms [171]. These were further verified using metabolite analyses of pigments, lipids and fatty acids [171]. In another high light stress study using 2-DE analysis, 105 differentially expressed proteins were identified between wild-type and very high light-resistant mutants of *C. reinhardtii* [172].

Temperature stress in microalgae has also been studied using proteomic profiling. The cold adaptation mechanisms in regard to central carbon metabolism, including glycolysis, gluconeogenesis, tricarboxylic acid cycle, oxidative phosphorylation, pentose-phosphate-pathway, and CO_2 fixation, have been investigated in *C. reinhardtii* using label-free shotgun proteomics [173]. This study also detected novel cold stress proteins in *C. reinhardtii* using proteomic analyses [173].

Overall, the studies above suggest that proteomic profiling is a useful high-throughput investigative tool in determining differential protein expression between growth conditions in microalgae, as well as identifying missing proteins from incomplete pathways, even when the genomic data is not available, with shotgun proteomics approaches providing better coverage of the proteome than 2-DE.

1.9 Objectives of this research

The main objective of the present study was to identify proteins expressed in *Euglena gracilis* under photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) growth conditions via proteomic profiling and investigate the regulation of metabolite biosynthesis and central carbon metabolic pathways under those conditions. An additional aim was to analyse these secondary metabolites, and identify proteins belonging to their biosynthetic pathways that have not been reported thus far, and assign a putative functional annotation to them.

The more specific aims were:

1. To analyse biomass and the metabolites paramylon, ascorbate, α-tocopherol and free amino acids in three strains of *E. gracilis* (*E. gracilis* var. *saccharophila*, *E. gracilis* Z and *E. gracilis* ZSB, an in-house developed streptomycin-bleached mutant strain) under PT, MT and HT growth conditions.
2. To determine the preliminary metabolic pathways of ascorbate, α-tocopherol, free amino acids, and central carbon and paramylon in *E. gracilis* under PT, MT and HT growth conditions.
3. To examine and compare how these pathways are differentially expressed under PT, MT and HT growth conditions, using label-free shotgun proteomics.
4. To provide putative functional annotations to unidentified proteins from the above pathways.
5. To investigate biomass accumulation, and the synthesis of proteins, free amino acids and the antioxidants ascorbate, α-tocopherol and β-carotene in a bioreactor cultivation of *E. gracilis* var. *saccharophila* under MT growth condition, to assess this strain as a potential host for supplement production.

This thesis has been organised into seven chapters including this introductory chapter. Chapter 2 details the materials and methods used in the experiments for this study that have not been described in the other chapters. Chapter 3 describes the ascorbate, α-tocopherol and free amino acids metabolic pathways of *E. gracilis* var. *saccharophila*. Chapter 4 describes the central carbon metabolic pathways of *E. gracilis* Z, comparing some of the enzymes to the *E. gracilis* var. *saccharophila* strain using data from the supplementary files of Chapter 3. Chapter 5 describes the paramylon biosynthesis and degradation pathways of *E. gracilis* var. *saccharophila, E. gracilis* Z and the streptomycin-bleached mutant strain *E. gracilis* ZSB. All three chapters analyse detailed proteomic profiles of the respective strains, in an effort to provide insight into proteomic responses of these pathways under PT, MT and HT growth conditions. The biochemical determination of each metabolite is also provided in the relevant chapter, along with the qRT-PCR analysis of paramylon synthase candidates in Chapter 5. Chapter 6 describes ascorbate, α-tocopherol and free amino acid analyses for *E. gracilis* Z and *E. gracilis* ZSB, which were not reported for these strains in Chapter 3. In addition, Chapter 6 provides biomass and metabolite analyses of paramylon, ascorbate, α-tocopherol, β-carotene, proteins and free amino acids in *E. gracilis* var. *saccharophila*, grown mixotrophically in a laboratory-scale bioreactor. Finally, Chapter 7 provides a conclusive summary and future directions of this research.

References

[1] Laurance WF, Sayer J and Cassman KG. 2014. Agricultural expansion and its impacts on tropical nature. *Trends Ecol Evol*. 29(2): 107-116.

[2] Davis DR. 2009. Declining fruit and vegetable nutrient composition: What is the evidence? *HortSci*. 44(1): 15-19.

[3] Cummings JH and Bingham SA. 1998. Clinical review. Diet and the prevention of cancer. *BMJ*. 317(7173): 1636-1640.

[4] Alissa EM and Ferns GA. 2017. Dietary fruits and vegetables and cardiovascular diseases risk. 2017. *Crit Rev Food Sci Nutr*. 57(9):1950-1962.

[5] Oh YS and Jun HS. 2014. Role of bioactive food components in diabetes prevention: Effects on beta-cell function and preservation. *Nutr Metab Insights*. 7: 51-59.

[6] Abdali D, Samson SE and Grover AK. 2015. How effective are antioxidant supplements in obesity and diabetes? *Med Princ Pract*. 24(3): 201-215.

[7] Kollmar M. 2015. Polyphyly of nuclear lamin genes indicates an early eukaryotic origin of the metazoan-type intermediate filament proteins. *Sci Rep*. 5: 10652.

[8] O'Neill EC, Trick M, Henrissat B and Field RA. 2015. *Euglena* in time: Evolution, control of central metabolic processes and multi-domain proteins in carbohydrate and natural product biochemistry. *Perspect Sci*. 6: 84-93.

[9] Buetow DE. Euglena. 2011. *Encyclopedia of Life Sciences*. John Wiley & Sons, Chichester. 1-5.

[10] Meyer MT, Genkov T, Skepper JN, Jouhet J, Mitchell MC, Spreitzer RJ and Griffiths H. 2012. Rubisco small-subunit α-helices control pyrenoid formation in *Chlamydomonas*. *Proc Natl Acad Sci USA*. 109(47): 19474-19479.

[11] Badger MR. 1987. The CO_2-concentrating mechanism in aquatic phototrophs. In Eds. Hatch MT and Boardman NK, *Photosynthesis*. Academic Press, Elsevier, MA, USA. 10: 219-

274.

[12] Kroth PG. 2015. The biodiversity of carbon assimilation. *J Plant Physiol*. 172: 76-81.

[13] Moroney JV and Ynalvez RA. 2007. Proposed carbon dioxide concentrating mechanism in *Chlamydomonas reinhardtii*. *Eukaryot Cell*. 6(8): 1251-1259.

[14] Rodríguez-Zavala JS, Ortiz-Cruz MA, Mendoza-Hernández G and Moreno-Sánchez R. 2010. Increased synthesis of α-tocopherol, paramylon and tyrosine by *Euglena gracilis* under conditions of high biomass production. *J Appl Microbiol*. 109(6): 2160-2172.

[15] US Food and Drug Administration. *GRN No. 513*. https://wayback.archive-it.org/7993/20171031004309/https://www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/NoticeInventory/ucm410031.htm (Date accessed: 27/05/2018).

[16] Nakano Y, Miyataka K, Yamaji R, Nishizawa A, Shigeoka S, Hosotani K, Inui H, Watanabe F, Enomoto T and Takenaka, S. 1995-1996. A protist, *Euglena gracilis* Z, functions as a sole nutrient source in a closed ecosystem. *Eco-Eng*. 8(1): 7-12.

[17] Azizullah A, Richter P and Häder DP. 2012. Sensitivity of various parameters in *Euglena gracilis* to short-term exposure to industrial wastewaters. *J Appl Phycol*. 24(2): 187-200.

[18] Hargreaves JW, Lloyd EJH and Whitton BA. 1975. Chemistry and vegetation of highly acidic streams. *Freshwater Biol*. 5(6): 563-576.

[19] Mahapatra DM, Chanakya HN and Ramachandra TV. 2013. *Euglena* sp. as a suitable source of lipids for potential use as biofuel and sustainable wastewater treatment. *J Appl Phycol*. 25(3): 855-65.

[20] Olaveson MM and Nalewajko C. 2000. Effects of acidity on the growth of two *Euglena* species. *Hydrobiologia*. 433(1-3): 39-56.

[21] Rodríguez-Zavala JS, García-García JD, Ortiz-Cruz MA and Moreno-Sánchez R. 2007. Molecular mechanisms of resistance to heavy metals in the protist *Euglena gracilis*. *J Environ Sci Health A Tox Hazard Subst Environ Eng*. 42(10): 1365-1378.

[22] Mendoza-Cózatl DG and Moreno-Sánchez R. 2005. Cd^{2+} transport and storage in the chloroplast of *Euglena gracilis*. *Biochim Biophys Acta*. 1706(1-2): 88-97.

[23] Mendoza-Cózatl DG, Devars S, Loza-Tavera H and Moreno-Sánchez R. 2002. Cadmium accumulation in the chloroplast of *Euglena gracilis*. *Physiol Plant*. 115(2): 276-283.

[24] Mendoza-Cózatl DG, Rodríguez-Zavala JS, Rodríguez-Enríquez S, Mendoza-Hernandez G, Briones-Gallardo R and Moreno-Sánchez R. 2006. Phytochelatin-cadmium-sulfide high-molecular-mass complexes of *Euglena gracilis*. *FEBS J*. 273(24): 5703-5713.

[25] Cervantes C, Espino-Saldaña AE, Acevedo-Aguilar F, León-Rodríguez IL, Rivera-Cano ME, Avila-Rodríguez M, Wróbel-Kaczmarczyk K, Wróbel-Zasada K, Gutiérrez-Corona JF, Rodríguez-Zavala JS and Moreno-Sánchez R. 2006. Interacciones microbianas con metales pesados. *Rev Latinoam Microbiol*. 48(2): 203-210.

[26] Castro-Guerrero NA, Rodríguez-Zavala JS, Marín-Hernández A, Rodríguez-Enríquez S and Moreno-Sánchez R. 2008. Enhanced alternative oxidase and antioxidant enzymes under Cd^{2+} stress in *Euglena*. *J Bioenerg Biomembr*. 40(3): 227-235.

[27] García-García JD, Rodríguez-Zavala JS, Jasso-Chávez R, Mendoza-Cozatl D and Moreno-Sánchez R. 2009. Chromium uptake, retention and reduction in photosynthetic *Euglena gracilis*. *Arch Microbiol*. 191(5): 431-440.

[28] Avilés C, Torres-Marquez ME, Mendoza-Cozatl D and Moreno-Sanchez R. 2005. Timecourse development of the Cd^{2+} hyper-accumulating phenotype in *Euglena gracilis*. *Arch Microbiol*. 184: 83-92.

[29] Avilés C, Loza-Tavera H, Terry N and Moreno-Sánchez R. 2003. Mercury pretreatment selects an enhanced cadmium-accumulating phenotype in *Euglena gracilis*. *Arch Microbiol*. 180(1): 1-10.

[30] Chae SR, Hwang EJ and Shin HS. 2006. Single cell protein production of *Euglena gracilis* and carbon dioxide fixation in an innovative photo-bioreactor. *Bioresour Technol*. 97(2): 322-329.

[31] Hayashi H, Narumi I, Wada S, Kikuchi M, Furuta M, Uehara K and Watanabe H. 2004. Light dependency of resistance to ionizing radiation in *Euglena gracilis*. *J Plant Physiol*. 161(10): 1101-1106.

[32] Beneragama CK and Goto K. 2010. Chlorophyll *a*: *b* ratio increases under low-light in 'shade-tolerant' *Euglena gracilis*. *Trop Agric Res*. 22(1): 12-25.

[33] Ahmed H and Häder DP. 2010. Rapid ecotoxicological bioassay of nickel and cadmium using motility and photosynthetic parameters of *Euglena gracilis*. *Environ Exp Bot*. 69(1): 68-75.

[34] Ahmed H and Häder DP. 2010. A fast algal bioassay for assessment of copper toxicity in water using *Euglena gracilis*. *J Appl Phycol*. 22(6): 785-792.

[35] Li M, Gao X, Wu B, Qian X, Giesy JP and Cui Y. 2014. Microalga *Euglena* as a bioindicator for testing genotoxic potentials of organic pollutants in Taihu Lake, China. *Ecotoxicology*. 23(4): 633-40.

[36] Grimm P, Risse JM, Cholewa D, Müller JM, Beshay U, Friehs K and Flaschel E. 2015. Applicability of *Euglena gracilis* for biorefineries demonstrated by the production of α-tocopherol and paramylon followed by anaerobic digestion. *J Biotechnol*. 215: 72-79.

[37] Anderson BB. 1964. Investigations into the *Euglena* method for the assay of the vitamin B_{12} in serum. *J Clin Pathol*. 17(1): 14-26.

[38] Häder DP. 1991. Phototaxis and gravitaxis in *Euglena gracilis*. In Eds. Lenci F, Ghetti F, Colombetti G, Häder DP and Song PS, *Biophys Photoreceptors Photomovements Microorg*. NATO ASI Series (Series A: Life Sciences). Springer, Boston, MA. 211: 203-221.

[39] Häder DP. 1987. Polarotaxis, gravitaxis and vertical phototaxis in the green flagellate, *Euglena gracilis*. *Arch Microbiol*. 147(2): 179-183.

[40] Brayner R, Coradin T, Beaunier P, Grenèche JM, Djediat C, Yéprémian C, Couté A and Fiéveta F. 2012. Intracellular biosynthesis of superparamagnetic 2-lines ferri-hydrite nanoparticles using *Euglena gracilis* microalgae. *Colloids Surf B Biointerfaces*. 93: 20-23.

[41] Shibakami M, Tsubouchi G, Nakamura M and Hayashi M. 2013. Polysaccharide nanofiber made from Euglenoid alga. *Carbohydr Polym*. 93(2): 499-505.

[42] Shibakami M, Tsubouchi G, Nakamura M and Hayashi M. 2013. Preparation of carboxylic acid-bearing polysaccharide nanofiber made from Euglenoid β-1,3-glucans. *Carbohydr Polym*. 98(1): 95-101.

[43] Shibakami M, Tsubouchi G, Sohma M and Hayashi M. 2016. Synthesis of nanofiber-formable carboxymethylated *Euglena*-derived β-1,3-glucan. *Carbohydr Polym*. 152: 468-478.

[44] Shibakami M, Tsubouchi G, Sohma M and Hayashi M. 2015. Preparation of transparent self-standing thin films made from acetylated Euglenoid β-1,3-glucans. *Carbohydr Polym*. 133: 421-428.

[45] Shibakami M, Tsubouchi G and Hayashi M. 2014. Thermoplasticization of Euglenoid β-1,3-glucans by mixed esterification. *Carbohydr Polym*. 105(1): 90-96.

[46] Lodén M, Ungerth L and Serup J. 2007. Changes in European legislation make it timely to introduce a transparent market surveillance system for cosmetics. *Acta Derm Venereol*. 87(6): 485-492.

[47] Ogawa T, Tamoi M, Kimura A, Mine A, Sakuyama H, Yoshida E, Maruta T, Suzuki K, Ishikawa T and Shigeoka S. 2015. Enhancement of photosynthetic capacity in *Euglena gracilis* by expression of cyanobacterial fructose-1,6-/sedoheptulose-1,7-bisphosphatase leads to increases in biomass and wax ester production. *Biotechnol Biofuels*. 8: 80.

[48] Doetsch NA, Favreau MR, Kuscuoglu N, Thompson MD and Hallick RB. 2001. Chloroplast transformation in *Euglena gracilis*: splicing of a group III twintron transcribed from a transgenic *psbK* operon. *Curr Genet*. 39(1): 49-60.

[49] Boehler RA and Danforth WF. 1968. Glucose utilization by *Euglena gracilis* var. *bacillaris*: Short-term metabolic studies. *J Protozool*. 15(1): 153-158.

[50] Hall RP and Schoenborn HW. 1939. The question of autotrophic nutrition in *Euglena gracilis*. *Physiol Zool*. 12(1): 76-84.

[51] Cook JR. 1967. Photo-Assimilation of acetate by an obligate phototrophic strain of *Euglena gracilis*. *J Protozool*. 14(3): 382-384.

[52] Barsanti L, Vismara R, Passarelli V and Gualtieri P. 2001. Paramylon (β-1,3-glucan) content in wild type and WZSL mutant of *Euglena gracilis*. Effects of growth conditions. *J Appl Phycol*. 13(1): 59-65.

[53] Mainx F. 1928. Beiträge zur Morphologie und Physiologie der *Euglenen*. II. Morphologische Beobachtung, Methode, und Erfolge der Reinkultur. II. Untersuchungen über die Ernährungs- und Reizphysiologie. *Arch für Protistenkd*. 60(2): 355-414.

[54] Wolken JJ. 1956. A molecular morphology of *Euglena gracilis* var. *bacillaris*. *J Protozool*. 3(4): 211-221.

[55] Cramer M and Myers J. 1952. Growth and photosynthetic characteristics of *Euglena gracilis*. *Arch Microbiol*. 17(1-4): 384-402.

[56] Pringsheim EG. 1955. Kleine Mitteilungen über Flagellaten und Algen. II. *Euglena gracilis* var. *saccharophila* n. var. und eine vereinfachte Nährlösung zur Vitamin B12-Bestimmung. *Arch für Mikrobiol*. 21: 414-419.

[57] Ogbonna JC, Tomiyama S and Tanaka H. 1998. Heterotrophic cultivation of *Euglena gracilis* Z for efficient production of α-tocopherol. *J Appl Phycol*. 10(1): 67-74.

[58] Schwarz T, Bartholmes P and Kaufmann M. 1995. Large-scale production of algal biomass for protein purification: Tryptophan synthase from *Euglena gracilis*. *Biotechnol Appl Biochem*. 22(2): 179-190.

[59] Radakovits R, Jinkerson RE, Darzins A and Posewitz MC. 2010. Genetic engineering of algae for enhanced biofuel production. *Eukaryot Cell*. 9(4): 486-501.

[60] Melis A. 2009. Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. *Plant Sci*. 177(4): 272-280.

[61] Ogbonna JC, Yada H, Masui H and Tanaka H. 1996. A novel internally illuminated stirred tank photobioreactor for large-scale cultivation of photosynthetic cells. *J Ferment*

Bioeng. 82(1): 61-67.

[62] Bäumer D, Preisfeld A and Ruppel HG. 2001. Isolation and characterization of paramylon synthase from *Euglena gracilis* (Euglenophyceae). *J Phycol*. 37(1): 38-46.

[63] Ogbonna JC, Tomiyama S and Tanaka H. 1999. Production of α-tocopherol by sequential heterotrophic-photoautotrophic cultivation of *Euglena gracilis*. *J Biotechnol*. 70(1-3): 213-221.

[64] Marchessault RH and Deslandes Y. 1979. Fine structure of (1→3)-β-D-glucans: curdlan and paramylon. *Carbohydr Res*. 75: 231-242.

[65] Kiss JZ, Vasconcelos AC and Triemer RE. 1988. The intramembranous particle profile of the paramylon membrane during paramylon synthesis in *Euglena* (Euglenophyceae). *J Phycol*. 24(2): 152-157.

[66] Briand J and Calvayrac R. 1980. Paramylon synthesis in heterotrophic and photoheterotrophic *Euglena* (Euglenophyceae). *J Phycol*. 16(2): 234-239.

[67] Marechal LR and Goldemberg SH. 1964. Uridine diphosphate glucose-β-1,3-glucan β-3-glucosyltransferase from *Euglena gracilis*. *J Biol Chem*. 239(10): 3163-3167.

[68] Gissibl A. 2018. *Enzymatic production of soluble bioactive β-1,3-glucans from paramylon*. Doctor of Philosophy Thesis, Department of Molecular Sciences, Macquarie University, New South Wales, Australia.

[69] Kataoka K, Muta T, Yamazaki S, and Takeshige K. 2002. Activation of macrophages by linear (1→3)-β-D-glucans. Implications for the recognition of fungi by innate immunity. *J Biol Chem*. 277(39): 36825-36831.

[70] Koizumi N, Sakagami H, Utsumi A, Fujinaga A, Takeda M, Asano K, Sugawara I, Ichikawa A, Kondo H, Mori S, Miyatake K, Nakano Y, Nakashima H, Murakami T, Miyano N and Yamamoto N. 1993. Anti-HIV (human immunodeficiency virus) activity of sulfated paramylon. *Antiviral Res*. 21(1): 1-14.

[71] Kondo Y, Kato A, Hojo H, Nozoe S, Takeuchi M and Ochi K. 1992. Cytokine-related

immunopotentiating activities of paramylon, a β-(1-->3)-D-glucan from *Euglena gracilis*. *J Pharmacobiodyn*. 15(11): 617-621.

[72] Sugiyama A, Suzuki K, Mitra S, Arashida R, Yoshida E, Nakano R, Yabuta Y and Takeuchi T. 2009. Hepatoprotective effects of paramylon, a β-1,3-D-glucan isolated from *Euglena gracilis* Z, on acute liver injury induced by carbon tetrachloride in rats. *J Vet Med Sci*. 71(7): 885-890.

[73] Sugiyama A, Hata S, Suzuki K, Yoshida E, Nakano R, Mitra S, Arashida R, Asayama Y, Yabuta Y and Takeuchi T. 2010. Oral administration of paramylon, a β-1,3-D-glucan isolated from *Euglena gracilis* Z inhibits development of atopic dermatitis-like skin lesions in NC/Nga mice. *J Vet Med Sci*. 72(6): 755-763.

[74] Xiao Z, Trincado CA and Murtaugh MP. 2004. β-glucan enhancement of T cell IFNγ response in swine. *Vet Immunol Immunopathol*. 102(3): 315-320.

[75] Wang L, Behr, SR, Newman RK and Newman CW. 1997. Comparative cholesterol-lowering effects of barley β-glucan and barley oil in golden Syrian hamsters. *Nutr Res*. 17(1): 77-88.

[76] Wood PJ. 1994. Evaluation of oat bran as a soluble fibre source. Characterization of oat β-glucan and its effects on glycaemic response. *Carbohydr Polym*. 25(4): 331-336.

[77] Yoshida Y, Tomiyama T, Maruta T, Tomita M, Ishikawa T and Arakawa K. 2016. De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics*. 17: 182.

[78] O'Neill EC, Trick M, Hill L, Rejzek M, Dusi RG, Hamilton CJ, Zimba PV, Henrissat B and Field RA. 2015. The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Mol BioSyst*. 11(10): 2808-2820.

[79] Tanaka Y, Ogawa T, Maruta T, Yoshida Y, Arakawa K and Ishikawa T. 2017. Glucan synthase-like 2 is indispensable for paramylon synthesis in *Euglena gracilis*. *FEBS Lett*.

591(10): 1360-1370.

[80] Ishiguro J, Saitou A, Durán A and Ribas JC. 1997. *cps1*$^{+}$, a *Schizosaccharomyces pombe* gene homolog of *Saccharomyces cerevisiae FKS* genes whose mutation confers hypersensitivity to cyclosporin A and papulacandin B. *J Bacteriol*. 179(24): 7653-7662.

[81] Roemer T and Bussey H. 1991. Yeast β-glucan synthesis: *KRE6* encodes a predicted type II membrane protein required for glucan synthesis *in vivo* and for glucan synthase activity *in vitro*. *Proc Natl Acad Sci USA*. 88(24): 11295-11299.

[82] Stasinopoulos SJ, Fisher PR, Stone BA and Stanisich VA. 1999. Detection of two loci involved in (1-->3)-β-glucan (curdlan) biosynthesis by *Agrobacterium* sp. ATCC31749, and comparative sequence analysis of the putative curdlan synthase gene. *Glycobiology*. 9(1): 31-41.

[83] Dwyer MR and Smillie RM. 1970. A light-induced β-1,3-glucan breakdown associated with the differentiation of chloroplasts in *Euglena gracilis*. *Biochim Biophys Acta*. 216(2): 392-401.

[84] Dwyer MR and Smillie RM. 1971. β-1,3-glucan: A source of carbon and energy for chloroplast development in *Euglena gracilis*. *Aust J Biol Sci*. 24(1): 15-22.

[85] Kitaoka M, Sasaki T and Tamiguchi H. 1993. Purification and properties of laminaribiose phosphorylase (EC 2.4.1.31) from *Euglena gracilis* Z. *Arch Biochem Biophys*. 304(2): 508-514.

[86] Goldemberg SH, Maréchal LR and De Souza BC. 1966. β-1,3-oligoglucan: Orthophosphate glucosyltransferase from *Euglena gracilis*. *Biochim Biophys Acta*. 146(2): 417-430.

[87] Takeda T, Nakano Y, Takahashi M, Konno N, Sakamoto Y, Arashida R, Marukawa Y, Yoshida E, Ishikawa T and Suzuki K. 2015. Identification and enzymatic characterization of an *endo*-1,3-β-glucanase from *Euglena gracilis*. *Phytochemistry*. 116: 21-27.

[88] Ray PD, Huang BW and Tsuji Y. 2012. Reactive oxygen species (ROS) homeostasis

and redox regulation in cellular signaling. *Cell Signal*. 24(5): 981-990.

[89] Smirnoff N. 2000. Ascorbic acid: Metabolism and functions of a multi-facetted molecule. *Curr Opin Plant Biol*. 3(3): 229-235.

[90] Ishikawa T and Shigeoka S. 2008. Recent advances in ascorbate biosynthesis and the physiological significance of ascorbate peroxidase in photosynthesizing organisms. *Biosci Biotechnol Biochem*. 72(5): 1143-1154.

[91] Zorov DB, Juhaszova M and Sollott SJ. 2014. Mitochondrial reactive oxygen species (ROS) and ROS-induced ROS release. *Physiol Rev*. 94(3): 909-950.

[92] Sharma P, Jha AB, Dubey RS and Pessarakli M. 2012. Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. *J Bot*. 2012: Article ID 217037.

[93] Görlach A, Dimova EY, Petry A, Martínez-Ruiz A, Hernansanz-Agustín P, Rolo AP, Palmeira CM and Kietzmann T. 2015. Reactive oxygen species, nutrition, hypoxia and diseases: Problems solved? *Redox Biol*. 6: 372-385.

[94] Csányi G and Miller Jr. FJ. 2014. Oxidative stress in cardiovascular disease. *Int J Mol Sci*. 15(4): 6002-6008.

[95] Valko M, Rhodes CJ, Moncol J, Izakovic M and Mazur M. 2006. Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem Biol Interact*. 160(1): 1-40.

[96] Nunomura A, Castellani RJ, Zhu X, Moreira PI, Perry G and Smith MA. 2006. Involvement of oxidative stress in Alzheimer disease. *J Neuropathol Exp Neurol*. 65(7): 631-641.

[97] Davey MW, van Montagu M, Inzé D, Sanmartin M, Kanellis A, Smirnoff N, Benzie IJJ, Strain JJ, Favell D and Fletcher J. 2000. Plant L-ascorbic acid: Chemistry, function, metabolism, bioavailability and effects of processing. *J Sci Food Agric*. 80(7): 825-860.

[98] Shigeoka S, Nakano Y and Kitaoka S. 1980. Occurrence of L-ascorbic acid in *Euglena gracilis* Z. *Bull Univ Osaka Pref Ser B*. 32: 43-48.

[99] Njoku PC, Ayuk AA and Okoye CV. 2011. Temperature effects on vitamin C content in citrus fruits. *Pakistan J Nutr*. 10(12): 1168-1169.

[100] Rekha C, Poornima G, Manasa M, Abhipsa V, Devi JP, Kumar HTV and Kekuda TRP. 2012. Ascorbic acid, total phenol content and antioxidant activity of fresh juices of four ripe and unripe citrus fruits. *Chem Sci Trans*. 1(2): 303-310.

[101] Foyer C, Rowell J and Walker D. 1983. Measurement of the ascorbate content of spinach leaf protoplasts and chloroplasts during illumination. *Planta*. 157(3): 239-244.

[102] Yuan Y V and Walsh NA. 2006. Antioxidant and antiproliferative activities of extracts from a variety of edible seaweeds. *Food Chem Toxicol*. 44(7): 1144-1150.

[103] Burtin P. 2003. Nutritional value of seaweeds. *Electron J Environ Agric Food Chem*. 2(4): 498-503.

[104] Bánhegyi G, Braun L, Csala M, Puskás F and Mandl J. 1997. Ascorbate metabolism and its regulation in animals. *Free Radic Biol Med*. 23(5): 793-803.

[105] Wheeler GL, Jones MA and Smirnoff N. 1998. The biosynthetic pathway of vitamin C in higher plants. *Nature*. 393(6683): 365-369.

[106] Ishikawa T, Nishikawa H, Gao Y, Sawa Y, Shibata H, Yabuta Y, Maruta T and Shigeoka S. 2008. The pathway via D-galacturonate/L-galactonate is significant for ascorbate biosynthesis in *Euglena gracilis*: Identification and functional characterization of aldonolactonase. *J Biol Chem*. 283(45): 31133-31141.

[107] Wheeler G, Ishikawa T, Pornsaksit V and Smirnoff N. 2015. Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *Elife*. 4: e06369.

[108] Ishikawa T, Tajima N, Nishikawa H, Gao Y, Rapolu M, Shibata H, Sawa Y and Shigeoka S. 2010. *Euglena gracilis* ascorbate peroxidase forms an intramolecular dimeric structure: Its unique molecular characterization. *Biochem J*. 426(2): 125-134.

[109] Tani Y and Tsumura H. 1989. Screening for tocopherol-producing microorganisms and

α-tocopherol production by *Euglena gracilis* Z. *Agric Biol Chem*. 53(2): 305-312.

[110] Taketomi H, Soda K and Katsui G. 1983. Results of screening test in tocopherols in microbial realm. *Vitamins (Japan)*. 57: 133-138.

[111] Ogbonna JC. 2009. Microbiological production of tocopherols: Current state and prospects. *Appl Microbiol Biotechnol*. 84(2): 217-225.

[112] Shigeoka S, Onishi T, Nakano Y and Kitaoka S. 1986. The contents and subcellular distribution of tocopherols in *Euglena gracilis*. *Agric Biol Chem*. 50(4): 1063-1065.

[113] Threlfall DR and Goodwin TW. 1967. Nature, intracellular distribution and formation of terpenoid quinones in *Euglena gracilis*. *Biochem J*. 103(2): 573-588.

[114] Kusmic C, Barsacchi R, Barsanti L, Gualtieri P and 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. *J Appl Phycol*. 10(6): 555-559.

[115] Carballo-Cárdenas EC, Tuan PM, Janssen M and Wijffels RH. 2003. Vitamin E (α-tocopherol) production by the marine microalgae *Dunaliella tertiolecta* and *Tetraselmis suecica* in batch cultivation. *Biomol Eng*. 20(4-6): 139-147.

[116] Takeyama H, Kanamaru A, Yoshino Y, Kakuta H, Kawamura Y and Matsunaga T. 1997. Production of antioxidant vitamins, beta-carotene, vitamin C, and vitamin E, by two-step culture of *Euglena gracilis* Z. *Biotechnol Bioeng*. 53(2): 185-190.

[117] Shigeoka S, Ishiko H, Nakano Y and Mitsunaga T. 1992. Isolation and properties of gamma-tocopherol methyltransferase in *Euglena gracilis*. *Biochim Biophys Acta*. 1128(2-3): 220-226.

[118] Qi Q, Hao M, Ng W, Slater SC, Baszis SR, Weiss JD and Valentin HE. 2005. Application of the *Synechococcus nirA* promoter to establish an inducible expression system for Eegineering the *Synechocystis* tocopherol pathway. *Appl Environ Microbiol*. 71(10): 5678-5684.

[119] Gross JA and Stroz RJ. 1975. The carotenoid hydrocarbons of *Euglena gracilis* and

derived mutants. *Plant Physiol*. 55(2): 175-177.

[120] Krinsky NI and Goldsmith TH. 1960. The carotenoids of the flagellated alga, *Euglena gracilis*. *Arch Biochem Biophys*. 91(2): 271-279.

[121] Watanabe F, Yoshimura K and Shigeoka S. 2017. Biochemistry and physiology of vitamins in *Euglena*. In Eds. Schwartzbach SD and Shigeoka S, *Euglena: Biochemistry, Cell and Molecular Biology*. *Adv Exp Med Biol.* Springer International Publishing, Cham. 979: 65-90.

[122] Britton G, Liaaen-Jensen S and Pfander H. 2008. Special Molecules, Special Properties. In Eds. Britton G, Liaaen-Jensen S and Pfander H, *Carotenoids Vol. 4: Natural Functions*. Birkhäuser Verlag, Basel. 4: 1-6.

[123] Hosotani K and Kitaoka S. 1984. Determination of provitamin A in *Euglena gracilis* Z by high performance liquid chromatography and changes of the contents under various culture conditions. *J Jpn Soc Nutr Food Sci*. 37(6): 519-524.

[124] Dolphin WD. 1970. Photoinduced carotenogenesis in chlorotic *Euglena gracilis*. *Plant Physiol*. 46(5): 685-691.

[125] Young VR. 1994. Adult amino acid requirements: the case for a major revision in current recommendations. *J Nutr*. 124(8 Suppl): 1517S-1523S.

[126] Imura K and Okada A. 1998. Amino acid metabolism in pediatric patients. *Nutrition*. 14(1): 143-148.

[127] Park BS, Hirotani A, Nakano Y and Kitaoka, S. 1983. The physiological role and catabolism of arginine in *Euglena gracilis*. *Agric Biol Chem*. 47: 2561-2567.

[128] Shigeoka S, Onishi T, Nakano Y and Kitaoka S. 1987. Characterization and physiological function of glutathione reductase in *Euglena gracilis* Z. *Biochem J*. 242(2): 511-515.

[129] Overbaugh JM and Fall R. 1985. Characterization of a selenium-independent glutathione peroxidase from *Euglena gracilis*. *Plant Physiol*. 77(2): 437-442.

[130] Shigeoka S, Yasumoto R, Onishi T, Nakano Y and Kitaoka S. 1987. Properties of monodehydroascorbate reductase and dehydroascorbate reductase and their participation in the regeneration of ascorbate in *Euglena gracilis*. *J Gen Microbiol*. 133: 227-232.

[131] Kanematsu S and Asada K. 1979. Ferric and manganic superoxide dismutases in *Euglena gracilis*. *Arch Biochem Biophys*. 195(2): 535-545.

[132] EMBL-EBI. 2018. *What is proteomics? Proteomics: An introduction to EMBL-EBI resources.* https://www.ebi.ac.uk/training/online/course/proteomics-introduction-ebi-resources/what-proteomics (Date accessed: 27/05/2018).

[133] Vesteg M, Vacula R, Burey S, Löffelhardt W, Drahovská H, Martin W and Krajcovic J. 2009. Expression of nucleus-encoded genes for chloroplast proteins in the flagellate *Euglena gracilis*. *J Eukaryot Microbiol*. 56(2):159-166.

[134] Hoffmeister M, van der Klei A, Rotte C, van Grinsven KWA, van Hellemond JJ, Henze K, Tielens AGM and Martin W. 2004. *Euglena gracilis* rhodoquinone: Ubiquinone ratio and mitochondrial proteome differ under aerobic and anaerobic conditions. *J Biol Chem*. 279(21): 22422-22429.

[135] Keller M, Chan RL, Tessier LH, Weil JH and Imbault P. 1991. Post-transcriptional regulation by light of the biosynthesis of *Euglena* ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit. *Plant Mol Biol*. 17(1): 73-82.

[136] Kishore R and Schwartzbach SD. 1992. Translational control of the synthesis of the *Euglena* light harvesting chlorophyll *a*/*b* binding protein of photosystem II. *Plant Sci*. 85(1): 79-89.

[137] Madhusudhan R, Ishikawa T, Sawa Y, Shigeoka S and Shibata H. 2003. Post-transcriptional regulation of ascorbate peroxidase during light adaptation of *Euglena gracilis*. *Plant Sci*. 165(1): 233-238.

[138] Levasseur PJ, Meng Q and Bouck GB. 1994. Tubulin genes in the algal protist *Euglena gracilis*. *J Eukaryot Microbiol*. 41(5): 468-477.

[139] Saint-Guily A, Schantz ML and Schantz R. 1994. Structure and expression of a cDNA encoding a histone H2A from *Euglena gracilis*. *Plant Mol Biol*. 24(6): 941-948.

[140] Vacula R, Steiner JM, Krajcovic J, Ebringer L and Löffelhardt W. 2001. Plastid state- and light-dependent regulation of the expression of nucleus-encoded genes for chloroplast proteins in the flagellate *Euglena gracilis*. *Folia Microbiol (Praha)*. 46(5): 433-441.

[141] Záhonová K, Füssy Z, Oborník M, Eliáš M and Yurchenko V. 2016. RuBisCO in non-photosynthetic alga *Euglena longa*: divergent features, transcriptomic analysis and regulation of complex formation. *PLoS One*. 11(7): e0158790.

[142] Ebenezer TE, Carrington M, Lebert M, Kelly S and Field MC. 2017. *Euglena gracilis* genome and transcriptome: Organelles, nuclear genome assembly strategies and initial features. In Eds. Schwartzbach SD and Shigeoka S, *Euglena: Biochemistry, Cell and Molecular Biology*. *Adv Exp Med Biol.* Springer International Publishing, Cham. 979: 125-140.

[143] Hasan MT, Sun A, Mirzaei M, Te'o J, Hobba G, Sunna A and Nevalainen H. 2017. A comprehensive assessment of the biosynthetic pathways of ascorbate, α-tocopherol and free amino acids in *Euglena gracilis* var. *saccharophila*. *Algal Res*. 27: 140-151.

[144] Koziol AG and Durnford DG. 2008. *Euglena* light-harvesting complexes are encoded by multifarious polyprotein mRNAs that evolve in concert. *Mol Biol Evol*. 25(1): 92-100.

[145] Nakazawa M, Minami T, Teramura K, Kumamoto S, Hanato, S, Takenaka S, Ueda M, Inui H, Nakano Y and Miyatake K. 2005. Molecular characterization of a bifunctional glyoxylate cycle enzyme, malate synthase/isocitrate lyase, in *Euglena gracilis*. *Comp Biochem Physiol Part B*. 141(4): 445-452.

[146] Brawerman G and Konigsberg N. 1960. On the formation of the TPN requiring glyceraldehyde-3-phosphate dehydrogenase during the production of chloroplasts in *Euglena gracilis*. *Biochim Biophys Acta*. 43: 374-381.

[147] Laval-Martin D, Farineau J, Pineau B and 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(2): 157-167.

[148] Montandon PE and Stutz E. 1990. Structure and expression of the *Euglena gracilis* nuclear gene coding for the translation elongation factor EF-1α. *Nucleic Acids Res*. 18(1): 75-82.

[149] Scheltema RA, Hauschild JP, Lange O, Hornburg D, Denisov E, Damoc E, Kuehn A, Makarov A and Mann M. 2014. The Q Exactive HF, a benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field orbitrap analyzer. *Mol Cell Proteomics*. 13(12): 3698-3708.

[150] Rabilloud T. 2014. How to use 2D gel electrophoresis in plant proteomics. *Methods Mol Biol*. 1072: 43-50.

[151] Feret R and Lilley KS. 2014. Protein profiling using two-dimensional difference gel electrophoresis (2-D DIGE). In Eds. Coligan JE, Dunn BM, Speicher DW and Wingfield PT, *Current Protocols in Protein Science*. John Wiley & Sons, Inc., Hoboken, NJ, USA. 75:22.2.1–22.2.17.

[152] Neilson KA, George IS, Emery SJ, Muralidharan S, Mirzaei M and Haynes PA. 2014. Analysis of rice proteins using SDS-PAGE shotgun proteomics. *Methods Mol Biol*. 1072: 289-302.

[153] Choi H, Fermin D and Nesvizhskii AI. 2008. Significance analysis of spectral count data in label-free shotgun proteomics. *Mol Cell Proteomics*. 7(12): 2373-2385.

[154] Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, Conaway JW, Florens L and Washburn MP. 2006. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci USA*. 103(50): 18928-18933.

[155] Chahrour O, Cobice D and Malone J. 2015. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *J Pharm Biomed Anal*. 113: 2-20.

[156] Braun CR, Bird GH, Wühr M, Erickson BK, Rad R, Walensky LD, Gygi SP and Haas W. 2015. Generation of multiple reporter ions from a single isobaric reagent increases multiplexing capacity for quantitative proteomics. *Anal Chem*. 87(19): 9855-9863.

[157] Rai V, Muthuraj M, Gandhi MN, Das D and Srivastava S. 2017. Real-time iTRAQ-based proteome profiling revealed the central metabolism involved in nitrogen starvation induced lipid accumulation in microalgae. *Sci Rep*. 7: 45732.

[158] Guarnieri MT, Nag A, Smolinski SL, Darzins A, Seibert M and Pienkos PT. 2011. Examination of triacylglycerol biosynthetic pathways via *de novo* transcriptomic and proteomic analyses in an unsequenced microalga. *PLoS One*. 6(10): e25851.

[159] Depraetere O, Deschoenmaeker F, Badri H, Monsieurs P, Foubert I, Leys N, Wattiez R and Muylært K. 2015. Trade-off between growth and carbohydrate accumulation in nutrient-limited *Arthrospira* sp. PCC 8005 studied by integrating transcriptomic and proteomic approaches. *PLoS One*. 10(7): e0132461.

[160] Patel AK, Huang EL, Low-Décarie E and Lefsrud MG. 2015. Comparative shotgun proteomic analysis of wastewater-cultured microalgae: Nitrogen sensing and carbon fixation for growth and nutrient removal in *Chlamydomonas reinhardtii*. *J Proteome Res*. 14(8): 3051-3067.

[161] Lee DY, Park JJ, Barupal DK and Fiehn O. 2012. System response of metabolic networks in *Chlamydomonas reinhardtii* to total available ammonium. *Mol Cell Proteomics*. 11(10): 973-988.

[162] Longworth J, Noirel J, Pandhal J, Wright PC and Vaidyanathan S. 2012. HILIC- and SCX-Based quantitative proteomics of *Chlamydomonas reinhardtii* during nitrogen starvation induced lipid and carbohydrate accumulation. *J Proteome Res*. 11(12): 5959-5971.

[163] Park JJ, Wang H, Gargouri M, Deshpande RR, Skepper JN, Holguin FO, Juergens MT, Shachar-Hill Y, Hicks LM and Gang DR. 2015. The response of *Chlamydomonas reinhardtii* to nitrogen deprivation: A systems biology analysis. *Plant J*. 81(4): 611-624.

[164] Morales-Sánchez D, Kyndt J, Ogden K and Martinez A. 2016. Toward an understanding of lipid and starch accumulation in microalgae: A proteomic study of *Neochloris oleoabundans* cultivated under N-limited heterotrophic conditions. *Algal Res*. 20: 22-34.

[165] Hsieh SI, Castruita M, Malasarn D, Urzica E, Erde J, Page MD, Yamasaki H, Casero D, Pellegrini M, Merchant SS and Loo JA. 2013. The Proteome of copper, iron, zinc, and manganese micronutrient deficiency in *Chlamydomonas reinhardtii*. *Mol Cell Proteomics*. 12(1): 65-86.

[166] Wang X, Huang B and Zhang H. 2014. Phosphorus deficiency affects multiple macromolecular biosynthesis pathways of *Thalassiosira weissflogii*. *Acta Oceanol Sin*. 33(4): 85-91.

[167] Patel AK. 2015. *Microalgae for wastewater treatment and biomass production: A comparative analysis of growth and nutrient removal including shotgun proteomics*. Doctor of Philosophy Thesis, Department of Bioresource Engineering, McGill University, Quebec, Canada.

[168] Le Bihan T, Martin SF, Chirnside ES, van Ooijen G, Barrios-Llerena ME, O'Neill JS, Shliaha PV, Kerr LE and Millar AJ. 2011. Shotgun proteomic analysis of the unicellular alga *Ostreococcus tauri*. *J Proteomics*. 74(10): 2060-2070.

[169] Pandhal J, Snijders AP, Wright PC and Biggs CA. 2008. A cross-species quantitative proteomic study of salt adaptation in a halotolerant environmental isolate using ^{15}N metabolic labelling. *Proteomics*. 8(11): 2266-2284.

[170] Mastrobuoni G, Irgang S, Pietzke M, Assmus HE, Wenzel M, Schulze WX and Kempa S. 2012. Proteome dynamics and early salt stress response of the photosynthetic organism *Chlamydomonas reinhardtii*. *BMC Genomics*. 13: 215.

[171] Dong HP, Dong YL, Cui L, Balamurugan S, Gao J, Lu SH and Jiang T. 2016. High light stress triggers distinct proteomic responses in the marine diatom *Thalassiosira pseudonana*. *BMC Genomics*. 17: 994.

[172] Förster B, Mathesius U and Pogson BJ. 2006. Comparative proteomics of high light stress in the model alga *Chlamydomonas reinhardtii*. *Proteomics*. 6(15): 4309-4320.

[173] Valledor L, Furuhashi T, Hanak AM and Weckwerth W. 2013. Systemic cold stress adaptation of *Chlamydomonas reinhardtii*. *Mol Cell Proteomics*. 12(8): 2032-2047.

CHAPTER 2

Materials and methods

2.1 General laboratory practices

All chemicals were purchased from Sigma-Aldrich (Australia) unless stated otherwise. Media, buffers and reagents were prepared using Millipore Milli-Q filtered water. All glassware, microcentrifuge tubes, micropipette tips and boxes, pipette tips, MilliQ water, solutions, reagents and buffers were autoclaved at 121°C at 15 psi for 20 min. The media were autoclaved under the same conditions, but for 30 min. The heat-sensitive solutions, reagents or buffers were sterilised by membrane filtration through Minisart® filters of 0.2 µm pore size.

2.2 *Euglena gracilis* strains

Three strains of *Euglena gracilis* were used for this project. *E. gracilis* Z was obtained from Southern Biological Pty Ltd, Australia and *E. gracilis* var. *saccharophila* (UTEX 752) was obtained from the University of Texas Culture Collection, USA. The streptomycin-bleached mutant, *E. gracilis* ZSB, was developed in-house by bleaching *E. gracilis* Z with streptomycin as described in Chapter 5. *E. gracilis* Z and *E. gracilis* var. *saccharophila* are mixotrophic strains, capable of photosynthesis as well as heterotrophic nutrient assimilation. However, the strains differ in their degree of preference for organic carbon sources, with *E. gracilis* var. *saccharophila* preferring heterotrophic nutrition more. On the other hand, the bleached mutant *E. gracilis* ZSB lacks mature chloroplasts and cannot perform photosynthesis.

2.3 Media and culture conditions

2.3.1 Growth and maintenance media

All *Euglena gracilis* stocks were maintained on solid *Euglena* maintenance (EM) medium. The medium was modified from the University of Texas Culture Collection of Algae

website (https://utex.org/products/euglena-medium). The composition of the modified EM medium is indicated in Chapter 5. The EM medium was only used for long-term maintenance of axenic stock cultures of the strains, and during development of the streptomycin-bleached mutant, *E. gracilis* ZSB, as described in Chapter 5.

The basal growth medium (NY medium, containing ammonium chloride and yeast extract) was prepared as previously described [1], with some modifications. The composition of the NY medium is shown in Table 2.1, and the components of the trace mineral stock solutions A and B are shown in Table 2.2. The pH of the medium was adjusted to 3.5 with concentrated HCl, prior to autoclaving and cooling the medium to room temperature (RT), and adding vitamin supplements (Table 2.3). Where mentioned, the NY medium was supplemented with glucose to form GNY (containing glucose, ammonium chloride and yeast extract) medium (Table 2.1). The same amounts of the NY medium components were used to prepare 925 ml of medium, which was topped up with 75 ml of 20% (w/v) glucose stock solution. Glucose stock solution (20% w/v) was prepared by dissolving 20 g of glucose in 100 ml of MilliQ water and autoclaved separately from the NY medium.

2.3.2 Culture conditions

Cultivation was carried out under three different growth conditions: photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT), and is described in detail in Chapter 3. For the cultures grown in the laboratory-scale bioreactor (Chapter 6), the same medium formulations and growth conditions were used.

Table 2.1: Composition of growth media (NY and GNY).

Components	**Amount (per litre)**
NY medium:	
$CaCO_3$	0.2 g
$MgSO_4$	0.5 g
$(NH_4)_2HPO_4$	0.4 g
KH_2PO_4	0.2 g
NH_4Cl	1.81 g
yeast extract	10 g
trace mineral stock solution A	2 ml
trace mineral stock solution B	1 ml
<u>After autoclaving and cooling:</u>	
filter-sterilised vitamin B_1	2 ml
filter-sterilised vitamin B_{12}	50 µl
GNY medium:	
autoclaved NY medium*	925 ml
20% (v/v) glucose autoclaved separately	75 ml

**components of the NY medium were dissolved in 925 ml of MilliQ water*

Table 2.2: Components of the trace mineral stock solutions A and B required for the preparation of the NY medium.

Components	**Amount (per 100 ml)**
Trace mineral stock solution A*:	
$ZnSO_4 \cdot 7H_2O$	4.4 g
$MnSO_4 \cdot H_2O$	3.04 g
$Na_2MoO_4 \cdot 2H_2O$	1 g
$CoCl_2 \cdot 6H_2O$	0.08 g
Trace mineral stock solution B:	
$CuSO_4 \cdot 5H_2O$	0.078 g
H_3BO_3	0.057 g

**components were dissolved in several drops of HCl, prior to making up the volume with MilliQ water; both trace mineral stock solutions were stored at RT in the dark*

Table 2.3: Vitamin B_1 and B_{12} stock solutions required for the NY and GNY media.

Components	**Amount (per 10 ml)**
Vitamin B_1*: thiamine hydrochloride	50 mg
Vitamin B_{12}*: cyanocobalamin	10 mg

**components were dissolved in MilliQ water and filter-sterilised, prior to adding to the autoclaved and cooled NY medium; both vitamin stock solutions were stored at -20°C for up to 6 months*

2.4 Determination of dry mass and metabolites

The primary and secondary metabolites measured in this study included paramylon, ascorbate, α-tocopherol, β-carotene, proteins and free amino acids. The methods for determination of these metabolites along with the dry mass and unused glucose in the medium are described within the chapters as mentioned in Table 2.4 (Section 2.6).

2.4.1 Extraction of α-tocopherol and β-carotene

Alpha tocopherol and β-carotene extraction was carried out in a single extraction of lipid-soluble vitamins using an α-tocopherol extraction procedure described previously [1]. An aliquot of five ml of cells was harvested at 5000 *g* for five min at RT, and the pellet was resuspended in 500 µl of N_2-saturated MilliQ water. Fifty µl of 0.1 mg ml^{-1} butylated hydroxytoluene (BHT) in N_2-saturated methanol was added to the suspension, which was then mixed thoroughly to prevent oxidation of α-tocopherol. Four ml of ice-cold N_2-saturated methanol stored at -20°C was added to the suspension, followed by vortexing for one min. Three ml of ice-cold petroleum ether (PET) stored at -20°C was then added to the mixture, followed by vortexing for one min. The mixture was centrifuged at 500 *g* for five min, followed by the transfer of the upper phase to a fresh tube. A further two ml of PET was added to the

lower phase remaining in the old tube, followed by vortexing for one min. The mixture was again centrifuged at 500 *g* for five min and the upper phase pooled together with that from the first extraction. The pooled upper phases were mixed with two ml of 95% (v/v) N_2-saturated methanol by gentle vortexing for 30 sec. This mixture was again centrifuged at 500 *g* for five min. The upper phase was transferred to a fresh tube and dried down under a stream of N_2 gas. The resulting lipid residue was resuspended in one ml of 1:1 (v/v) mixture of methanol:ethanol and filtered through a 0.45 μm pore-diameter filter (Merck Millipore, Australia). Alpha tocopherol was then measured using a reverse phase high-performance liquid chromatography (RP-HPLC) as described in Chapter 3. The β-carotene (Chapter 6) was measured simultaneously on the RP-HPLC instrument by reading the UV absorbance at 455 nm.

The identity of peaks was confirmed by two separate methods: spiking the sample with respective standards to identify the peak, and collecting the peak fractions and running them through tandem mass spectrometry (MS/MS).

2.5 Proteomic analysis (label-free shotgun proteomics)

A schematic diagram of the label-free shotgun proteomics workflow is provided in Figure 2.1, which included the cultivation of samples, extraction, fractionation and in-gel digestion of proteins, analysis of peptides by nanoflow liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS), peptide identification, and functional annotation.

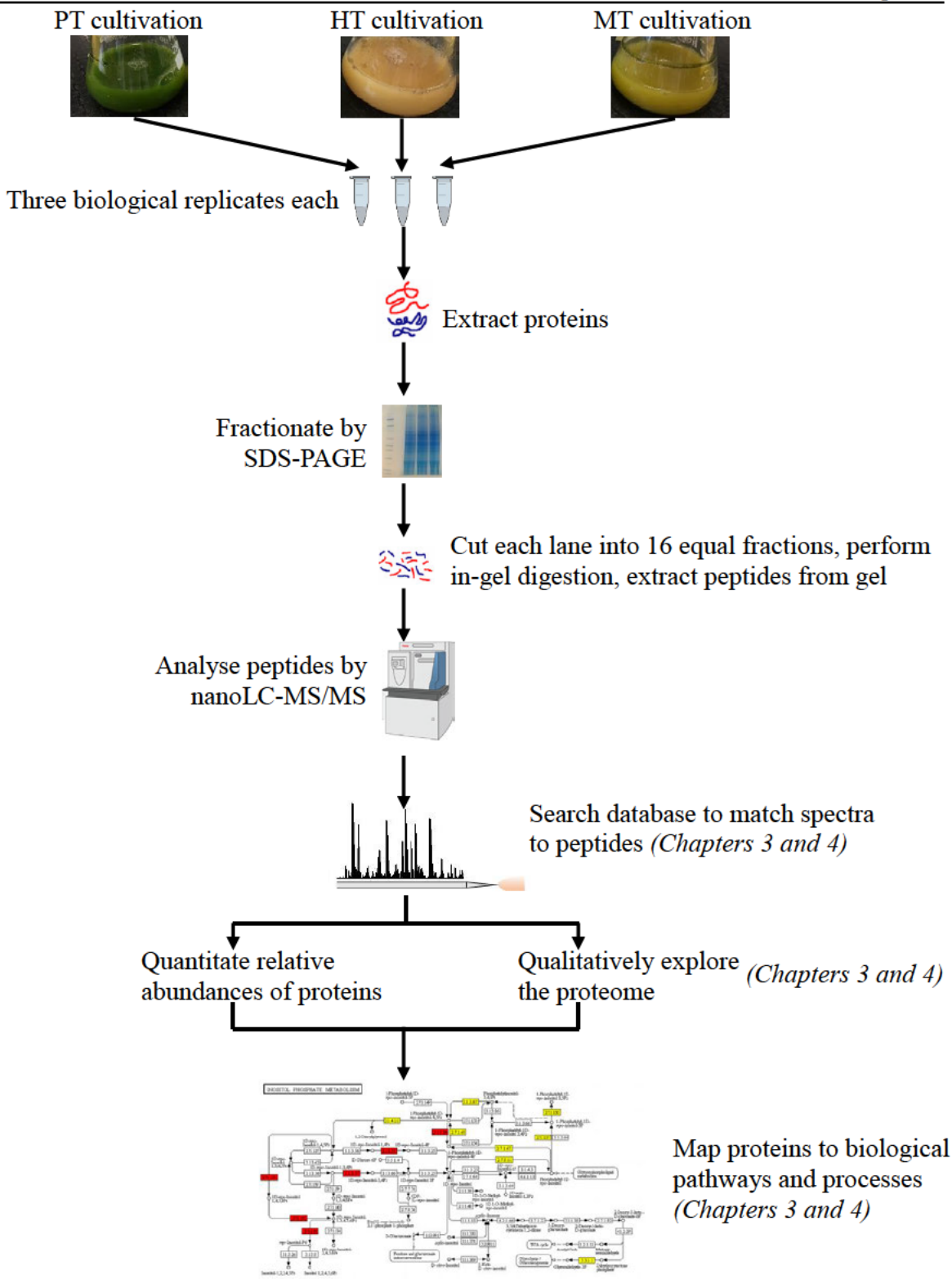

Figure 2.1: Schematic diagram of the label-free shotgun proteomics workflow including cultivation of cells, extraction and digestion of proteins, nanoLC-MS/MS analysis of peptides, identification of peptides, qualitative and quantitative analysis, and pathway mapping. General methodologies are provided in respective sections of this chapter and elaborated in the chapters indicated in italics.

2.5.1 Protein extraction and precipitation

Proteins were extracted together with ascorbate and free amino acids as described in Chapter 3. The lysed cells were centrifuged (Chapter 3) and the proteins in the supernatant were precipitated by aqueous methanol/chloroform precipitation method. Cold water, methanol and chloroform were added to the supernatant in the ratio of 3:4:1 (volume ratio). The mixture was vortexed vigorously for 10 sec until the solution was uniformly cloudy. The solution was centrifuged at 8,000 *g* for two min at 4°C. The proteins precipitated at the interface between the aqueous upper phase and the organic lower phase. The aqueous phase was discarded without disturbing the protein pellet at the interface and four volumes of methanol were added. The mixture was vortexed vigorously for 10 sec to wash the pellet, and then centrifuged at 8,000 *g* for two min. The white protein pellet was precipitated at the bottom of the tube. The methanol:chloroform solution was carefully removed and discarded. The protein pellet was air-dried for 5 to 10 min in a fume hood until the remaining methanol:chloroform evaporated, without over-drying the pellet. The dried pellet was stored at -20°C until required.

2.5.2 Protein quantification by bicinchoninic acid (BCA) assay

The air-dried pellet (Section 2.5.1) was re-dissolved in up to 500 µl of 5% (w/v) SDS by shaking for an hour at RT and the protein concentration was measured by using the BCA assay kit as per the manufacturer's protocol (Thermo Scientific Pierce™, Australia). Briefly, bovine serum albumin (BSA) standards were prepared in 5% (v/v) SDS in the range of 0 to 2 mg ml^{-1}. Three technical replicates of 25 µl each were pipetted into wells of a Greiner CELLSTAR® 96-well flat-bottomed polystyrene plate for the BSA standards and the unknown protein samples. To each well, 200 µl of the BCA working reagent was added and the plate was covered and shaken on a micro-plate shaker for 30 sec. The plate was incubated at 37°C

for 30 min to allow the colour to develop. The plate was cooled to RT and the absorbance was measured at 562 nm in a BMG FLUOstar Galaxy multi-functional plate reader. The BSA standards were used to plot a standard curve against which the protein concentrations of the unknown samples were determined, and the average of the technical replicates of each biological replicate was calculated.

2.5.3 Protein fractionation by SDS-PAGE and gel staining

For label-free shotgun proteomics analysis, 100 µg of proteins were fractionated per sample in the reduced state by SDS-PAGE. The volume of protein mixture in 5% (v/v) SDS (Section 2.5.2) containing 100 µg protein, was mixed with 7.5 µl of NuPAGE® 4x LDS Sample Buffer (Invitrogen) and three µl of NuPAGE® 10x Sample Reducing Agent (Invitrogen) and topped up to 30 µl with MilliQ water. The samples were heated at 70°C for 10 min and cooled to RT before fractionating in precast 4-12% gradient NuPAGE® Bis-Tris Mini Gels at 200 V for about 50 min in 1x 3-N-morpholinopropansulphonic acid (MOPS) running buffer. Each biological replicate was loaded onto a different lane and a lane of Novex® Sharp Pre-stained Protein Standard was run parallel to the samples.

The gels were stained with Coomassie Brilliant Blue for visualisation. They were first immersed in fixing solution containing 7% (v/v) acetic acid and 10% (v/v) methanol and incubated at RT with gentle shaking for 30 min. The fixing solution was then discarded and replaced with Coomassie staining solution containing 0.14% (w/v) of Coomassie Brilliant Blue G-250, 19.85% (v/v) methanol, 1.98% (v/v) phosphoric acid and 10% (w/v) ammonium sulphate. The gels were incubated for two hours at RT with gentle shaking. The Coomassie staining solution was then discarded and the gels were destained with 1% (v/v) acetic acid for 24 h at RT with gentle shaking. Kimwipes (Kimtech® Science, Australia) were immersed in the destaining solution to absorb the excess stain.

2.5.4 In-gel tryptic digestion

In-gel tryptic digestion was carried out as previously described [2] with slight modifications in incubation times. The gels containing stained protein bands were washed twice with MilliQ water, with 10 min incubation for each wash. Each lane, except the protein ladder, was excised from the gel with a scalpel and cut into 16 equally sized pieces. Each of these pieces was further chopped finely and placed in a well of a Greiner CELLSTAR® 96-well flat-bottomed polystyrene plate. The gel pieces were covered at all times except when required in order to minimise contamination from external proteins such as keratin.

To equilibrate the pH of the gel pieces, 200 µl of 100 mM NH_4HCO_3 was added to each well, and the gel pieces were incubated for 10 min at RT. The NH_4HCO_3 was discarded and the gel pieces were destained with 200 µl of wash buffer containing 50 mM NH_4HCO_3 and 50% (v/v) acetonitrile (ACN) for 20 min at RT on a shaker. This destaining step was repeated twice, followed by dehydration of the gel pieces with 200 µl of 100% (v/v) ACN for 10 min at RT on a shaker. The ACN was discarded and the gel pieces were air-dried in a fume hood for about 15 min, so that they were white and shrunken.

The cystine disulphide bonds of proteins in the dehydrated gel pieces were reduced with 50 µl (per well) of 10 mM dithiothreitol (DTT) in 50 mM NH_4HCO_3 solution by shaking briefly and then incubating at 37°C for 60 min. Any residual DTT was discarded and the gel pieces were cooled to RT. The reduced proteins were further alkylated with 50 µl (per well) of 55 mM iodoacetamide (IAA) in 50 mM NH_4HCO_3 solution by shaking briefly and then incubating at RT for 45 min in the dark. The residual IAA was discarded and the pH of the gel pieces was equilibrated as before, followed by destaining, dehydrating and air-drying.

The gel pieces were then incubated at 4°C for 15 min to equilibrate the temperature, before adding trypsin. Trypsin solution was prepared by adding 20 µl of 50 mM acetic acid per 20 µg of Sequencing Grade Modified Trypsin (Promega, Australia) to bring the trypsin into

solution. A 12.5 ng μl^{-1} trypsin working solution was then prepared by adding 1580 µl of 50 mM NH_4HCO_3. Twenty µl of the trypsin working solution was added per well of the cooled gel pieces, which were then incubated at 4°C for 20 min. Digestion was then carried out at 37°C overnight.

2.5.5 Peptide extraction and purification

Peptide extraction from the gel pieces was carried out as described before [2] with slight modifications in working volumes and incubation times. Fifty µl of a solution of 50% (v/v) ACN and 2% (v/v) formic acid (FA) was added to each well containing the digested peptides. The gel pieces were incubated in the ACN/FA solution at RT for 30 min with gentle shaking. The solution per well was transferred to a fresh 0.65 ml tube. The extraction was carried out once more, pooling the solution per well with the solution from the first ACN/FA extraction. The combined extracted peptide solution was just below 100 µl from each well. The extracted peptide solution was centrifuged at 2000 *g* for one min before drying the peptide extracts in a vacuum centrifuge. The dried peptide extracts were reconstituted in 20 µl of 2% (v/v) FA and purified using ZipTip® Pipette Tips with C18 resin (Merck Millipore, US).

The ZipTips were activated by passing 20 µl of a solution of 90% (v/v) ACN and 0.1% (v/v) FA through the tips three times with short spinning. Twenty µl of 0.1% (v/v) FA was then passed through the tips three times with short spinning. The peptide extract was passed through the tip, with short spinning. The trapped peptides on the activated filter were washed by passing 20 µl of 0.1% (v/v) FA through the tip with short spinning. The peptide flow-through and the wash-step flow through were collected, combined and again passed through the tip with short spinning. The peptides were then eluted into a fresh collection tube with 20 µl of a solution of 90% (v/v) ACN and 0.1% (v/v) FA with short spinning. The elution was repeated twice, so that the final volume of the eluate was about 40 µl. The purified peptide extracts were then dried

using a vacuum centrifuge and reconstituted again in 20 µl of 2% (v/v) FA. The reconstituted peptides were centrifuged at 18,000 *g* for 15 min. The top 10 µl of the reconstituted peptides were used for nanoflow liquid chromatography-MS/MS (nanoLC-MS/MS).

2.5.6 NanoLC-MS/MS

Each peptide extract was analysed by reversed-phase nanoLC-MS/MS using a QExactive mass spectrometer (Thermo scientific, CA, USA), as described previously [3]. Samples from each fraction were separated over a 60 min gradient using an Easy Nano LC 1000 (Thermo scientific, CA, USA). Briefly, 10 µl of sample was injected onto a peptide trap column (3.5 cm x 100 µm) packed in-house with Halo C18 silica (Advanced Materials Tech) of particle size 2.7 µm. The peptide samples in the column were desalted with 20 µl of Buffer A comprising 0.1% (v/v) FA. The peptide trap was then switched on line with an in-house packed reversed-phase column (10 cm x 75 µm) with Halo C18 silica (Advanced Materials Tech) of particle size 2.7 µm. Peptides were eluted from the column using a linear solvent gradient with steps from 1 to 50% of Buffer B for 50 min, 50 to 85% of Buffer B for two min, and then holding at 85% for eight min with a flow rate of 300 nl min^{-1} across the gradient. Buffer B consisted of 99.9% (v/v) ACN, 0.1% (v/v) FA. A nanospray ionisation source was used to direct the column eluate into the mass spectrometer. A 2.0 kV electrospray voltage was applied via a liquid junction upstream of the C18 column. The spectra were scanned over the range 350 to 2000 amu, and automated peak recognition, dynamic exclusion and MS/MS of the top ten most intense precursor ions at 30% normalisation collision energy were performed using the Xcalibur software (version 2.06) (Thermo scientific, USA).

2.5.7 Protein/peptide identification (global proteome machine parameters)

Protein and peptide identification was carried out using the global proteome machine (GPM) and is detailed in Chapter 3. The GPM search parameters were set as previously described [2], and included a fragment mass error of 0.1 Da, a tolerance of up to three missed tryptic cleavages and K/R-P cleavages, complete modification as carbamidomethylation of cysteine, and potential modifications as oxidation of methionine and tryptophan. The 16 fractions of each biological replicate were processed consecutively on the GPM with output files for each fraction, as well as a simple merged non-redundant output file for peptide and protein identifications with log(e) values less than -1.

2.5.8 Data processing and quality control

Data processing and quality control is described in Chapter 3. Peptide and protein false discovery rates (FDRs) were calculated as follows: peptide FDR was 2 x (reversed peptide identifications)/(total peptide identifications) x 100; protein FDR was (reversed protein identifications/total protein identifications) x 100.

A spectral fraction of 0.5 was added to all spectral counts to compensate for null values and enable log transformation for the statistical analysis. The normalised spectral abundance factor (NSAF) value was calculated for each protein as follows [4]:

$$(NSAF)_k = \frac{(SpC/L)_k}{\sum_{i=1}^{N}(SpC/L)_k}$$

where SpC is the total number of MS/MS spectra matching peptides from protein k, L is the length of protein k, and N indicates all the proteins present in the experiment. Quantitative proteomic analysis and statistical analysis were carried out using summed NSAF values as a measure of relative protein abundance [5].

2.6 Other materials and methods used in this project

The general methods used for the work described here are shown in Table 2.4. Detailed description has been outlined in the corresponding chapters.

Table 2.4: List of methods as described in the respective chapters of this thesis.

Method	**Chapter**
Culture conditions of *E. gracilis* strains	3 and 4
Development of the streptomycin-bleached mutant strain, *E. gracilis* ZSB	5
Determination of dry mass	3 and 4
Determination of ascorbate by colorimetric assay	3
Determination of α-tocopherol by reverse phase high-performance liquid chromatography (RP-HPLC)	3
Determination of free amino acids by reverse phase ultra-performance liquid chromatography (RP-UPLC)	3
Label-free shotgun proteomics: protein/peptide identification, statistical analysis, functional annotation	3 and 4
Determination of paramylon by phenol-sulphuric acid assay	5
Determination of unused glucose in the medium by phenol-sulphuric acid assay	5
Real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR)	5
Cultivation in a laboratory-scale bioreactor	6
Light microscopy	6

References

[1] Rodríguez-Zavala JS, Ortiz-Cruz MA, Mendoza-Hernández G and Moreno-Sánchez R. 2010. Increased synthesis of α-tocopherol, paramylon and tyrosine by *Euglena gracilis* under conditions of high biomass production. *J Appl Microbiol*. 109(6): 2160-2172.

[2] Wu Y, Mirzaei M, Pascovici D, Chick JM, Atwell BJ and Haynes PA. 2016. Quantitative proteomic analysis of two different rice varieties reveals that drought tolerance is correlated with reduced abundance of photosynthetic machinery and increased abundance of ClpD1 protease. *J Proteomics*. 143: 73-82.

[3] Thompson EL, O'Connor W, Parker L, Ross P and Raftos DA. 2015. Differential proteomic responses of selectively bred and wild-type Sydney rock oyster populations exposed to elevated CO_2. *Mol Ecol*. 24(6): 1248-1262.

[4] Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, Conaway JW, Florens L and Washburn MP. 2006. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci USA*. 103(50): 18928-18933.

[5] Neilson KA, George IS, Emery SJ, Muralidharan S, Mirzaei M and Haynes PA. 2014. Analysis of rice proteins using SDS-PAGE shotgun proteomics. *Methods Mol Biol*. 1072: 289-302.

CHAPTER 3

A comprehensive assessment of the biosynthetic pathways of ascorbate, α-tocopherol and free amino acids in *Euglena gracilis* var. *saccharophila*

3.1 Introduction

Euglena gracilis var. *saccharophila* is a sugar-loving variant that grows well in the presence of glucose. While much has been reported about metabolites produced in the model strain *E. gracilis* Z, the ability of the *E. gracilis* var. *saccharophila* strain to produce metabolites has not been investigated. In this chapter, proteomic profiling of *E. gracilis* var. *saccharophila* is described along with the identification of enzymes involved in the biosynthesis of the metabolites ascorbate, α-tocopherol and free amino acids. A comparison among the proteomic expression profiles during photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivations is also reported. Differential expression of isozymes of several enzyme classes was observed among the cultivation conditions, including those involved in the biosynthesis of aromatic amino acids, serine and cysteine among others. Putative functional annotations for many of the unidentified proteins have also been provided.

The results of this research were reported in a peer-reviewed paper that was published in *Algal Research* in 2017.

3.2 Contribution to publication

The concept for this publication was developed in partnership with my supervisors Helena Nevalainen, Anwar Sunna and Angela Sun, my adjunct supervisor Junior Te'o (Queensland University of Technology), and the industrial partner representative for this project Graham Hobba (Agritechnology Pty Ltd). The experiments were designed by me with supervision from Helena Nevalainen and Angela Sun. I performed all experiments, with help from Angela Sun for α-tocopherol extraction and measurements. Data analysis was carried out by me with the assistance of Mehdi Mirzaei, along with my supervisors. The manuscript was prepared by me and reviewed by Angela Sun, Junior Te'o, Graham Hobba, Anwar Sunna and Helena Nevalainen.

Table 3.1: Author contribution summary for publication.

	MTH	AS[1]	MM	JT	GH	AS[2]	HN
Conception	•	•		•	•	•	•
Experimental design	•	•					•
Data collection	•	•					
Data analysis	•	•	•				•
Manuscript	•	•		•	•	•	•

MTH = Mafruha Tasnin Hasan, AS[1] = Angela Sun, MM = Mehdi Mirzaei, JT = Junior Te'o, GH = Graham Hobba, AS[2] = Anwar Sunna, HN = Helena Nevalainen

3.3 Publication

Hasan MT, Sun A, Mirzaei M, Te'o J, Hobba G, Sunna A and Nevalainen H. 2017. A comprehensive assessment of the biosynthetic pathways of ascorbate, α-tocopherol and free amino acids in *Euglena gracilis* var. *saccharophila*. *Algal Res*. 27: 140-151.

Algal Research 27 (2017) 140–151

Contents lists available at ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

A comprehensive assessment of the biosynthetic pathways of ascorbate, α-tocopherol and free amino acids in *Euglena gracilis* var. *saccharophila*

Mafruha T. Hasan[a,b,c], Angela Sun[a,b,c], Mehdi Mirzaei[a,b,c,d,e], Junior Te'o[b,f], Graham Hobba[b,g], Anwar Sunna[a,b,c], Helena Nevalainen[a,b,c,*]

[a] Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia
[b] ARC-MQ Industrial Transformation Training Centre (ITTC) for Molecular Technology in the Food Industry, Sydney, Australia
[c] Biomolecular Discovery and Design Research Centre, Macquarie University, Sydney, Australia
[d] Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia
[e] Australian Proteome Analysis Facility, Macquarie University, Sydney, Australia
[f] School of Earth, Environmental and Biological Sciences, Queensland University of Technology, Brisbane, Australia
[g] Agritechnology Pty Ltd, Borenore, Australia

ARTICLE INFO

Keywords:
Euglena gracilis var. *saccharophila*
Proteomics
Biosynthetic pathways
Ascorbate
α-Tocopherol
Free amino acid biosynthesis

ABSTRACT

Euglena gracilis produces several important health-enhancing metabolites including ascorbate, α-tocopherol and free amino acids (faa). The yield of metabolites is dependent on the strain of *E. gracilis* and the metabolic growth condition. Here we investigated the effects of photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivation on the synthesis of ascorbate, α-tocopherol and faa in *E. gracilis* var. *saccharophila*, using label-free shotgun proteomics, and metabolite analysis using colourimetric assay, high-performance and ultra-performance liquid chromatography (HPLC/UPLC). PT cultivation resulted in the production of more antioxidants (up to 4.13 mg g^{-1} ascorbate and 2.52 mg g^{-1} α-tocopherol) than the MT and HT growth conditions (up to 0.97 and 0.50 mg g^{-1} ascorbate, and 1.40 and 0.21 mg g^{-1} α-tocopherol, respectively). The relative abundance of several faa varied between mid-log and initial stationary growth phases, but the total amount of faa remained about the same, with arginine as the most abundant amino acid. Proteomic analysis revealed a total of 3843 non-redundant proteins in *E. gracilis* var. *saccharophila*, of which 1890 were common among all cultivations. Gene ontology annotations suggested derivatisation of metabolic pathways from different organisms, such as lysine biosynthesis from fungi and serine biosynthesis from plants, while a few pathways were unique to *Euglena*, such as those of ascorbate and arginine. Some enzymes exhibited several isoforms that were influenced by the metabolic growth condition. For example, one of the isozymes of threonine aldolase was expressed in HT/MT cultures only, and one of the isozymes of phosphoglycerate dehydrogenase was expressed in PT cultures only. This is the first proteomic study of *E. gracilis* var. *saccharophila*, which provides a mechanistic insight into the biosynthetic pathway dynamics of primary metabolites (antioxidants and faa). This new information can serve as a framework for further development of *Euglena* as a producer of nutraceuticals.

1. Introduction

Microalgae have recently attracted interest as a platform for several industrial applications owing to their metabolic versatility. *Euglena gracilis* is a freshwater microalga that can be employed for bioremediation of heavy metals and toxic substances from polluted water [1,2], CO_2 sequestration [3], toxicity testing [4], biodiesel and biogas production [5,6], vitamin B_{12} assays [7], photo-, gravi- and polaro-taxis studies [8,9], nanoparticle and nano-fiber synthesis [10,11], and nutraceutical biosynthesis [12]. This wide range of uses can be attributed to the remarkably flexible metabolism of *E. gracilis*, acquired from a complex evolutionary history, with genes assimilated from proteobacteria, red algae and photosynthetic organisms [13]. *E. gracilis* thus has a vast armoury of genes that it can access to adapt to different environmental conditions.

E. gracilis naturally produces several compounds that have health-enhancing properties, including antioxidants that can remediate oxidative stress. Oxidative stress inside cells has been reported to have a strong correlation with numerous diseases, including cancer [14], cardiovascular diseases [15] and Alzheimer's disease [16]. *E. gracilis* also

* Corresponding author at: Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia.
E-mail address: helena.nevalainen@mq.edu.au (H. Nevalainen).

http://dx.doi.org/10.1016/j.algal.2017.08.029
Received 17 April 2017; Received in revised form 18 August 2017; Accepted 30 August 2017

synthesises all twenty of the protein-building amino acids, eight of which are essential in the human diet and four required by infants and young children [12]. The amount of health-enhancing organic compounds produced in *E. gracilis* depends largely on the strain and cultivation method used. Metabolic adaptability allows this alga to grow under photoautotrophic (PT) cultivation using inorganic CO_2 for photosynthesis, heterotrophic (HT) cultivation using organic carbon sources from its environment, as well as mixotrophic (MT) cultivation using both CO_2 and organic compounds. Several studies have been carried out to identify metabolites produced by *E. gracilis* under the influence of light and dark, and various carbon sources [12,17,18]. However, the effect of these cultivation conditions on the enzymes and pathways that produce the metabolites has not been examined in *E. gracilis* var. *saccharophila* before. The most common strain of *E. gracilis* used for metabolite studies is *E. gracilis* Z, which favours photosynthetic growth, although it can grow heterotrophically as well. Little is known about its sugar-loving counterpart, *E. gracilis* var. *saccharophila* and its metabolite production capabilities.

The initial features of the *E. gracilis* genome have recently been described [19]. However, the genome has not been fully annotated or made publicly available yet. The recent publication of the transcriptome has revealed insight into its exhaustive metabolic competency, with 32,128 predicted proteins [20]. A comprehensive translated proteome of *E. gracilis* will be difficult to compile as some proteins are only expressed when certain conditions are met. Moreover, the potentially expressed proteins described in the transcriptomic paper may differ from the actual proteins that are translated since *E. gracilis* extensively alters mRNA sequences prior to translation [20]. Nevertheless, in this study we aimed to identify the enzymes and pathways that *E. gracilis* var. *saccharophila* utilises to produce ascorbate, α-tocopherol and the twenty protein-building amino acids under PT, HT and MT cultivations. Label-free shotgun proteomics was used to find enzymes expressed at the proteomic level, and this was compared to data from metabolite analysis. We believe that this work will help to elucidate those transcripts that are preferentially translated in this strain, under PT, MT and HT cultivation respectively, from the wide range of predicted protein isoforms. This will in turn provide a valuable opportunity to explore metabolic networks in *E. gracilis*, which can be used for functional analysis and overexpression of the metabolites of interest.

2. Materials and methods

2.1. Growth medium

E. gracilis var. *saccharophila* (UTEX 752) was obtained from the University of Texas Culture Collection, USA. The strain was maintained as a pure axenic culture using modified Hutner medium, as the basal medium (pH 3.5), supplemented with vitamins B_1 and B_{12}, as described previously [12], with the addition of 10 g/l yeast extract (YE). For PT cultures (Section 2.2 below), the basal medium was prepared by dissolving all minerals and bases [12] in 1 l of MilliQ water, whereas for MT and HT cultures (Section 2.2 below), the same quantities of minerals and bases were dissolved in 925 ml of MilliQ water. The autoclaved basal medium for PT cultures was used directly for inoculation, whereas for the MT and HT cultures, the autoclaved basal medium was supplemented with 75 ml of separately autoclaved 20% D-(+)-glucose before inoculation, so that the final concentration of glucose in the medium was 1.5%. The glucose for the MT and HT cultures, and the vitamins for all cultures were added only after cooling the medium to room temperature, before inoculation. All chemicals and reagents were purchased from Sigma Aldrich, Australia unless stated otherwise.

2.2. Culture conditions

Cultivation was carried out under three different nutritional modes: PT (photoautotrophic), MT (mixotrophic) and HT (heterotrophic). The difference between the cultures was the source of carbon. PT cultures were able to utilise mainly atmospheric CO_2 via photosynthesis, with some carbon from the amino acids present in YE. The HT cultures (in the dark) utilised an organic carbon source (glucose) with YE. MT cultures were able to utilise both atmospheric CO_2 and glucose from the medium, along with YE.

Each set of cultivations was performed with three biological replicates. The PT cultures were grown in basal medium containing no glucose and were exposed to white light illumination (2000 lx) for a photoperiod of 14 h light/10 h darkness. The MT cultures were grown in the same lighting condition as PT, but the basal medium was supplemented with glucose. The HT cultures were grown in the same medium as MT cultures, but in complete darkness. Light-adapted and dark-adapted stock cultures, maintained under the conditions mentioned above, for over 2 years, were used to inoculate the PT and MT cultures, and HT cultures respectively. All stock cultures were starved (basal medium not replenished) for 3 days prior to inoculation, and no glucose was added to the stock cultures for 10 days prior to inoculation. For each culture, an initial concentration of 1.5×10^8 cells of the respective stock was inoculated in 150 ml of medium in 250 ml Erlenmeyer flasks. All cultures were incubated at 23 °C with orbital shaking at 150 rpm.

2.3. Determination of dry mass

Every 24 h, 5 ml of cells were harvested by centrifugation at 5000*g* for 10 min and the supernatant was discarded. The pellet was dried in a pre-weighed aluminium evaporating dish at 70 °C for 48 h. The dish was cooled to room temperature in a desiccator for 10 min before weighing.

2.4. Determination of ascorbate

Ascorbate was determined every 24 h, for 72 h, from inoculation to the start of the stationary phase. The cells were harvested, washed twice with 10 ml of cold Milli-Q water, and resuspended in 3 ml of cold lysis buffer. The lysis buffer was prepared by dissolving one PBS tablet (MP Biomedicals, Australia) in 100 ml of Milli-Q water, supplemented with 0.04% (v/v) β-mercaptoethanol and protease inhibitor (2 tablets of cOmplete™ Protease Inhibitor Cocktail, Roche) before use. The samples were lysed by French press treatment (Thermo Spectronic French Pressure Cell Press, MA, USA) at 15,000 psi. The lysate was centrifuged at 2000*g* for 5 min. One hundred μl of the lysate supernatant was deproteinised by ultrafiltration through a 10 kDa nominal MWCO centrifugal filter (Amicon® Ultra Centrifugal Filters, Merck Millipore, Australia) by centrifuging at 14,000*g* for 1 h. The remaining lysate supernatant was used for free amino acid determination (Section 2.6) and protein extraction (Section 2.7) described below, for the respective days. Ascorbate from the deproteinised flow-through was then measured using the Ascorbic Acid Assay Kit (MAK074, Sigma Aldrich, Australia) as per the manufacturer's protocol.

2.5. Determination of α-tocopherol

Alpha tocopherol was determined at the same time points as ascorbate. Extraction was carried out as previously described [12]. An Agilent 1260 Infinity Semi-Preparative High Performance Liquid Chromatography-Diode Array Detector (HPLC-DAD) system (Agilent Technologies, Australia) was used to determine α-tocopherol. A Symmetry C18 150 mm × 3.5 μm reverse phase HPLC column (i.d. 2.1 mm, Waters, Australia), with an injection volume of 10 μl and a flow rate of 0.5 ml min^{-1} was employed at room temperature. An isocratic method was used, eluting the components with a methanol:acetonitrile (50:50 v/v) mobile phase, and α-tocopherol was detected by UV absorbance at 294 nm. Quantitation was carried out using a standard calibration

curve. The identity of the α-tocopherol peak was confirmed by two separate methods: spiking the sample with α-tocopherol standard to identify the peak, and collecting the peak fractions and running them through tandem mass spectrometry (MS/MS).

2.6. Determination of free amino acids

Free amino acid (FAA) was determined for the mid-log phase (24 h for MT and HT cultures, 48 h for PT cultures) and the beginning of the stationary phase (48 h for MT and HT cultures, and 72 h for the PT cultures). Concentrations of the common 20 proteinogenic amino acids were determined using pre-column derivatisation amino acid analysis with 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (AQC, Waters, Australia).

One hundred μl of the leftover lysate supernatant of the ascorbate assay (Section 2.4) from the respective time points were used. This was diluted with an internal standard of DL-norvaline solution and deproteinised, as described in Section 2.4. Twenty μl of the filtrate was then subjected to AQC derivatisation at 50 °C for 10 min and analysis based on the method described by Cohen [21], but modified for compatibility with a Waters ACQUITY™ reverse phase Ultra Performance Liquid Chromatography (RP-UPLC) system. An ACQUITY UPLC BEH C18 100 mm × 1.7 μm column (i.d. 2.1 mm, Waters, Australia) was used, with a binary gradient flow rate of 0.7 ml min^{-1} at 60 °C. All amino acids were detected by UV absorbance at 260 nm. Chromatographic analysis and quantitation was carried out using the Empower software (Waters, Australia), using a set of prepared standards, with DL-norvaline as the internal standard.

2.7. Protein extraction and fractionation by SDS-PAGE

Proteomic analysis was carried out during the mid-log phase only (24 h for MT and HT cultures and 48 h for PT cultures). The remaining lysate supernatants from the ascorbate assay (Section 2.4) of the respective days were used for protein precipitation as previously described [22]. The air-dried pellet was solubilised in 100 μl of 5% SDS and the protein concentration was measured by BCA assay as per the manufacturer's protocol (Thermo Scientific Pierce, Australia).

One hundred μg of protein per sample were fractionated in the reduced state by SDS-PAGE. The samples were separated in precast 4–12% gradient NuPAGE® Bis-Tris Mini Gels at 200 V for 50 min in MOPS running buffer. The gels were stained with Coomassie Brilliant Blue for visualisation [22] and each lane was cut into 16 portions of about the same size, which were further sliced into smaller pieces before in-gel digestion.

2.8. Trypsin in-gel digestion, peptide extraction and purification

Trypsin in-gel digestion and peptide extraction were carried out as previously described [22], with the modification that the dried peptide extracts were reconstituted in 20 μl of 2% (v/v) formic acid (FA, Merck, Australia). The extracts were purified using ZipTip® Pipette Tips with C_{18} resin (Merck Millipore, US). The purified extracts were dried using a vacuum centrifuge, reconstituted in 20 μl of 2% (v/v) FA and centrifuged at 18,000g for 15 min. The top 10 μl of the reconstituted peptides were used for nanoflow liquid chromatography (nanoLC)-MS/MS.

2.9. NanoLC-MS/MS and protein/peptide identification

NanoLC-MS/MS was carried out as described previously [23], using a QExactive mass spectrometer (Thermo Scientific, CA, USA), to obtain raw MS files from the Xcalibur software (version 2.06, Thermo Scientific, US). The raw files were converted to mzXML format and searched using the global proteome machine (GPM) software version 2.2.1 with the X!Tandem algorithm (http://www.thegpm.org). The *Euglena* non-redundant protein database [20], and the NCBI *Euglena* database were stitched together, and the mzXML files were searched against this stitched database, along with common human and trypsin peptide contaminants. Reverse database searching was also used for estimating false discovery rates (FDR). The *Euglena* non-redundant protein database from the John Innes Centre website (http://jicbio.nbi.ac.uk/euglena/) and the *Euglena* protein database from NCBI (http://www.ncbi.nlm.nih.gov/) were acquired in September 2016. The GPM search parameters were set as previously described [22]. The 16 fractions of each replicate were processed consecutively on the GPM with output files for each fraction, as well as a simple merged non-redundant output file for peptide and protein identifications with log(*e*) values less than −1.

2.10. Data processing, quality control, and quantitative and statistical analyses

The low-quality hits from the GPM were filtered out to produce a high stringency dataset of reproducibly identified proteins present in each cultivation. The criteria for filtering were that the proteins must be present in all three biological replicates, and the total spectral count of each protein must be five or more. Peptide and protein FDRs were then calculated (FDR threshold was set lower than 1%), followed by the removal of the reversed database hits and calculation of the normalised spectral abundance factors (NSAF) [24]. Quantitative proteomic analysis and statistical analysis were carried out, using summed NSAF values as a measure of relative protein abundance [24]. A spectral fraction of 0.5 was added to all spectral counts to compensate for null values and enable log transformation for the statistical analysis. Two sample unpaired *t*-tests were performed on the log-transformed NSAF dataset, and proteins with a p-value < 0.05 were considered differentially expressed, to yield a set of up-regulated, down-regulated and unchanged proteins. An analysis of variance (ANOVA) was also performed on proteins present reproducibly across all 3 cultivation conditions. For ANOVA, again the log-transformed NSAF values were used and the proteins with a p-value < 0.05 were regarded as having a significant change in abundance among the cultivations.

2.11. Functional annotation

The resulting set of differentially expressed and unchanged proteins were functionally annotated using BLASTP. The corresponding sequences from the *Euglena* non-redundant peptide database were compared to the UniProtKB/Swiss-Prot database using BLASTP with an *E*-value cut-off of $1e^{-10}$. This similarity analysis showed that of the expressed proteins identified from the *Euglena* non-redundant protein database via nanoLC-MS/MS, about 30% had unknown function. The Gene Ontology (GO) information of the proteins that could be functionally annotated was retrieved from the UniProt database.

3. Results and discussion

3.1. Dry mass of E. gracilis var. saccharophila

The exponential growth phase for MT and HT cultures, and PT culture occurred approximately between 0 and 48 h, and 0 and 72 h respectively (Fig. 1). The average dry mass for HT cultures reached up to 12.91 g l^{-1} at its highest, while that of MT and PT cultures reached only about 9.42 g l^{-1} and 2.38 g l^{-1} respectively. The growth pattern was used to determine the mid-log phase, which was at about 24 h after inoculation for the MT and HT cultures, and about 48 h for the PT cultures. The mid-log phase represents the most suitable time for studying proteomic profiles related to metabolite production during rapid growth.

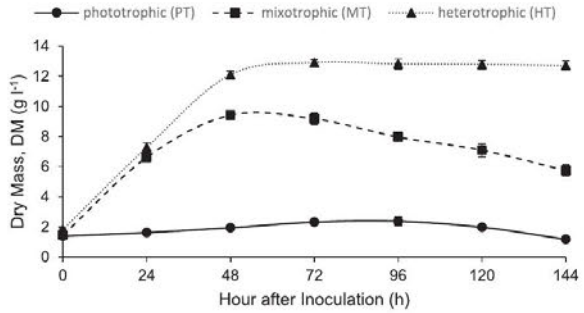

Fig. 1. Growth patterns of *Euglena gracilis* var. *saccharophila* under different cultivation conditions: phototrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivation respectively. The exponential growth phase occurs between 0 and 48 h for HT and MT cultures, and up to 72 h for the PT cultures. HT cultures were marked by a prolonged stationary phase, while the MT cultures and the PT cultures experienced a shorter stationary phase.

3.2. Analysis of label-free shotgun proteomics

The GPM results for all replicates are included in Supplementary Table S1. A total of 3843 proteins were identified across all three growth conditions, of which 1890 non-redundant proteins were reproducibly identified among the three growth conditions (Fig. 2). This number is much lower than the total number of proteins predicted in the transcriptome [20], and part of it could be owing to the limitation in detection of the tandem MS. However, the vast difference is likely to indicate that *E. gracilis* does not control protein expression at the transcription level.

The highest number of proteins was expressed under PT cultivation (3108), while the HT cultures showed the fewest number of proteins (2442), with MT cultures in between (2851). About 30% of the proteins could not be annotated with confidence, having a BLASTP *E*-value of greater than $1e^{-10}$. Protein classification of the remaining proteins, using GO information, showed that most of the missing proteins in HT cultures, compared to PT cultures, were photosynthesis-related. Several other proteins belonging to the biological categories of generation of precursor metabolites and energy, and transport were also missing or down-regulated. Compared to PT cultures, most of the proteins missing in MT cultures belonged to the biological category of cellular metabolic process, including organic acid, cellular aromatic compound, and cellular nitrogen compound metabolic processes. Several other proteins from lipid and macromolecule metabolism, and generation of precursor metabolites and energy were also absent or down-regulated in MT cultures.

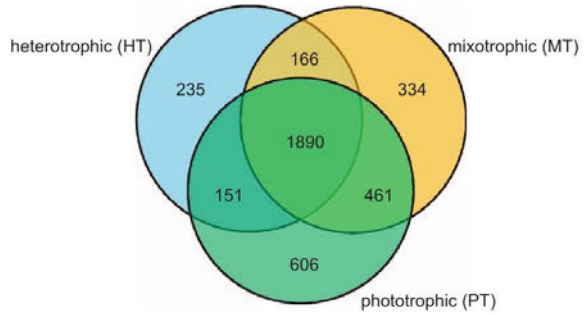

Fig. 2. Venn diagram of the number of proteins common between and among growth conditions: phototrophic (PT), mixotrophic (MT) and heterotrophic (HT). PT cultures had the highest number of proteins (3108), while HT cultures had the lowest (2442). More proteins were shared between PT and MT cultures than any other combination (2351).

The redundant count of peptides (rC_p), which is the sum of the number of peptides identified for proteins in each biological replicate, was consistent among the replicates of each treatment (Table 1), with the coefficient of variation ranging between 1.3 and 4.4%. The FDR for peptides was < 0.16% for all samples, while the FDR for proteins was < 1% (Table 1), indicating that there is low variability and high reproducibility among replicates, and that the data did not require any further filtering.

3.3. Differentially expressed proteins among the three growth conditions and pairwise comparison (2-sample test)

Of the 1890 proteins, reproducible in all three cultures, 910 showed significant differential expression at the 0.05 confidence level from ANOVA across all cultivations (Supplementary Table S2).

For the 2-sample pairwise comparison between PT and the other two cultures (Supplementary Table S2), the calculated NSAF values from Section 2.10 were used. The NSAF value for each protein indicated the abundance of that protein, relative to the entire pool of proteins, at the time of protein extraction. The NSAF values for proteins from the PT cultures were compared, using unpaired two sample *t*-test (Section 2.10), to the NSAF values for the same proteins in the other two cultures. This gave the relative abundance of the respective proteins in HT and MT cultures, compared to the PT cultures. Overall, HT culture showed 660 down-regulated and 677 up-regulated proteins, while MT culture showed 571 down-regulated and 519 up-regulated proteins. Using the confidence level of 0.05, a total of 1668 and 1965 proteins were classified as unchanged between PT and HT cultures, and PT and MT cultures, respectively.

The proteins were classified using their GO information to identify proteins belonging to the biosynthetic pathways of ascorbic acid, α-tocopherol and the twenty free amino acids. The *E*-value from BLASTP results, along with the EC numbers of the candidate proteins and fold changes with respect to PT cultures, are provided in Supplementary Table S3.

3.4. Ascorbate

The changes in ascorbate content over the first 72 h showed a similar trend across all cultivation conditions (Fig. 3). This is comparable to Shigeoka et al. [25], with the content increasing during the early log-phase, and then decreasing for the remainder of the log-phase. By the beginning of the stationary phase, the ascorbate content returned to the initial concentration per g of dry mass (DM). The highest amount of ascorbate was found in PT culture at 24 h (4.13 mg g^{-1} of DM), which is comparable to previously reported ascorbate contents in *E. gracilis* Z of 4.25 mg g^{-1} [25] and 4.16 mg g^{-1} [26]. The HT cultures produced insignificant amounts of ascorbate, with about 0.50 mg g^{-1} of DM at their highest yields, at 24 h. The MT cultures did not produce as much ascorbate in this strain (0.97 mg g^{-1} of DM), as in other reported strains [25].

There are three major ascorbate biosynthetic pathways in eukaryotes, including the animal pathway, the plant pathway and the *Euglena* pathway [27]. The last pathway is unique to *Euglena* and has been described previously [27,28]. The transcriptome of *E. gracilis* revealed candidates for the enzymes of all three pathways, although several were loosely assigned owing to the difficulty in assigning enzymes that have sugar-based substrates [20]. Here, we looked at the dominant pathway for ascorbate synthesis in *E. gracilis*, via galacturonate/L-galactonate [29]. All enzymes in the pathway could be identified from the transcriptome, except for the galacturonate reductase that converts D-galacturonate to L-galactonate [30]. This enzyme has been characterised in *E. gracilis*, and its N-terminus sequenced, but it is very closely related to malate dehydrogenase and thus difficult to annotate [30]. Some of the ascorbate pathway enzymes were not found in the proteome of HT cells (Table 2), while a few others did not meet the

Table 1
Summary of proteins identified under phototrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivations for *E. gracilis* var. saccharophila, and their corresponding FDR values.

Growth condition[a]	No. of identified proteins[b]			Redundant count of peptides (rC_p)[b]			No. of proteins[c]	Protein FDR[d] (%)	Peptide FDR[d] (%)
	R1	R2	R3	R1	R2	R3			
PT	4991	4719	4767	115,022	114,741	112,275	3108	0.35	0.07
MT	4649	4796	4670	108,454	116,789	109,454	2851	0.73	0.15
HT	3871	3891	4084	83,001	90,614	87,619	2442	0.97	0.11

[a] PT, MT and HT: phototrophic, mixotrophic and heterotrophic respectively.
[b] R1, R2 and R3: Replicates 1, 2 and 3 respectively.
[c] Total number of proteins common to all three replicates.
[d] FDR: false discovery rate.

filtering criteria used in this study (Section 2.10).

In PT and MT cultivations, the primary photochemical reactions that take place during photosynthesis produce ample O_2, which leads to the generation of a large amount of reactive oxygen species (ROS). ROS are also produced by the electron transport system in the mitochondria and plasma membrane [31], as well as other metabolic processes localised in different cellular compartments, such as the cytosol, endoplasmic reticulum, peroxisomes, Golgi apparatus, lysosomes, cell wall and apoplast [32–34]. Environmental stresses including hypoxia, drought, low temperature, salinity, radiation, high metal toxicity and pathogenic attacks can also elevate the generation of ROS [32–34], in order to signal a stress response through cellular communication [34]. Ascorbate is an important antioxidant as it is catabolised in the presence of H_2O_2 by ascorbate peroxidase. A single peroxidase exists in *E. gracilis*, and it was found across all cultivations, but was highly up-regulated under PT cultivation, followed by MT cultivation (Table 2). This enzyme also scavenges other ROS in *E. gracilis*, such as superoxide radicals and alkyl hydroperoxides, thus preventing oxidative stress and lipid hydroperoxidation that can lead to membrane damage [35]. This coincides with recent reports about higher polyunsaturated fatty acid (PUFA) composition of light-grown cells [36]. PUFA are more susceptible to oxidative stress due to the presence of double bonds and higher antioxidant levels can protect against free radical-induced lipid peroxidation. The requirement to combat higher levels of oxidation explains why there is more ascorbate found in PT cultivations.

3.5. α-Tocopherol

As reported for other strains [37,38], the concentration of α-tocopherol per dry mass of cells was highest in PT cultures, followed by MT cultures, while the HT cultures had significantly lower α-tocopherol concentration (Fig. 4). For the PT cultures, α-tocopherol increased over 72 h of cultivation, similar to other strains [39], reaching up to 2.52 mg g^{-1}. α-Tocopherol levels per g DM decreased during the first 24 h for the MT cultures, and increased from thereon, rising to 1.40 mg g^{-1}. At 72 h, the concentration of α-tocopherol in the MT cultures was still lower than the starting concentration. The HT cultures followed no specific trend and only reached a maximum of 0.21 mg g^{-1}, which was also lower than the starting concentration. The highest amount of α-tocopherol for all cultures is comparable to those reported in the literature for other strains [6,12,38].

There are contradictory reports in the literature regarding the distribution of tocopherols within cells and its relation to photosynthesis [26,37,39–41]. The variations in these reports, coupled with the presence of tocopherols in the dark, which increases during light cultivation, suggest that both mitochondria and chloroplasts may play important roles in α-tocopherol synthesis [42].

All the enzymes required for α-tocopherol synthesis could be found in the transcriptome [20]. However, in the current study, the candidate sequences for tocopherol cyclase (EC 5.5.1.24) and tocopherol *O*-methyltransferase (EC 2.1.1.95) were not found at the proteome level (Table 2). The remaining enzymes of the pathway were identified in the proteome and all were up-regulated in PT cultures (Table 2). This finding could be for the same antioxidant properties as ascorbate (Section 3.4), especially since this is a lipid soluble antioxidant.

3.6. Free amino acids

The synthesis patterns of the twenty amino acids showed discrepancy among the cultivations used in this study, with some present at higher concentrations under a specific cultivation, while others remained unchanged or diminished for the same cultivation (Table 3). The total amino acid content was about the same across cultivations, although the ratios of amino acids were different. The concentration of amino acids, however, may have been influenced by the uptake of amino acids from the medium, which contained yeast extract. A summary of the amino acid biosynthetic pathways discussed here is shown in Fig. 5.

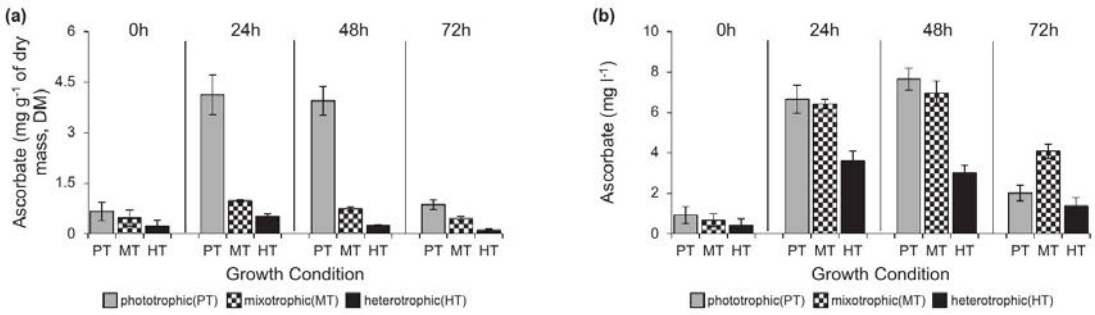

Fig. 3. Ascorbate content of *Euglena gracilis* var. *saccharophila* from inoculation to the beginning of the stationary growth phase, in (a) yield per g of dry mass (mg g^{-1} of DM) and (b) volumetric concentration per litre of culture (mg l^{-1}). The same trend of producing more ascorbate during the exponential phase (0–48 h), followed by a decline when the stationary phase commences (48–72 h), was observed across all three cultivation conditions used. Phototrophic (PT) cultures produced considerably more ascorbate per DM than the other two cultures.

Table 2
Tabulated heat map of proteins from the ascorbate, α-tocopherol and free amino acid pathways during the mid-log phase of growth for *E. gracilis* var. *saccharophila*, created using normalised spectral abundance factor (NSAF) values. The identities (IDs) represent those from the published transcriptome [20] or NCBI protein IDs.

PT	MT	HT	ID	Description
A Ascorbate biosynthesis				
			light_m.8509	UDP-glucose 6-dehydrogenase
			light_m.21407	UDP-glucose/glucuronate 4-epimerase
			light_m.20758	UTP-hexose-1-phosphate uridylyltransferase
			light_m.57669	Galactokinase
			gi|161727354|	Aldonolactonase
			gi|667667677|	L-galactonolactone dehydrogenase
			gi|229002753|	Ascorbate peroxidase
B α-tocopherol biosynthesis				
			dark_m.13964	4-hydroxyphenylpyruvate dioxygenase
			light_m.13577	
			light_m.48543	HPT/GGT
			dark_m.3700	2-MPHMT
			dark_m.38942	Tocopherol cyclase
			light_m.59814	Tocopherol O-methyltransferase
C Amino acid biosynthesis				
C.1 Trp, Phe, Tyr				
			light_m.32049	DAHPS
			light_m.52098	AROM complex
			dark_m.12955	Chorismate synthase
			light_m.97034	Shikimate dehydrogenase
			light_m.64084	Shikimate kinase
			dark_m.24194	Anthranilate synthase (phenazine specific)
			light_m.68214	
			dark_m.33057	MTSE
			dark_m.9334	Tryptophan synthase
			light_m.34856	
			light_m.67331	ChrM-PD; bifunctional
			light_m.43561	ChrM-PD/AD-PDH/ADH
			light_m.47105	L-amino-acid oxidase
			light_m.51611	Phenylalanine hydroxylase
C.2 Gly, Ser, Thr				
			light_m.15606	GLO
			light_m.25655	
			light_m.67200	L-threonine aldolase
			light_m.67529	
			dark_m.6710	Glycine dehydrogenase[1]; GDC
			light_m.17026	Aminomethyltransferase; GDC
			light_m.12684	Dihydrolipoyl dehydrogenase; GDC
			dark_m.20652	
			gi|47600753|	
			dark_m.31849	SHMT
			dark_m.19245	
			dark_m.7243	
			light_m.9457	3-Phosphoglycerate dehydrogenase
			dark_m.8278	
			light_m.9629	
			light_m.7044	Phosphoserine transaminase
			light_m.35742	PSP/HSK
			dark_m.38833	SAL
			light_m.15790	Bifunctional SAL/ThrAL
			light_m.48603	Aspartate kinase/homoserine dehydrogenase[2]
			light_m.25814	Aspartate-semialdehyde dehydrogenase
			light_m.20768	Homoserine dehydrogenase
			light_m.36284	Threonine synthase
			light_m.62724	ThrAL
C.3 Val, Ile, Leu				
			light_m.11204	Acetolactate synthase
			dark_m.25329	
			light_m.28738	
			dark_m.442	Ketol-acid reductoisomerase
			light_m.22604	Dihydroxy-acid dehydratase
			light_m.32069	Branched-chain-amino-acid transaminase
			dark_m.16450	2-isopropylmalate synthase
			light_m.48249[3]	
			light_m.11199	3-isopropylmalate dehydratase
			light_m.15833[4]	3-isopropylmalate dehydrogenase
			light_m.13576	
C.4 Met, Cys				
			light_m.28332	CBS
			dark_m.15720	CBL
			light_m.24557	CGL/CBL/CGS
			dark_m.18447	
			light_m.16136	
			light_m.10253	Methionine synthase
			dark_m.32578	
			gi|296784641|	Methionine S-adenosyl transferase
			dark_m.51101	SAT
			light_m.43584	
			dark_m.17386	CysS
			dark_m.17385	
			light_m.22452	
			dark_m.37469	
			light_m.56039	
			dark_m.26706	
			light_m.38496	
C.5 Asp, Asn, Glu, Gln, Ala				
			light_m.77518	probable Aspartate transaminase
			dark_m.21257	L-aspartate oxidase
			dark_m.21260	
			light_m.26971	
			dark_m.31760	Adenylosuccinate synthase
			light_m.26836	
			light_m.26849	
			light_m.26845	
			dark_m.21830	Adenylosuccinate lyase
			dark_m.19787	Asparagine synthase (gln-hydrolysing)
			light_m.37192	Asparaginase
			light_m.46896	
			dark_m.462	Glutamate synthase (NADH/NADPH)
			dark_m.24254	
			light_m.6736	Glutamate dehydrogenase, NAD-linked
			light_m.13579	
			dark_m.86	
			dark_m.150	Glutamate-ammonia ligase
			light_m.89	Alanine dehydrogenase
C.6 Arg, Pro				
			light_m.13208	CarbPS (glu-hydrolysing)
			light_m.9616	
			gi|1005746202|	Ornithine carbamoyltransferase 1
			gi|1005746216|	Ornithine carbamoyltransferase 2
			dark_m.3269	Argininosuccinate synthase
			dark_m.26272	Argininosuccinate lyase
			dark_m.20593	Arginine deiminase
			light_m.12713	Glutamate-5-kinase[5]
			dark_m.10754	
			dark_m.10755	
			light_m.25543	
			dark_m.52519	Pyrroline-5-carboxylate reductase
			dark_m.2778	Ornithine aminotransferase
			light_m.36885	L-glutamate γ-semialdehyde dehydrogenase
			dark_m.3408	Proline dehydrogenase
			light_m.30651	
C.7 Lys, His				
			light_m.24228	Homocitrate synthase
			light_m.36101	Homoaconitase-homoaconitate hydratase
			light_m.38051	Homoisocitrate dehydrogenase
			light_m.23118	L-aminoadipate-semialdehyde dehydrogenase
			light_m.21957	
			light_m.12750	SacD (NAD+/NADP+, L-glu/lys-forming)
			light_m.14473	
			dark_m.6924	Ribose-phosphate diphosphokinase
			light_m.32896	
			dark_m.24604	ATP phosphoribosyltransferase
			dark_m.59999	ProFAR-I
			light_m.68800	
			light_m.49662	IGPS
			light_m.23842	Imidazoleglycerol-phosphate dehydratase
			light_m.14924	Histidinol-phosphate transaminase
			dark_m.65978	Histidinol dehydrogenase
			light_m.58833	

Not Detected Down Up

Footnote: Enzymes that could not be identified from the transcriptome by amino acid sequence alone, have not been shown here, such as galacturonate reductase, citrullinase and most of the aminotransferases/transaminases. [1]Aminomethyl-transferring; [2]similar multifunctionality as in *E. coli*; [3]may be bifunctional with EC 2.3.1.182; [4]may be multifunctional with EC 1.1.1.87/42; [5]predicted to be multifunctional. HPT: homogentisate phytyltransferase; GGT: geranylgeranyltransferase; 2-MPHMT: 2-methyl-6-phytyl-1,4-hydroquinone methyltransferase; DAHPS: 3-deoxy-7-phosphoheptulonate synthase; MTSE: multifunctional tryptophan-synthesising enzyme; ChrM: chorismate mutase; PD: prephenate dehydratase; AD: arogenate dehydratase; PDH: prephenate dehydrogenase; ADH: arogenate dehydrogenase; GLO: (*S*)-2-hydroxy-acid oxidase; GDC: glycine decarboxylase; SHMT: serine hydroxymethyltransferase; PSP: phosphoserine phosphatase; HSK: homoserine kinase; SAL: L-serine ammonia-lyase; ThrAL: L-threonine ammonia-lyase; CGL: cystathionine gamma-lyase; CBL: cystathionine beta-lyase; CGS: cystathionine gamma-synthase; CBS: cystathionine beta-synthase; SAT: serine *O*-acetyltransferase; CysS: cysteine synthase; CarbPS: carbamoyl-phosphate synthase; SacD: saccharopine dehydrogenase; ProFAR-I: 1-(5-phosphoribosyl)-5-((5-phosphoribosylamino)methylideneamino)imidazole-4-carboxamide isomerase; IGPS: imidazoleglycerol-phosphate synthase.

3.6.1. Tryptophan, phenylalanine and tyrosine

The preliminary steps for the biosynthesis of aromatic compounds, including tryptophan, phenylalanine and tyrosine, involve the shikimate pathway via chorismate, like in plants and other algae [43]. This pathway in *E. gracilis* contains two monofunctional enzymes (EC 2.5.1.54 and 4.2.3.5) for the first and last steps, with a pentafunctional enzyme (AROM complex) carrying out the intermediate steps, which is similar to that in fungi [44]. All three enzymes were up-regulated under HT cultivation (Table 2). This indicates a high demand for aromatic compound biosynthesis during rapid cell proliferation.

The conversion of chorismate to tryptophan is unique in *E. gracilis* as it uses a key regulatory enzyme to convert chorismate to anthranilate (anthranilate synthase, EC 4.1.3.27), while the rest of the steps are carried out by a single multifunctional tryptophan-synthesising enzyme (MTSE) [45]. Two isoforms of anthranilate synthase were found in the proteome, with one up-regulated during HT cultivation, and one during PT cultivation. MTSE was up-regulated during HT and MT cultivations, resulting in more tryptophan synthesis. This was also reflected in the FAA analysis (Table 3) and may be due to the faster growth rate and higher protein synthesis rate during these cultivations. Tryptophan is also required for the production of other aromatic compounds such an indoleacetate (IAA), which encourages faster growth and stimulates paramylon accumulation [46]. The enzymes for this conversion were also identified in our proteomic profiles (data not shown).

The enzymes for the other two aromatic amino acids were shared or correlated. A probable multifunctional enzyme was identified for chorismate mutase-prephenate/arogenate dehydratase-prephenate/arogenate dehydrogenase (EC 5.4.99.5, 4.2.1.51/91, 1.3.1.78, 1.3.1.12/13). Although it appears that most of the enzymes were up-regulated during HT and MT cultivations, the FAA analysis shows an upregulation of phenylalanine only (Table 3), while tyrosine was 2-fold higher in PT cultures. This observation is likely due to the higher conversion of phenylalanine to tyrosine in PT cultures by phenylalanine hydroxylase (Table 2). The FAA analysis further shows that phenylalanine reduces across all cultivations as the stationary phase commences and growth slows down. Tyrosine is required as an electron donor in photosynthesis, and as a precursor to α-tocopherol. The yield remains approximately the same between log phase and the beginning of stationary phase for HT cultures, but increases for PT and MT cultures. This is consistent with the increase in α-tocopherol production.

Table 3
Concentration of free amino acids (mg g^{-1} of dry mass, DM) in *Euglena gracilis* var. *saccharophila*, during the mid-log phase of phototrophic (PT, 48 h), mixotrophic (MT, 24 h) and heterotrophic (HT, 24 h) cultivations and the start of the stationary phase (72 h for PT, 48 h for MT and HT).

Amino acid	PT (mg g^{-1} DM)		MT (mg g^{-1} DM)		HT (mg g^{-1} DM)	
	48 h	72 h	24 h	48 h	24 h	48 h
Trp	0.24[a]	0.27[a]	0.44[a]	0.05	0.51[a]	0.16[a]
Phe	0.55[a]	0.25[a]	0.68[a]	0.22	0.91[b]	0.32[a]
Tyr	0.24[a]	0.90[a]	0.12	0.63[a]	0.11	0.11
Gly	0.20[a]	0.21[a]	0.18	0.19	0.18[a]	0.20
Ser	0.44[a]	0.23[a]	0.18[a]	0.13	0.13	0.10
Thr	0.51[a]	0.42[a]	0.31	0.09[a]	0.37[a]	0.07
Val	0.33[a]	0.29[a]	0.48[a]	0.20[a]	0.50[a]	0.14
Ile	0.24[a]	0.28[a]	0.20	0.45[a]	0.17[a]	0.44[a]
Leu	0.27[a]	0.21[a]	0.45[a]	0.04	0.42	0.04
Met	0.52[a]	0.11	0.11	0.05	0.08	0.05
Cys	0.64[b]	0.72[a]	0.45[a]	1.14[a]	0.46[a]	0.51[a]
Asp	2.65[b]	2.62[b]	3.01[b]	4.45[b]	4.01[b]	5.22[b]
Asn	7.22[b]	4.63[b]	10.88[b]	1.85[b]	11.50[c]	3.47[b]
Glu	0.86[a]	0.79[a]	0.93[a]	0.90[a]	0.84[a]	0.86[a]
Gln	2.84[b]	1.01[b]	2.49[a]	1.44[b]	2.02[b]	1.41[a]
Ala	0.34[a]	0.39[a]	1.48[b]	2.16[b]	1.16[b]	2.37[b]
Arg	23.37[c]	19.97[d]	18.41[b]	15.44[b]	18.45[c]	14.01[c]
Pro	0.19[a]	0.20[a]	0.23[a]	0.25[a]	0.21[a]	0.22
Lys	0.82[a]	1.66[b]	1.13[a]	0.36	1.06[a]	0.42[a]
His	3.17[b]	2.62[b]	1.30[a]	0.66[a]	1.61[b]	0.73[a]
Total	45.64	37.79	43.44	30.71	44.72	30.85

All other SD < ± 0.00981.
PT: phototrophic; MT: mixotrophic; HT: heterotrophic.
[a] SD < ± 0.0997.
[b] SD < ± 0.678.
[c] SD < ± 1.39.
[d] SD = ± 2.97.

3.6.2. Glycine, serine and threonine

Glycine and serine are required for protein biosynthesis, as precursors to several compounds such as pyrimidines and purines, as one-carbon units in many systems, and as intermediates of the

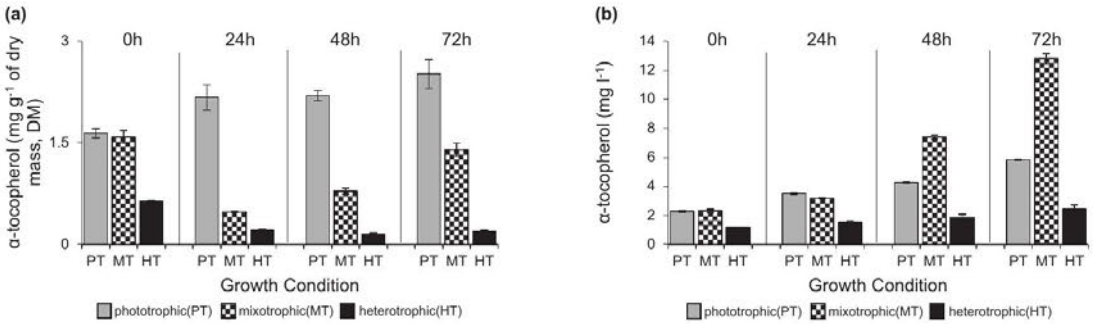

Fig. 4. α-Tocopherol content of *Euglena gracilis* var. *saccharophila* from inoculation to the initial stationary growth phase, in (a) yield per g of dry mass (mg g^{-1} of DM) and (b) volumetric concentration per litre of culture (mg l^{-1}). Phototrophic (PT) cultures had the highest yield per DM, although mixotrophic (MT) cultures had more α-tocopherol per volume of culture by mid-log phase (48 h).

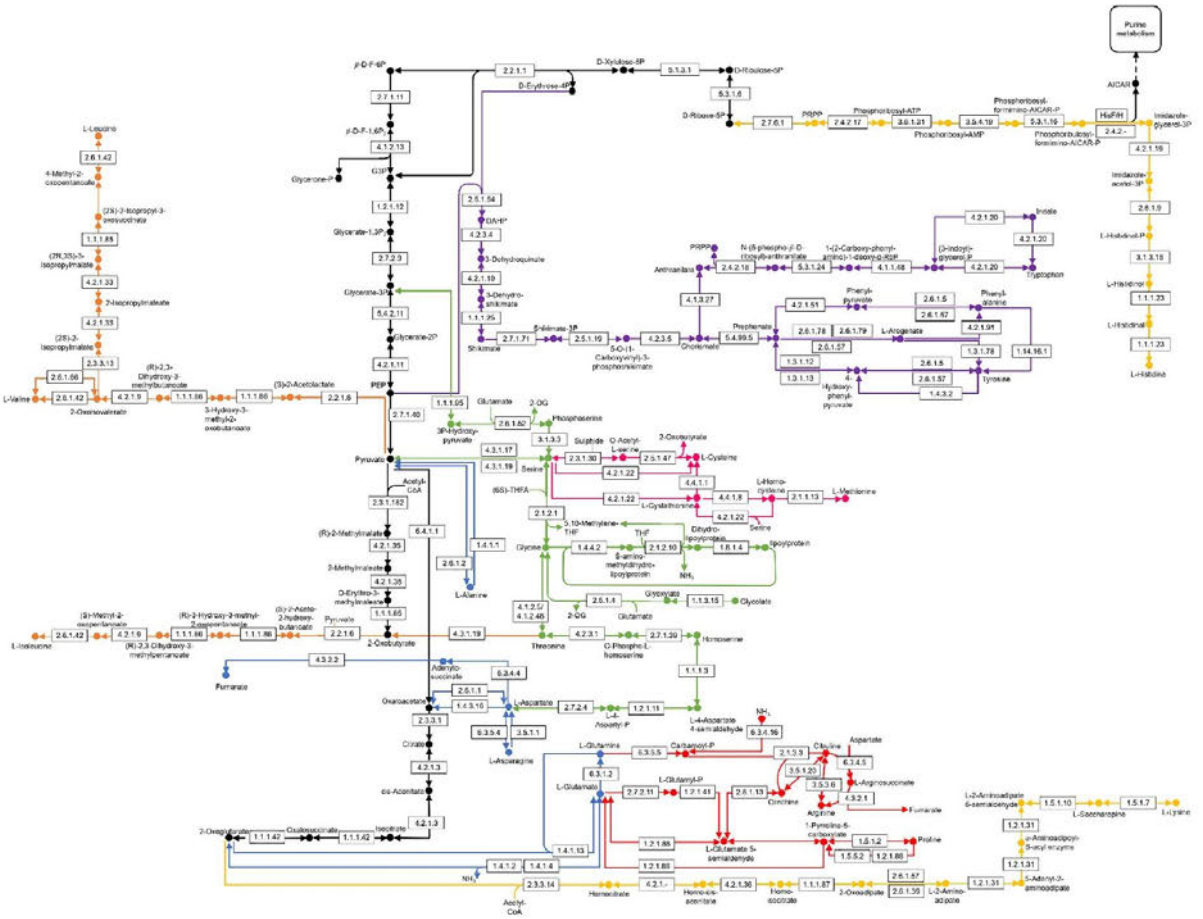

Fig. 5. The possible routes for amino acid biosynthesis in *E. gracilis* var. *saccharophila*: (●→) tryptophan, phenylalanine and tyrosine; (●→) glycine, serine and threonine; (●→) valine, isoleucine and leucine; (●→) methionine and cysteine; (●→) aspartic acid, asparagine, glutamic acid, glutamine and alanine; (●→) arginine and proline; and (●→) lysine and histidine. Footnote: The abbreviations in the pathway are β-D-fructose 6-phosphate (β-D-F-6P); glyceraldehyde 3-phosphate (G3P); phosphoenolpyruvic acid (PEP); acetyl-coenzyme A (Acetyl-CoA); 2-oxoglutarate (2-OG); 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP); phosphoribosyl pyrophosphate (PRPP); 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR); tetrahydrofolate (THF); tetrahydrofolic acid (THFA).

photorespiratory cycle in photosynthetic organisms [47]. Glycine yields were only marginally higher in PT cultures than HT and MT cultures during mid-log phase, and its content did not change significantly between this phase and the initial stationary phase (Table 3). In photosynthesising cells, glycine is produced from the by-products of photorespiration by (*S*)-2-hydroxy-acid oxidase (GLO, EC 1.1.3.15) and glutamate:glyoxylate aminotransferase (EC 2.6.1.4). Two translated isozyme candidates were found for GLO, although they had high similarity with lactate dehydrogenase, and both were up-regulated in PT cultures. The aminotransferase could not be assigned with confidence to any sequence(s), as such enzymes are difficult to functionally assign through bioinformatics alone. Glycine is also synthesised via a secondary route through the conversion of threonine by threonine aldolase (EC 4.1.2.5/48). Two isozymes for this enzyme were found, of which one was absent in PT cultures, and the other remained unchanged through all cultivations. The lack of a significant change in glycine content during the two growth phases indicates that glycine probably does not accumulate, even though its synthesis is up-regulated in photosynthetic cells. This may be because it is immediately oxidised to serine by the glycine decarboxylase-serine hydroxymethyltransferase system (GDC-SHMT), similar to that in higher plants [47].

Serine biosynthesis via GDC-SHMT system is the major route of serine formation in higher plants under light conditions [48], and it may be the same in *E. gracilis* var. *saccharophila*, as GDC (EC 2.1.2.10, 1.4.4.2 and 1.8.1.4) was up-regulated under the PT condition (Table 2). Three isozymes of SHMT (EC 2.1.2.1) were found, of which one was up-regulated under the PT condition (probably plastidic [49]), and another was up-regulated for the HT and MT conditions. The expression level of the other SHMT isoenzyme was unchanged across the three growth conditions. The overall upregulation of the GDC-SHMT system explains why serine content during PT cultivation was considerably higher than MT and HT cultivations (Table 3). Under dark conditions, and in non-photosynthetic tissues of higher plants, serine biosynthesis is proposed to occur from glycerate-3P via either phosphorylated or non-phosphorylated intermediates [48,50]. In *E. gracilis* var. *saccharophila* all three enzymes for this pathway are present under both light and dark conditions. Three isozymes were found for the first enzyme of the pathway, 3-phosphoglycerate dehydrogenase (EC 1.1.1.95), of which two were up-regulated under HT cultivation, while the third isozyme was unique to PT cultures. The remaining two enzymes of this pathway were unchanged. From this, it can be assumed that in *E. gracilis*, under light condition, serine biosynthesis occurs mainly via the GDC-SHMT system, producing large quantities of serine. In contrast, under dark conditions, serine biosynthesis occurs via glycerate-3P catabolism from

glycolysis, producing smaller quantities of serine. Catabolism of serine to pyruvate by serine ammonia lyase (EC 4.3.1.17) was also up-regulated in PT cultivation, but it appears that serine biosynthesis occurs more rapidly than its degradation in the presence of light.

Threonine biosynthesis from aspartate via homoserine involves five enzymes, which were unchanged across the conditions, except aspartate-semialdehyde dehydrogenase (EC 1.2.1.11), which was up-regulated under the HT cultivation (Table 2). However, the FAA analysis showed about 1.5-fold more threonine during PT cultivation than the MT and HT conditions (Table 3). This observation could mean that under HT and MT cultivations, threonine is broken down relatively fast. While the conversion of threonine to glycine, described above, may have played a small role in this, the major catabolism was possibly carried out by threonine ammonia-lyase (EC 4.3.1.19), to form 2-oxobutyrate in isoleucine biosynthesis as described below (Section 3.6.3).

3.6.3. Valine, isoleucine and leucine

The biosynthesis of valine and isoleucine are carried out by the same set of enzymes (Fig. 5). The control of biosynthesis of these amino acids is thus substrate quantity and substrate affinity-based. As described above, the conversion of threonine to 2-oxobutyrate plays an important role. Two isozymes of threonine ammonia-lyase (EC 4.3.1.19) were found, one of which matches to a bifunctional enzyme with additional serine ammonia-lyase activity (EC 4.3.1.17). The monofunctional isozyme was up-regulated under both HT and MT conditions during the mid-log phase, compared to PT condition (Table 2), which implies a higher rate of catabolism of threonine to 2-oxobutyrate during those cultivations. Moreover, there was already a higher amount of isoleucine per gram of dry mass during PT cultivation at mid-log phase, which may have caused a negative feedback on threonine ammonia-lyase [51], slowing down further catabolism of threonine in PT cultures.

The FAA analysis shows that during mid-log phase, a significant amount of valine was present in the cells under all cultivation conditions but by the beginning of the stationary phase, valine was notably lower for HT and MT cultivations (Table 3, 48 h). This can be explained through the expression of acetolactate synthase (EC 2.2.1.6), which plays a key role in controlling cell growth through valine biosynthesis. All three isoforms of this enzyme were up-regulated under either HT or MT or both cultivations (Table 2). This enzyme has two substrates: pyruvate and 2-oxobutyrate, and it has a higher affinity for the latter. During log-phase, when more pyruvate is present in the cells, valine biosynthesis via pyruvate continues normally, resulting in cell growth [51,52]. As threonine conversion to 2-oxobutyrate increases, the latter builds up in the cells, and more acetolactate synthase would be dedicated to converting this toxic compound to isoleucine [51,52]. In the late log phase, this is expected to hinder valine biosynthesis, augment isoleucine biosynthesis and slow down cell growth, triggering the stationary phase of growth. This is also supported by the FAA analysis, which shows a substantial rise in isoleucine during the start of the stationary phase, compared to the log phase for both HT and MT cultivations (Table 3). However, the FAA analysis also shows that the threonine and valine decrease, and isoleucine increase in the PT cultures were less prominent, possibly as a result of the gradual transition to the stationary phase for PT cultures, compared to the other two cultivations. Thus, the balance between valine and isoleucine within cells may be involved in maintaining cell growth and controlling growth phases.

The common biosynthetic pathway for leucine and valine uses the same enzymes for the initial steps (Fig. 5), with pyruvate as the primary substrate. As such, the 2-oxobutyrate inhibition of valine biosynthesis also affects leucine biosynthesis. This was also reflected in the FAA data (Table 3), which indicated a 10.5-fold and 11.25-fold decrease in leucine between mid-log phase (24 h) and stationary phase (48 h) for HT and MT cultivations respectively, compared to PT cultivation, which only showed a 1.29-fold decrease. The synthesis of leucine branches from the valine pathway at 2-oxoisovalerate, beyond which an additional three enzymes are required to complete its biosynthesis (Fig. 5). Probable candidates for these steps, including multifunctional ones, are listed in Table 2.

3.6.4. Methionine and cysteine

Two isozymes of methionine synthase (EC 2.1.1.13) were identified at the proteome level, one of which was uniquely expressed in HT and MT cultivations, while the second one was expressed in PT and MT cultivations only (Table 2). The FAA analysis showed that PT cultivation had a 6.5-fold and 4.73-fold higher methionine content than HT and MT cultivations respectively (Table 3). This could mean that although HT and MT cultivations are synthesising methionine at a faster rate, they are catabolising methionine more rapidly as well. Evidence of this could be found through the expression of methionine adenosyltransferase (EC 2.5.1.6), which converts methionine to *S*-adenosylmethionine, a compound that plays a key role in *trans*-methylation of many different substrates, *trans*-sulphuration and polyamine synthesis [53]. This enzyme was highly up-regulated under HT and MT cultivations, and may play a major role in methionine catabolism (Table 2). Apart from the three isozymes of EC 2.1.1.13, *E. gracilis* Z is reported to have a cobalamin-independent methionine synthase (EC 2.1.1.14) in the cytosol [54], but no sequence assigned to this enzyme could be deciphered from the transcriptome of *E. gracilis* var. *saccharophila* at this point.

In plants and bacteria, the synthesis of cysteine is catalysed by two enzymes: serine *O*-acetyltransferase (SAT) and cysteine synthase (CysS) [55]. Two isozymes of SAT (EC 2.3.1.30) and seven isozymes of CysS (EC 2.5.1.47) were translated from the transcriptome, which showed a variety of expression patterns (Table 2). For all cultivations, cysteine increased slightly by the start of the stationary phase, especially for the MT cells (48 h), in which the cysteine content doubled (Table 3). An interesting observation was that MT cells initially turned yellow when glucose was added to the medium, after which they greened over the next 48 h (data not shown). Cysteine has previously been reported to encourage green sector development and enhance pigment accretion in variegated plants exposed to light [56]. The excessive accumulation of cysteine in MT cells may thus be related to chloroplast development during transition from log to stationary growth phase. The less prominent increase of cysteine for all other cultivations follow the general trend reported previously for mid-log to end-log growth phase of *E. gracilis* [12,57].

3.6.5. Aspartate, asparagine, glutamate, glutamine and alanine

The FAA analysis shows that in PT cultures aspartate yields remain about the same through mid-log and beginning of stationary phases, while aspartate increases 1.3-fold and 1.5-fold for HT and MT cultures, respectively (Table 3). Aspartate is synthesised via the transfer of an amino group from glutamate to oxaloacetate by aspartate transaminase (EC 2.6.1.1) [43]. As mentioned before, aminotransferases are difficult to assign by amino acid sequence alone mainly because of their broad specificities. Asparagine is synthesised from aspartate by using either NH_3 (aspartate-ammonia ligase, EC 6.3.1.1) or glutamine (asparagine synthase, EC 6.3.5.4) as the amino group donor [43]. Although no candidate sequence could be found for the former, which usually occurs in prokaryotes [58], two candidates were found for the latter, with one translated in the proteome. This enzyme was unchanged across all cultivations (Table 2), but asparagine levels were higher for MT and HT cultures, and decreased drastically between the mid-log and initial stationary phases for all cultivations (Table 3). This catabolism could have been caused by the two putative asparaginases (EC 3.5.1.1) identified, one of which was expressed under light cultivation only (Table 2).

On the other hand, FAA analysis showed that glutamate does not change dramatically between the two phases of growth and among growth conditions (Table 3). Glutamate seems to be synthesised in *E.*

gracilis from 2-oxoglutarate and glutamine (glutamate synthase, EC 1.4.1.13/14) or NH_3 (glutamate dehydrogenase, EC 1.4.1.2/3/4), while glutamine is synthesised from glutamate and NH_3 (glutamate-ammonia ligase, EC 6.3.1.2). This latter enzyme was up-regulated in PT cultures (Table 2) as was the glutamine content from the FAA analysis at mid-log phase (Table 3). This event is similar to that in higher plants and other microalgae where exposure to light for a sufficient time stimulated the activity of this enzyme and it is dependent on photosynthesis [59,60]. Most photosynthetic organisms have two isoforms of this enzyme [43,61], located in the cytosol and chloroplast. However, only one isoform was identified in the current study.

Aspartate, asparagine, glutamate and glutamine are required for the distribution of nitrogen in all other pathways, and synthesis of other amino acids. Additionally, aspartate plays an important role in the malate-aspartate shuttle [62] and glutamate in the synthesis of δ-aminolevulinic acid (precursor for porphyrin and chlorophyll) [63], and possibly in the alternative TCA cycle in *E. gracilis*. The complex involvement of these amino acids in numerous cellular metabolic reactions means that further investigation is critical to elucidate their fate in *E. gracilis* and understand the discrepancies between the FAA content and synthase expression patterns.

Alanine is synthesised either from aspartate by aspartate-4-decarboxylase (EC 4.1.1.12) or from pyruvate by alanine transaminase (EC 2.6.1.2). The decarboxylase enzyme was not found in *E. gracilis*, while two transaminases could only be weakly assigned, as they are similar to other aminotransferases, especially EC 2.6.1.1. A putative sequence for alanine dehydrogenase (EC 1.4.1.1) was also found slightly up-regulated during PT cultivation (Table 2), which may explain the higher alanine content of HT and MT cultivations.

3.6.6. Arginine and proline

Arginine, the most abundant amino acid from all time points and different growth conditions (Table 3), is the nitrogen reserve equivalent to the paramylon carbon reserve, in *E. gracilis* [64]. Unlike most eukaryotes, *E. gracilis* does not appear to possess the set of enzymes to convert glutamate to ornithine for arginine synthesis via the urea cycle. Instead it utilises the enzymes to convert glutamine and NH_3 to carbamoyl phosphate (carbamoyl-phosphate synthase, EC 6.3.5.5 and EC 6.3.4.16), which is a substrate, along with ornithine, for citrulline formation by ornithine carbamoyltransferase (EC 2.1.3.3). Citrulline is then converted to argininosuccinate (argininosuccinate synthase, EC 6.3.4.5), which is sequentially converted to arginine by argininosuccinate lyase (EC 4.3.2.1). As previously reported in other strains of *E. gracilis*, our study also confirms the incomplete urea cycle in *E. gracilis* (Fig. 5), with the missing arginase and glycine amidinotransferase enzymes to convert arginine back to ornithine [64]. *E. gracilis* instead relies on the arginine dihydrolase pathway for arginine metabolism to ornithine [65], which requires arginine deiminase (EC 3.5.3.6) and citrullinase (EC 3.5.1.20). A candidate for the deiminase, which was unchanged across all cultivations during mid-log phase, was identified. Citrullinase has only been characterised in a few organisms, and currently no sequence is available on UniProt. It is thus not possible to identify this enzyme in *E. gracilis* at this point through amino acid sequence matching alone. The FAA analysis shows that PT cultures accumulated more arginine (Table 3), possibly owing to more nitrogen dedicated to cellular metabolism and growth of the rapidly dividing HT and MT cells. However, most of the enzymes involved in arginine biosynthesis were slightly up-regulated in HT and MT cultivations compared to PT (Table 2). This finding could indicate that the cells are compensating for the lower levels of nitrogen reserve during these cultivations. Further investigation needs to be carried out to determine if the arginine dihydrolase pathway up-regulation is responsible for the decline in arginine reserve at the start of the stationary phase, to provide nitrogen for metabolism.

For proline biosynthesis in *E. gracilis*, based on sequence homology, there seem to be two pathways, analogous to those in plants [66]. The main plant pathway uses glutamate-5-kinase (EC 2.7.2.11) and glutamate-5-semialdehyde dehydrogenase (EC 1.2.1.41) to convert glutamate to glutamate-5-semilaldehyde [66]. Four bifunctional isozymes were found to catalyse these reactions, one of which was slightly up-regulated under MT cultivation (Table 2). Proline is also synthesised from ornithine, which is transaminated to glutamate-5-semialdehyde by ornithine aminotransferase (EC 2.6.1.13) [66]. A single enzyme was identified for this reaction, which was slightly up-regulated under HT and MT cultivations (Table 2). FAA analysis showed that proline content did not vary much across all growth conditions (Table 3), and proline metabolism to glutamate (proline dehydrogenase, EC 1.5.5.2 and L-glutamate γ-semialdehyde dehydrogenase, EC 1.2.1.88) also remained unchanged across all cultivations (Table 2).

3.6.7. Lysine and histidine

Lysine biosynthesis in *E. gracilis* takes place through the α-aminoadipate pathway, which is unique to higher fungi, euglenoids, some archaea and *Thermus* spp. [67–70]. The enzymes for this pathway were difficult to annotate as their amino acid sequences overlap with enzymes from other pathways, and some of the enzymes are bifunctional and multifunctional, similar to those in fungi [69,71]. Nevertheless, the possible candidates are shown in Table 2. As in fungi, homocitrate synthase (EC 2.3.3.14), homoaconitate hydratase (EC 4.2.1.36) and homoisocitrate dehydrogenase (EC 1.1.1.87) had high sequence similarity to isopropylmalate and citramalate synthase, isopropylmalate dehydratase, and isocitrate and isopropylmalate dehydrogenase, respectively [71]. Most of the enzymes of this pathway were up-regulated under HT and MT cultivations during mid-log phase, which corresponds well with the lysine content from the FAA analysis (Table 3). At the initial stationary phase, however, lysine content of PT cultivations increased, while that in HT and MT cultivations decreased. Lysine is known to have at least nine catabolic fates [71], which need to be investigated further in *E. gracilis* to determine the causes of these changes at the start of the stationary phase.

Histidine biosynthesis in *E. gracilis* appears to be incomplete from this study, with four of the nine enzymes missing [72]. No sequences corresponding to histidinol phosphatase (EC 3.1.3.15) and the bifunctional phosphoribosyl-ATP diphosphatase (EC 3.6.1.31)/phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) were found. One of the isozymes for histidinol dehydrogenase (EC 1.1.1.23) shows some sequence similarity to the bifunctional enzyme (Supplementary Table S3). This isoenzyme, along with phosphoribosylformimino-AICAR-P isomerase (ProFAR-I, EC 5.3.1.16), and imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19) were, however, not detected at the proteomic level. With such little information and sequence similarity, very few of the enzymes for this pathway could be annotated. From the FAA analysis, PT cultivations had 1.97-fold and 2.44-fold higher histidine yields than HT and MT cultivations respectively at mid-log phase, with overall decreased amounts across all cultivations by the start of the stationary phase (Table 3). This finding is indicative of *E. gracilis* undergoing a higher level of oxidative stress during growth under the PT cultivations, as a key role of histidine is to inhibit oxidative stress [73], and reinforces the higher ascorbate and α-tocopherol yields in cells also grown under PT condition.

4. Conclusion

In conclusion, from the 32,128 proteins predicted by the transcriptome [20] and 1482 proteins acquired from NCBI, a combined total of 3843 proteins were detected across all three cultivation conditions (PT, MT and HT) in this strain of *E. gracilis*. The vast difference in the number of proteins between the transcriptomic report [20] and the current proteomic profiles suggests that this organism proactively synthesises mRNA and selectively translates the required mRNA during response to biotic and abiotic stressors. The proteins identified in this study help to establish some of the metabolic pathways employed by *E.*

gracilis var. *saccharophila* and clarify the effect of cultivation on their expression patterns. The biosynthesis of serine is affected by the synthesis route taken during different growth conditions. Under HT cultivation, *E. gracilis* var. *saccharophila* appears to utilise glycerate-3P from glycolysis for serine biosynthesis, while under PT and MT cultivation, it mainly employs the enzymes of the GDC-SHMT system, which helps it to accumulate more serine. Three isozymes for acetolactate synthase were identified, which play a key role in regulating cell growth through a balance between valine and isoleucine synthesis. There was a 2.5-fold increase in cysteine yield during MT cultivation of this strain of *E. gracilis*, which may be related to slower chloroplast development in the presence of glucose in the growth medium. Two pathways for proline biosynthesis were identified, with one predominant pathway similar to plants. Only some of the key isoforms from the large pool of isoenzymes predicted in the transcriptome [20] were translated. For instance, of the six ribose-phosphate diphosphokinase (histidine biosynthesis) in the transcriptome, only two were detected from this study. Further work on enzyme purification and activity assays can clarify the data presented in this paper. To the best of our knowledge, this is the first proteomic study of *E. gracilis*, to comprehensively characterise the biosynthetic pathways of amino acids and antioxidants produced by the strain *E. gracilis* var. *saccharophila*. There remains some ambiguities and missing enzymes, such as those for ascorbate and aminotransferases, and histidine biosynthesis respectively. Nevertheless, our data provides a blueprint for further studies to elucidate metabolic pathways and shed light into possible strain engineering for yield improvement.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.algal.2017.08.029.

Conflict of interest

The authors declare that they have no conflict of interest.

Contribution of authors

Conception and design: MTH, AS[1], HN, JT, GH, AS[2]. Collection and assembly of data: MTH, AS[1]. Analysis and interpretation of data: MTH, AS[1], MM, HN. Statistical expertise: MM. Drafting of the article: MTH, AS[1], HN. Critical revision of the article for important intellectual content: HN, MM, JT, AS[2], GH. Obtaining of funding: HN, JT, GH.

Note: AS[1] refers to Angela Sun, and AS[2] refers to Anwar Sunna.

Acknowledgements

This work was supported by the Australian Research Council's Industrial Transformation Training Centre funding scheme (Project Number: IC130100009). The authors would like to thank Dr. Dana Pascovici for her help with bioinformatics, and acknowledge the contribution of APAF in facilitating this research, using infrastructure provided by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).

References

[1] J.S. Rodríguez-Zavala, J.D. García-García, M.A. Ortiz-Cruz, R. Moreno-Sánchez, Molecular mechanisms of resistance to heavy metals in the protest *Euglena gracilis*, J. Environ. Sci. Health A Tox. Hazard. Subst. Environ. Eng. 42 (10) (2007) 1365–1378.

[2] M.M. Olaveson, C. Nalewajko, Effects of acidity on the growth of two *Euglena* species, Hydrobiologia 433 (1) (2000) 39–56.

[3] H. Hayashi, I. Narumi, S. Wada, M. Kikuchi, M. Furuta, K. Uehara, H. Watanabe, Light dependency of resistance to ionizing radiation in *Euglena gracilis*, J. Plant Physiol. 161 (10) (2004) 1101–1106.

[4] M. Li, X. Gao, B. Wu, X. Qian, J.P. Giesy, Y. Cui, Microalga *Euglena* as a bioindicator for testing genotoxic potentials of organic pollutants in Taihu Lake, China, Ecotoxicology 23 (4) (2014) 633–640.

[5] D.M. Mahapatra, H.N. Chanakya, T.V. Ramachandra, *Euglena* sp. as a suitable source of lipids for potential use as biofuel and sustainable wastewater treatment, J. Appl. Phycol. 25 (3) (2013) 855–865.

[6] P. Grimm, J.M. Risse, D. Cholewa, J.M. Müller, U. Beshay, K. Friehs, E. Flaschel, Applicability of *Euglena gracilis* for biorefineries demonstrated by the production of α-tocopherol and paramylon followed by anaerobic digestion, J. Biotechnol. 215 (2015) 72–79.

[7] B.B. Anderson, Investigations into the *Euglena* method for the assay of the vitamin B_{12} in serum, J. Clin. Pathol. 17 (1) (1964) 14–26.

[8] D.P. Häder, Phototaxis and gravitaxis in *Euglena gracilis*, in: F. Lenci, F. Ghetti, G. Colombetti, D.P. Häder, P.S. Song (Eds.), Biophysics of Photoreceptors and Photomovements in Microorganisms, 211 Springer US, Boston, 1991, pp. 203–221.

[9] D.P. Häder, Polarotaxis, gravitaxis and vertical phototaxis in the green flagellate, *Euglena gracilis*, Arch. Microbiol. 147 (2) (1987) 179–183.

[10] R. Brayner, T. Coradin, P. Beaunier, J.M. Grenèche, C. Djediat, C. Yéprémian, A. Couté, F. Fiévet, Intracellular biosynthesis of superparamagnetic 2-lines ferrihydrite nanoparticles using *Euglena gracilis* microalgae, Colloids Surf. B: Biointerfaces 93 (2012) 20–23.

[11] M. Shibakami, G. Tsubouchi, M. Sohma, M. Hayashi, Synthesis of nanofiber-formable carboxymethylated *Euglena*-derived β-1,3-glucan, Carbohydr. Polym. 152 (2016) 468–478.

[12] J.S. Rodríguez-Zavala, M.A. Ortiz-Cruz, G. Mendoza-Hernández, R. Moreno-Sánchez, Increased synthesis of α-tocopherol, paramylon and tyrosine by *Euglena gracilis* under conditions of high biomass production, J. Appl. Microbiol. 109 (6) (2010) 2160–2172.

[13] E.C. O'Neill, M. Trick, B. Henrissat, R.A. Field, *Euglena* in time: evolution, control of central metabolic processes and multi-domain proteins in carbohydrate and natural product biochemistry, Perspect. Sci. 6 (2015) 84–93.

[14] M. Valko, C.J. Rhodes, J. Moncol, M. Izakovic, M. Mazur, Free radicals, metals and antioxidants in oxidative stress-induced cancer, Chem. Biol. Interact. 160 (1) (2006) 1–40.

[15] G. Csányi, F.J. Miller Jr., Oxidative stress in cardiovascular disease, Int. J. Mol. Sci. 15 (4) (2014) 6002–6008.

[16] A. Nunomura, R.J. Castellani, X. Zhu, P.I. Moreira, G. Perry, M.A. Smith, Involvement of oxidative stress in Alzheimer disease, J. Neuropathol. Exp. Neurol. 65 (7) (2006) 631–641.

[17] F. Ivušić, B. Šantek, Optimization of complex medium composition for heterotrophic cultivation of *Euglena gracilis* and paramylon production, Bioprocess Biosyst. Eng. 38 (6) (2015) 1103–1112.

[18] J.C. Ogbonna, H. Tanaka, Cyclic autotrophic/heterotrophic cultivation of photosynthetic cells: a method of achieving continuous cell growth under light/dark cycles, Bioresour. Technol. 65 (1–2) (1998) 65–72.

[19] T.E. Ebenezer, M. Carrington, M. Lebert, S. Kelly, M.C. Field, *Euglena gracilis* genome and transcriptome: organelles, nuclear genome assembly strategies and initial features, in: S.D. Schwartzbach, S. Shigeoka (Eds.), Euglena: Biochemistry, Cell and Molecular Biology, Adv. Exp. Med. Biol. vol. 979, Springer International Publishing, Cham, 2017, pp. 125–140.

[20] E.C. O'Neill, M. Trick, L. Hill, M. Rejzek, R.G. Dusi, C.J. Hamilton, P.V. Zimba, B. Henrissat, R.A. Field, The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry, Mol. BioSyst. 11 (10) (2015) 2808–2820.

[21] S.A. Cohen, Amino acid analysis using precolumn derivatization with 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate, in: C. Cooper, N. Packer, K. Williams (Eds.), Amino Acid Anal Protoc. Methods in Molecular Biology™, 159 Humana Press, New York, 2000, pp. 39–47.

[22] Y. Wu, M. Mirzaei, D. Pascovici, J.M. Chick, B.J. Atwell, P.A. Haynes, Quantitative proteomic analysis of two different rice varieties reveals that drought tolerance is correlated with reduced abundance of photosynthetic machinery and increased abundance of ClpD1 protease, J. Proteome 143 (2016) 73–82.

[23] E.L. Thompson, W. O'Connor, L. Parker, P. Ross, D.A. Raftos, Differential proteomic responses of selectively bred and wild-type Sydney rock oyster populations exposed to elevated CO_2, Mol. Ecol. 24 (6) (2015) 1248–1262.

[24] K.A. Neilson, I.S. George, S.J. Emery, S. Muralidharan, M. Mirzaei, P.A. Haynes, Analysis of rice proteins using SDS-PAGE shotgun proteomics, Methods Mol. Biol. 1072 (2014) 289–302.

[25] S. Shigeoka, Y. Nakano, S. Kitaoka, Occurrence of L-ascorbic acid in *Euglena gracilis* Z, Bull. Univ. Osaka Pref. Ser. B 32 (1980) 43–48.

[26] H. Takeyama, A. Kanamaru, Y. Yoshino, H. Kakuta, Y. Kawamura, T. Matsunaga, Production of antioxidant vitamins, beta-carotene, vitamin C, and vitamin E, by two-step culture of *Euglena gracilis* Z, Biotechnol. Bioeng. 53 (2) (1997) 185–190.

[27] G. Wheeler, T. Ishikawa, V. Pornsaksit, N. Smirnoff, Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes, elife 4 (2015) e06369.

[28] S. Shigeoka, Y. Nakano, S. Kitaoka, The biosynthetic pathway of L-ascorbic acid in *Euglena gracilis* Z, J. Nutr. Sci. Vitaminol. 25 (4) (1979) 299–307.

[29] T. Ishikawa, H. Nishikawa, Y. Gao, Y. Sawa, H. Shibata, Y. Yabuta, T. Maruta, S. Shigeoka, The pathway via D-galacturonate/L-galactonate is significant for ascorbate biosynthesis in *Euglena gracilis*: identification and functional characterization of aldonolactonase, J. Biol. Chem. 283 (45) (2008) 31133–31141.

[30] T. Ishikawa, I. Masumoto, N. Iwasa, H. Nishikawa, Y. Sawa, H. Shibata, A. Nakamura, Y. Yabuta, S. Shigeoka, Functional characterization of D-galacturonic acid reductase, a key enzyme of the ascorbate biosynthesis pathway, from *Euglena gracilis*, Biosci. Biotechnol. Biochem. 70 (11) (2006) 2720–2726.

[31] T. Ishikawa, S. Shigeoka, Recent advances in ascorbate biosynthesis and the physiological significance of ascorbate peroxidase in photosynthesizing organisms, Biosci. Biotechnol. Biochem. 72 (5) (2008) 1143–1154.

[32] D.B. Zorov, M. Juhaszova, S.J. Sollott, Mitochondrial reactive oxygen species (ROS) and ROS-induced ROS release, Physiol. Rev. 94 (3) (2014) 909–950.

[33] P. Sharma, A.B. Jha, R.S. Dubey, M. Pessarakli, Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions, J. Bot. 2012 (2012) 217037.

[34] A. Görlach, E.Y. Dimova, A. Petry, A. Martínez-Ruiz, P. Hernansanz-Agustín, A.P. Rolo, C.M. Palmeira, T. Kietzmann, Reactive oxygen species, nutrition, hypoxia and diseases: problems solved? Redox Biol. 6 (2015) 372–385.

[35] T. Ishikawa, N. Tajima, H. Nishikawa, Y. Gao, M. Rapolu, H. Shibata, Y. Sawa, S. Shigeoka, *Euglena gracilis* ascorbate peroxidase forms an intramolecular dimeric structure: its unique molecular characterization, Biochem. J. 426 (2) (2010) 125–134.

[36] M. Zeng, W. Hao, Y. Zou, M. Shi, Y. Jiang, P. Xiao, A. Lei, Z. Hu, W. Zhang, L. Zhao, J. Wang, Fatty acid and metabolomic profiling approaches differentiate heterotrophic and mixotrophic culture conditions in a microalgal food supplement '*Euglena*', BMC Biotechnol. 16 (1) (2016) 49.

[37] S. Shigeoka, T. Onishi, Y. Nakano, S. Kitaoka, The contents and subcellular distribution of tocopherols in *Euglena gracilis*, Agric. Biol. Chem. 50 (4) (1986) 1063–1065.

[38] J.C. Ogbonna, S. Tomiyama, H. Tanaka, Production of α-tocopherol by sequential heterotrophic–photoautotrophic cultivation of *Euglena gracilis*, J. Biotechnol. 70 (1999) 213–221.

[39] D.R. Threlfall, T.W. Goodwin, Nature, intracellular distribution and formation of terpenoid quinones in *Euglena gracilis*, Biochem. J. 103 (2) (1967) 573–588.

[40] C. Kusmic, R. Barsacchi, L. Barsanti, P. Gualtieri, V. Passarelli, *Euglena gracilis* as source of the antioxidant vitamin E. Effects of culture conditions in the wild strain and in the natural mutant WZSL, J. Appl. Phycol. 10 (6) (1998) 555–559.

[41] E.C. Carballo-Cárdenas, P.M. Tuan, M. Janssen, R.H. Wijffels, Vitamin E (alpha-tocopherol) production by the marine microalgae *Dunaliella tertiolecta* and *Tetraselmis suecica* in batch cultivation, Biomol. Eng. 20 (4–6) (2003) 139–147.

[42] J.C. Ogbonna, Microbiological production of tocopherols: current state and prospects, Appl. Microbiol. Biotechnol. 84 (2) (2009) 217–225.

[43] M.A. Bromke, Amino acid biosynthesis pathways in diatoms, Meta 3 (2) (2013) 294–311.

[44] A. Schaller, M. van Afferden, V. Windhofer, S. Bülow, G. Abel, J. Schmid, N. Amrhein, Purification and characterization of chorismate synthase from *Euglena gracilis*: comparison with chorismate synthases of plant and microbial origin, Plant Physiol. 97 (4) (1991) 1271–1279.

[45] C.N. Hankins, S.E. Mills, A dimer of a single polypeptide chain catalyzes the terminal four reactions of the L-tryptophan pathway in *Euglena gracilis*, J. Biol. Chem. 252 (1) (1977) 235–239.

[46] E. Fernandez-Valiente, M. Rodriguez-Lopez, Action of indole-acetic acid and gibberellic acid on the paramylon synthesis in *Euglena gracilis*, Microbiol. Esp. 32-3 (1970) 11–20.

[47] J.M. Mouillon, S. Aubert, J. Bourguignon, E. Gout, R. Douce, F. Rébeillé, Glycine and serine catabolism in non-photosynthetic higher plant cells: their role in C1 metabolism, Plant J. 20 (2) (1999) 197–205.

[48] C.L. Ho, M. Noji, M. Saito, K. Saito, Regulation of serine biosynthesis in *Arabidopsis*. Crucial role of plastidic 3-phosphoglycerate dehydrogenase in non-photosynthetic tissues, J. Biol. Chem. 274 (1) (1999) 397–402.

[49] M. Sakamoto, T. Masuda, Y. Yanagimoto, Y. Nakano, S. Kitaoka, Y. Tanigawa, Purification and characterization of serine hydroxymethyltransferase from mitochondria of *Euglena gracilis* Z, Biosci. Biotechnol. Biochem. 60 (12) (1996) 1941–1944.

[50] C.L. Ho, M. Noji, K. Saito, Plastidic pathway of serine biosynthesis. Molecular cloning and expression of 3-phosphoserine phosphatase from *Arabidopsis thaliana*, J. Biol. Chem. 274 (16) (1999) 11007–11012.

[51] Y. Oda, Y. Nakano, S. Kitaoka, Occurrence and some properties of two threonine dehydratases in *Euglena gracilis*, J. Gen. Microbiol. 129 (1) (1983) 57–61.

[52] Y. Oda, Y. Nakano, S. Kitaoka, Properties and regulation of valine-sensitive acetolactate synthase from mitochondria of *Euglena gracilis*, J. Gen. Microbiol. 128 (6) (1982) 1211–1216.

[53] C. García-Estrada, Y. Pérez-Pertejo, D. Ordóñez, R. Balaña-Fouce, R.M. Reguera, Analysis of genetic elements regulating the methionine adenosyltransferase gene in *Leishmania infantum*, Gene 389 (2) (2007) 163–173.

[54] Y. Isegawa, F. Watanabe, S. Kitaoka, Y. Nakano, Subcellular distribution of cobalamin-dependent methionine synthase in *Euglena gracilis* Z, Phytochemistry 35 (1994) 59–61.

[55] E.R. Bonner, R.E. Cahoon, S.M. Knapke, J.M. Jez, Molecular basis of cysteine biosynthesis in plants: structural and functional analysis of O-acetylserine sulfhydrylase from *Arabidopsis thaliana*, J. Biol. Chem. 280 (46) (2005) 38803–38813.

[56] S. Rodermel, Arabidopsis variegation mutants, The Arabidopsis Book, vol. 1, American Society of Plant Biologists, 2002, p. e0079.

[57] O. Coppellotti, Glutathione, cysteine and acid-soluble thiol levels in *Euglena gracilis* cells exposed to copper and cadmium, Comp. Biochem. Physiol. 94C (1) (1989) 35–40.

[58] L. Gaufichon, M. Reisdorf-Cren, S.J. Rothstein, F. Chardon, A. Suzuki, Biological functions of asparagine synthetase in plants, Plant Sci. 179 (3) (2010) 141–153.

[59] R. Tischner, A. Hüttermann, Regulation of glutamine synthetase by light and during nitrogen deficiency in synchronous *Chlorella sorokiniana*, Plant Physiol. 66 (5) (1980) 805–808.

[60] S. Ghosh, Role of light in the appearance of glutamine synthetase in leaves of *Pennisetum glaucum*, J. Plant Biochem. Biotechnol. 19 (2) (2010) 239–242.

[61] S.F. McNally, B. Hirel, P. Gadal, A.F. Mann, G.R. Stewart, Glutamine synthetases of higher plants. Evidence for a specific isoform content related to their possible physiological role and their compartmentation within the leaf, Plant Physiol. 72 (1) (1983) 22–25.

[62] R. Jasso-Chávez, R. Moreno-Sánchez, Cytosol-mitochondria transfer of reducing equivalents by a lactate shuttle in heterotrophic *Euglena*, Eur. J. Biochem. 270 (24) (2003) 4942–4951.

[63] S.M. Mayer, S.I. Beale, δ-Aminolevulinic acid biosynthesis from glutamate in *Euglena gracilis*. Photocontrol of enzyme levels in a chlorophyll-free mutant, Plant Physiol. 97 (1991) 1094–1102.

[64] B.-S. Park, A. Hirotani, Y. Nakano, S. Kitaoka, The physiological role and catabolism of arginine in *Euglena gracilis*, Agric. Biol. Chem. 47 (1983) 2561–2567.

[65] B.-S. Park, A. Hirotani, Y. Nakano, S. Kitaoka, Purification and some properties of arginine deiminase in *Euglena gracilis* Z, Agric. Biol. Chem. 48 (1984) 483–489.

[66] L. Szabados, A. Savouré, Proline: a multifunctional amino acid, Trends Plant Sci. 15 (2) (2010) 89–97.

[67] H.J. Vogel, Distribution of lysine pathways among fungi: evolutionary implications, Am. Nat. 98 (1964) 435–446.

[68] H.J. Vogel, Lysine biosynthesis in *Chlorella* and *Euglena*: phylogenetic significance, Biochim. Biophys. Acta 34 (1959) 282–283.

[69] H. Nishida, M. Nishiyama, What is characteristic of fungal lysine synthesis through the alpha-aminoadipate pathway? J. Mol. Evol. 51 (3) (2000) 299–302.

[70] T. Kosuge, T. Hoshino, Lysine is synthesized through the α-aminoadipate pathway in *Thermus thermophilus*, FEMS Microbiol. Lett. 169 (1998) 361–367.

[71] T.M. Zabriskie, M.D. Jackson, Lysine biosynthesis and metabolism in fungi, Nat. Prod. Rep. 17 (1) (2000) 85–97.

[72] A. Stepansky, T. Leustek, Histidine biosynthesis in plants, Amino Acids 30 (2) (2006) 127–142.

[73] J. Lemire, Y. Milandu, C. Auger, A. Bignucolo, V.P. Appanna, V.D. Appanna, Histidine is a source of the antioxidant, alpha-ketoglutarate, in *Pseudomonas fluorescens* challenged by oxidative stress, FEMS Microbiol. Lett. 309 (2) (2010) 170–177.

CHAPTER 4

Comparative proteomics investigation of central carbon metabolism in *Euglena gracilis* grown under photoautotrophic, mixotrophic and heterotrophic cultivations

4.1 Introduction

Euglena gracilis has the capacity to use numerous carbon sources owing to its metabolic versatility. Central carbon metabolism is defined as those enzymatic pathways that are used to convert organic and inorganic carbon sources into precursors of other metabolic pathways. Although several enzymes involved in the central carbon metabolic pathways have been individually identified in *E. gracilis*, the overall relative abundance of these enzymes under different growth conditions has not been described. Moreover, enzymes from many of the reactions in the central carbon metabolic pathways have not been identified. In Chapter 4, label-free shotgun proteomics is used to investigate the effect of photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivations on central carbon metabolism of *E. gracilis* Z. Some key findings include the lack of hexokinase for the first step of glycolysis, the presence of a glucokinase, the dominance of the oxidative pentose phosphate pathway in glucose metabolism during MT and HT cultivations and heterotrophic CO_2 fixation in the dark even when surplus glucose is present in the medium.

The results from the work are reported in the form of a manuscript that has been prepared for submission.

4.2 Contribution to Manuscript

This manuscript was developed under the supervision of Helena Nevalainen. The conception of this manuscript was perceived with my supervisors Helena Nevalainen, Anwar Sunna and Angela Sun, along with my adjunct supervisor Junior Te'o (Queensland University of Technology), and the industrial partner representative for this project Graham Hobba (Agritechnology Pty Ltd). The experiments were designed by me with input from Helena Nevalainen and Angela Sun. I performed all of the experimental work, while data analysis was carried out by me with assistance from Angela Sun and Mehdi Mirzaei. The manuscript was prepared by me and reviewed by Angela Sun, Junior Te'o, Graham Hobba, Anwar Sunna and Helena Nevalainen.

Table 4.1: Author contribution summary for Chapter 4 (Manuscript 1).

	MTH	AS[1]	MM	JT	GH	AS[2]	HN
Conception	•	•		•	•	•	•
Experimental design	•	•					•
Data collection	•						
Data analysis	•	•	•				
Manuscript	•	•		•	•	•	•

MTH = Mafruha Tasnin Hasan, AS[1] = Angela Sun, MM = Mehdi Mirzaei, JT = Junior Te'o, GH = Graham Hobba, AS[2] = Anwar Sunna, HN = Helena Nevalainen

4.3 Manuscript

Hasan MT, Sun A, Mirzaei M, Te'o J, Hobba G, Sunna A and Nevalainen H. Comparative proteomics investigation of central carbon metabolism in *Euglena gracilis* grown under photoautotrophic, mixotrophic and heterotrophic cultivations. Manuscript prepared for submission.

Comparative proteomics investigation of central carbon metabolism in *Euglena gracilis* grown under photoautotrophic, mixotrophic and heterotrophic cultivations

Mafruha T Hasan[1,2,3], Angela Sun[1,2,3], Mehdi Mirzaei[1,2,3,4,5], Junior Te'o[2,6], Graham Hobba[2,7], Anwar Sunna[1,2,3], Helena Nevalainen[1,2,3]

[1]Department of Molecular Sciences, Macquarie University, Sydney, Australia

[2]ARC-MQ Industrial Transformation Training Centre (ITTC) for Molecular Technology in the Food Industry, Sydney, Australia

[3]Biomolecular Discovery and Design Research Centre, Macquarie University, Sydney, Australia

[4]Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia

[5]Australian Proteome Analysis Facility, Macquarie University, Sydney, Australia

[6]School of Earth, Environmental and Biological Sciences, Queensland University of Technology, Brisbane, Australia

[7]Agritechnology Pty Ltd, Borenore, Australia

Corresponding Author:

Helena Nevalainen

Email: helena.nevalainen@mq.edu.au

Keywords:

carbon metabolism, glucokinase, post-translational regulation, *Euglena gracilis* proteomics

Abstract

Euglena gracilis can use a wide range of organic carbon sources, as well as CO_2 from the atmosphere. This metabolic versatility is owed to the genome of *E. gracilis* that can encode a wide range of enzymes. Many of these enzymes are regulated post-transcriptionally, allowing the cells to adapt even more quickly to changes in their surroundings. Here we investigated the effect of photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivation on central carbon metabolism in *E. gracilis* Z using label-free shotgun proteomics. Differential expression between isozymes was observed for several enzymes based on the cultivation condition. A hexokinase enzyme identified in the published transcriptome was not expressed in the proteome. Instead, a high-specificity glucokinase appeared to conduct the first step of glycolysis. Proteomic analysis also revealed that the oxidative pentose phosphate pathway plays a key role in glucose metabolism under MT and HT cultivation, while glycolysis was predominant under PT cultivation. Some enzymes of the Calvin pathway were expressed under HT cultivation indicating regulation at the post-translational level. Even in the presence of sufficient glucose in the medium, HT cells may be carrying out CO_2 fixation using a different route in lieu of the Calvin pathway. The carbon metabolic pathways investigated here in terms of proteomic changes provide new information, as well as validate data presented elsewhere, adding to the existing knowledge of metabolism in *E. gracilis*. Putative functional annotations of several proteins that were previously unidentified are also provided.

1 Introduction

The freshwater microalga *Euglena gracilis* is a metabolically versatile organism capable of photosynthesising, as well as using organic carbon sources from its surroundings. During both these modes of nutrient assimilation, carbon is transformed into reserve molecules

that can be broken down to provide ATP for growth and cell maintenance, and carbon skeletons for other essential molecules.

Central carbon metabolism refers to a series of complex enzymatic reactions to convert sugars into precursors of other metabolic reactions [1]. These precursors and their consequential reactions are then used to synthesise the entire cellular biomass [1]. Although enzymes from a few carbon metabolic pathways in *E. gracilis* have been identified, a large portion of them remain unidentified. *E. gracilis* has been reported to utilise the Embden-Meyerhof-Parnas (EMP) pathway (glycolysis) for glucose catabolism [2]. Most of the enzymes for this pathway have been reported, including hexokinase (HK) [3], fructose-bisphosphate aldolase (ALDO) [4], triosephosphate isomerase (TPI) [5], glyceraldehyde-3-phosphate dehydrogenase (GAPDH) [6], chloroplastic phosphoglycerate kinase (PGK) [7] and enolase (ENO) [8]. The oxidative decarboxylation of pyruvate to acetyl-CoA has also been described in *E. gracilis* [9-10], along with some enzymes for gluconeogenesis [11-13], an alternate tricarboxylic acid (TCA) cycle [14-15], the glyoxylate cycle [16] and the oxidative pentose phosphate (oxPP) pathway [2].

However, the transcriptome has revealed that some of the enzymes have several isoforms and it is unclear if their expression in the proteome is affected by the mode of nutrient assimilation. Since several studies have identified that *E. gracilis* does not always regulate protein expression at the transcriptional level [17-24], proteomic changes of these enzymes can give a better idea of the effect of mode of nutrient assimilation on carbon metabolism. A high-throughput proteomic study can reveal the expression levels of all these enzymes at once, and comparative proteomics can then allow evaluation of differential expression among the modes of nutrition. Given the complexity of carbon metabolism in microalgae, a comprehensive description of the interactions between pathways and the fate of all the carbon molecules is not always clear. Nevertheless, the central carbon metabolic pathways of *E. gracilis* have been described here using a similar high-throughput label-free shotgun proteomics approach as in

our recent publication [25]. The aim of this study was thus to acquire insight into the global regulation of carbon metabolism in response to mode of nutrition acquisition in *E. gracilis*.

2 Materials and Methods

2.1 Growth medium, cultivation conditions and dry mass determination

Euglena gracilis Z strain was obtained from Southern Biological Pty Ltd, Australia. The strain was maintained as a pure axenic culture using autoclaved modified Hutner medium as the basal medium (pH 3.5), supplemented with vitamins B_1 and B_{12}, 10 g l^{-1} of yeast extract, and 1.5% (v/v) D-(+)-glucose. Cultivation was carried out in biological triplicates under three growth conditions: photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT). The cultivation conditions have been defined before [25]. Briefly, PT and MT cultures were exposed to white light illumination (2000 lx) for a photoperiod of 14 h light/10 h darkness, while HT cultures were kept in the dark. Prior to inoculation, the MT and HT cultures were supplemented with D-(+)-glucose to a final concentration of 1.5% (w/v), while PT cultivations received no glucose supplementation. Thus, the PT cultures utilised mainly atmospheric CO_2 via photosynthesis, with some carbon from the amino acids present in YE. The MT cultures utilised CO_2 from the atmosphere as well as glucose and YE from the medium, while the HT cultures utilised only the organic carbon sources of glucose and YE from the medium. The volumes of cultures were adjusted accordingly to ensure the concentration of all other nutrients and inoculum were about the same for all cultures. An initial concentration of 1.5 x 10^8 cells of light-adapted and dark-adapted axenic stock cultures, maintained under the conditions described above for over two years, were used to inoculate the PT and MT cultures, and HT cultures respectively. The stock cultures were starved by not replenishing basal medium for

three days and glucose for 10 days prior to inoculation. One hundred and fifty ml of cultures were grown in 250 ml Erlenmeyer flasks at 23°C with orbital shaking at 150 rpm.

Dry mass was measured every 24 h using pre-weighed aluminium evaporating dishes. Harvested cells were dried in these dishes at 70°C for 48 h, and cooled to room temperature in a desiccator for 10 min before weighing. All chemicals and reagents were purchased from Sigma Aldrich, Australia unless stated otherwise.

2.2 Label-free shotgun proteomics

Proteomic analysis was carried out at mid-log phase (24 h for MT and HT cultures and 48 h for PT cultures) using the label-free shotgun proteomics approach described previously [25]. Briefly, the harvested cells were washed twice with MilliQ water and resuspended in three ml of cold PBS buffer supplemented with 0.04% (v/v) β-mercaptoethanol and protease inhibitor (cOmplete™ Protease Inhibitor Cocktail, Roche; two tablets per 100 ml buffer). The cell suspension was lysed using French press (Thermo Spectronic French Pressure Cell Press, MA, USA) at 15,000 psi and then centrifuged at 2000 *g* for five min. Proteins from 800 µl of the supernatant were precipitated by methanol:chloroform precipitation by adding 4:3:2 (v/v) methanol:MilliQ water:chloroform, vortexing briefly and centrifuging at 8000 *g* for two min. The protein precipitate formed at the interphase was washed once with four volumes of methanol, air-dried and resolubilised in 5% (v/v) SDS. Protein concentration was measured by BCA assay (Thermo Scientific Pierce, Australia).

One hundred µg of proteins per replicate were fractionated in the reduced state by SDS-PAGE in precast 4-12% gradient NuPAGE® Bis-Tris Mini Gels at 200 V for 50 min in MOPS buffer. The gel was Coomassie-stained for visualisation [26] and each lane containing a replicate was cut into 16 portions of about the same size. Each portion was further sliced into smaller pieces and placed into a well of a 96-well plate before in-gel digestion. The gel pieces

were reduced with 10 mM DTT, alkylated with 55 mM iodoacetamide (IAA) and digested with 12.5 ng ml^{-1} trypsin overnight as described previously [26]. DTT, IAA and trypsin were prepared in 50 mM NH_4HCO_3 solutions [26]. Peptide extraction was carried out twice with 50%:2% (v/v) acetonitrile:formic acid (FA, Merck, Australia), dried in a vacuum centrifuge and reconstituted in 20 µl of 2% (v/v) FA [25]. The peptide extracts were purified using ZipTip® Pipette Tips with C_{18} resin (Merck Millipore, US), dried in a vacuum centrifuge, reconstituted in 20 µl of 2% (v/v) FA and centrifuged at 18,000 *g* for 15 min. The top 10 µl of the centrifuged peptides were then used for nanflow liquid chromatography tandem mass spectrometry (nanoLC-MS/MS).

NanoLC-MS/MS was carried out using a QExactive mass spectrometer (Thermo Scientific, CA, USA) and the spectra were analysed using the global proteome machine (GPM) software version 2.2.1 with the X!Tandem algorithm (http://www.thegpm.org) [27]. The *Euglena* non-redundant protein database [28] and the NCBI *Euglena* database were stitched together and used as the search database, along with common human and trypsin peptide contaminants. The *Euglena* non-redundant protein database from the John Innes Centre website (http://jicbio.nbi.ac.uk/euglena/) and the *Euglena* protein database from NCBI (http://www.ncbi.nlm.nih.gov/) were acquired in September 2016. Reverse database searching was also used for estimating false discovery rates (FDRs). The GPM search parameters were set as previously described [26]. The 16 fractions of each replicate were processed consecutively on the GPM with output files for each fraction, as well as a simple merged non-redundant output file for peptide and protein identifications with log(*e*) values less than -1.

2.3 Data processing, quality control, quantitative and statistical analyses

The low-quality hits from the GPM were filtered out to produce a high stringency dataset of reproducibly identified proteins present in each cultivation using the same criteria as

before [25]. Peptide and protein FDRs were calculated (FDR threshold was set to lower than 1%), followed by the removal of the reversed database hits and calculation of the Normalised Spectral Abundance Factor (NSAF) values [29]. Quantitative proteomic analysis and statistical analysis were carried out using the summed NSAF values as a measure of relative protein abundance [29]. A spectral fraction of 0.5 was added to all spectral counts to compensate for null values and enable log transformation for the statistical analysis. Two-sample unpaired *t*-tests and an analysis of variance (ANOVA) were performed on the log-transformed NSAF dataset, and proteins with a *p*-value less than 0.05 were considered to have a significant change in relative abundance among cultivations [25]. For the two-sample unpaired *t*-tests, proteins detected under the PT growth condition were used as the control against which proteins detected under the MT and HT growth conditions were compared.

2.4 Functional annotation

The resulting set of differentially expressed and unchanged proteins were functionally annotated using BLASTP. The corresponding sequences from the *Euglena* non-redundant peptide database were compared to the UniProtKB/Swiss-Prot database using BLASTP with an *E*-value cut-off of $1e^{-10}$. This similarity analysis showed that of the expressed proteins identified from the *Euglena* non-redundant protein database via nanoLC-MS/MS, about 30% had unknown function. The gene ontology (GO) information of the proteins that could be functionally annotated was retrieved from the UniProt database. This information was then summarised into 30 GO categories of interest for the up-regulated, down-regulated and unchanged proteins of *E. gracilis* Z, using proteins from the PT condition as the control.

3 Results and discussion

3.1 Growth pattern of E. gracilis Z

Glucose was used as the main organic carbon source for ease of growth and pathway analysis. The exponential growth phase of the MT culture of *E. gracilis* Z occurred between 0 and 48 h, while the exponential growth phase of the PT and HT cultures occurred between 0 and 72 h (Figure 1). Compared to the growth curve for *E. gracilis* var. *saccharophila* reported in our previous study [25], the PT cultures of *E. gracilis* Z had a higher accumulation of biomass, and unlike the *E. gracilis* var. *saccharophila* strain, *E. gracilis* Z maintained the stationary phase during PT cultivation even at 144 h. On the other hand, the MT cultures had an incomprehensible stationary phase with the growth declining beyond 48 h. Based on the growth patterns of the *E. gracilis* Z cultures, the mid-log phase was deduced to be around 24 h for MT and HT cultures, and around 48 h for PT cultures. The mid-log phase was selected for label-free shotgun proteomics analysis to detect changes in the core carbon metabolic pathways during carbon assimilation.

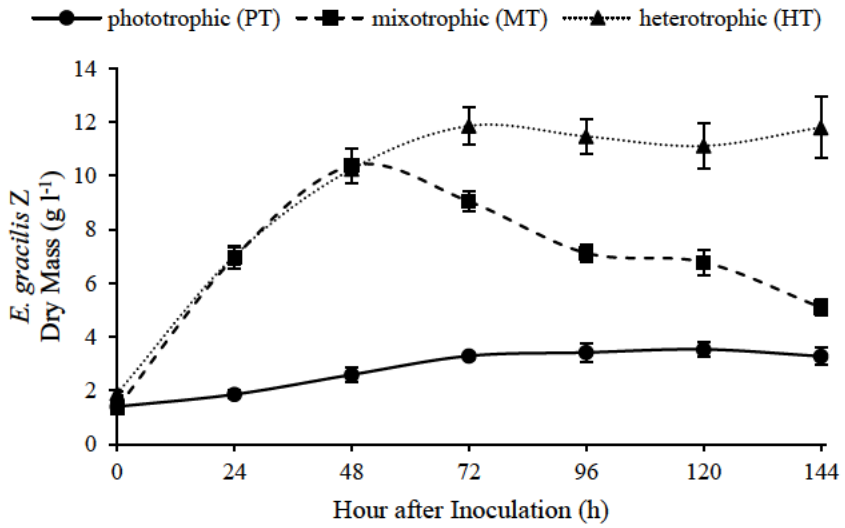

Figure 1: Growth pattern of *E. gracilis* Z under photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivations. The exponential growth phase occurred between 0 to 48 h for MT cultures, and up to 72 h for the PT and HT cultures. The fastest growth occurred at about 24 h for MT and HT cultures, and 48 h for PT cultures, and thus these were the selected time points for proteomic profiling.

3.2 Analysis of label-free shotgun proteomics

A total of 1564 non-redundant proteins were reproducibly identified across the three growth conditions of *E. gracilis* Z (Figure 2). The highest number of reproducibly identified proteins was found in the MT cultures (2868), while the PT and HT cultures showed fewer reproducible proteins (2662 and 2198 respectively). The highest number of unique proteins was also observed in MT cultures (501) compared to the PT (407) and HT (274) cultures.

The GPM data for all replicates are included in Supplementary Table S1. The redundant count of peptides, which is the total number of peptides found for proteins in a given sample, allowed for quantification of the data. The initial redundant count of peptides obtained from the nanoLC-MS/MS was consistent among the replicates of all treatments for *E. gracilis* Z, with a coefficient of variation of less than 10% (Supplementary Table S1). The FDR values for

proteins and peptides were less than or equal to 1% and 0.2% respectively, indicating that the data did not require any further filtering (Table 1).

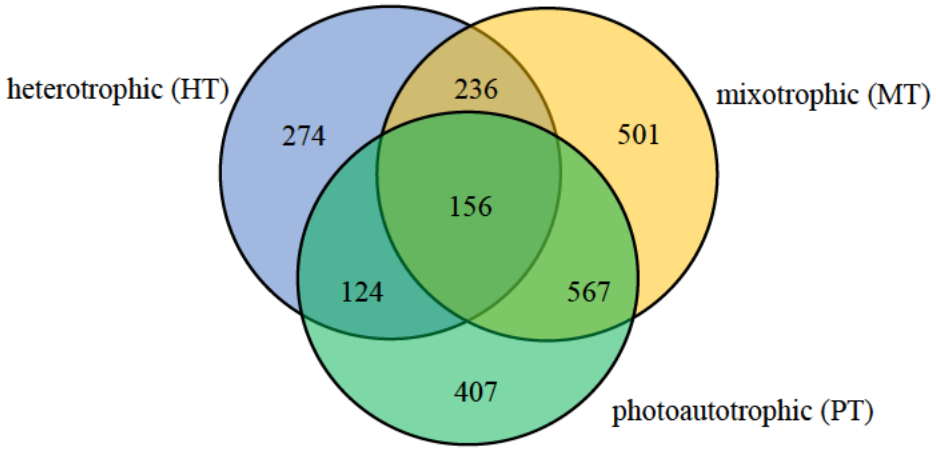

Figure 2: Venn diagram of the number of proteins common between and among cultures: photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT), in *E. gracilis* Z. MT cultures had the highest number of proteins (2868), while HT cultures had the lowest (2198). More proteins were shared between PT and MT cultures than any other combination (2131).

Table 1: Summary of proteins identified under photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivations of *E. gracilis* Z, and their corresponding FDR values.

Growth Condition[a]	No. of proteins[b]	Protein FDR[c] (%)	Peptide FDR[c] (%)
PT	2662	0.93	0.16
MT	2868	0.97	0.20
HT	2198	0.45	0.08

[a]PT, MT and HT: photoautotrophic, mixotrophic and heterotrophic respectively. [b]Total number of proteins common to all three replicates. [c]FDR: false discovery rate.

3.3 Differentially expressed proteins among the three types of cultures and pairwise comparison (two-sample unpaired t-test)

Of the 1564 proteins found to be reproducible under all three growth conditions, 827 showed significant differential expression at the 0.05 confidence level from ANOVA (Supplementary Table S2). This meant that about 22.5% of the 3673 proteins detected in the mid-log phase proteomic profile of *E. gracilis* Z were differentially expressed among the growth conditions, and suggests that the proteome of *Euglena gracilis* Z is significantly influenced by the mode of nutrient assimilation. This was comparable to the *E. gracilis* var. *saccharophila* strain from our previous study, in which about 21.5% of the 3843 proteins detected at the mid-log phase were differentially expressed [25].

The comparison between PT and the other cultures (MT and HT) of *E. gracilis* Z showed down-regulation of 316 and 369 proteins in the MT and HT cultures respectively, and up-regulation of 795 and 833 proteins in MT and HT cultures respectively. Using a confidence level of 0.05, 1771 proteins were unchanged between PT and MT cultures, and 1513 proteins were unchanged between PT and HT cultures (Supplementary Table S2).

About 30% of the proteins of *E. gracilis* Z could not be annotated with confidence, having a BLASTP *E*-value of greater than $1e^{-10}$. Protein classification of the remaining proteins using GO information for the up- and down-regulated proteins in 30 GO categories of interest are shown in Figure 3. The *E*-value from BLASTP results, along with the EC numbers of the candidate proteins and fold changes with respect to PT cultures of *E. gracilis* Z, are provided in Supplementary Table S3. The Supplementary Table S3 also contains the fold-changes of the candidate proteins from the proteome of *E. gracilis* var. *saccharophila*, compiled from the data provided in the Supplementary Table S2 of Hasan *et al*. (2017) [25]. The proteome of *E. gracilis* Z is significantly altered when grown under different nutritional treatments and lighting conditions. By analysing proteins from PT, MT and HT cultures, a more

comprehensive proteome for this model strain was established, and a wide range of proteins that were translated from the *E. gracilis* transcriptome was identified. Several proteins from carbon metabolism were differentially expressed, as discussed in the following sections.

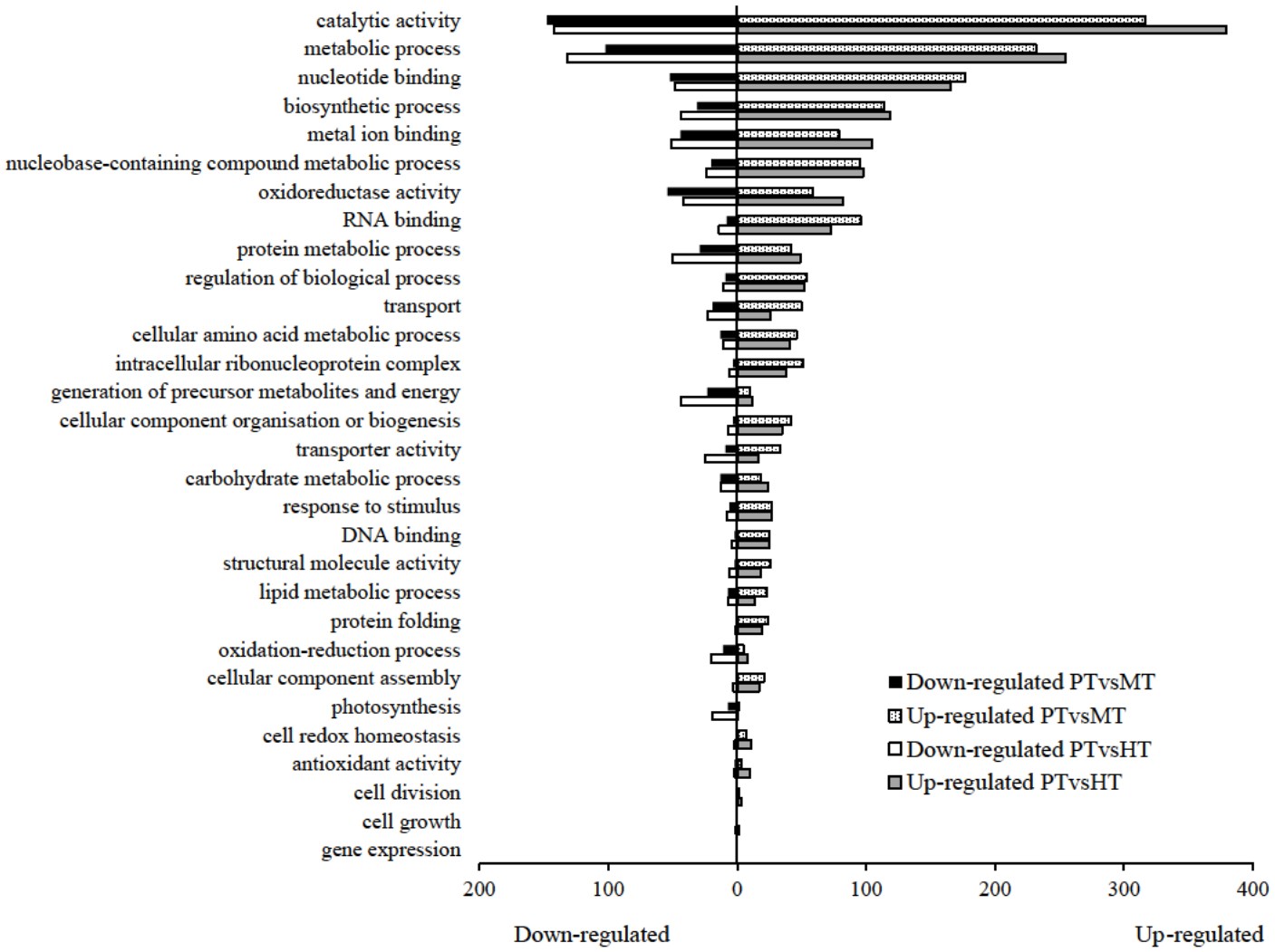

Figure 3: GO classifications of interest showing the up- and down-regulated proteins in the MT and HT cultures, compared to the PT culture of *E. gracilis Z*. The up- and down-regulated proteins were categorised into 30 different biological process and molecular function categories. Down-regulated proteins are plotted to the left, while the up-regulated proteins are plotted to the right of the central *y*-axis. The values indicate the number of proteins identified in each functional category.

3.4 Embden-Meyerhof-Parnas (EMP) pathway

Several isozymes were observed at the transcriptomic level for nine out of the ten core metabolic enzymes involved in the EMP pathway [28], and most of these isozymes were also found in the proteome (Figure 4). Some isozymes were differentially expressed among the growth conditions, while others were unique to either PT cultures (probably plastidic isoform) or to MT and HT cultures. For example, the second isozyme of pyruvate kinase (EC 2.7.1.40) was expressed only under PT condition, while the fifth isozyme of TPI (EC 5.3.1.1) was expressed only under MT and HT conditions. When compared to the *E. gracilis* var. *saccharophila* proteome from our previous study [25], a few isozymes also appeared to be strain-dependent and were significantly expressed in only one of these strains (Supplementary Table S3). For instance, a third isozyme of pyruvate kinase and a fourth isozyme of phosphoglycerate kinase (EC 2.7.2.3) were absent in *E. gracilis* Z, while the fifth isozyme of TPI and the third isozyme of ENO (EC 4.2.1.11) were absent in *E. gracilis* var. *saccharophila*.

Acetate flagellates such as *E. gracilis* have been previously reported to lack HK activity except under non-photosynthetic and acidic conditions, where they were required to utilise exogenous glucose [30]. A HK that acts on glucose, fructose and mannose substrates has been reported in *E. gracilis* [3], and a HK also exists in the transcriptomes of *E. gracilis* var. *saccharophila* and *E. gracilis* Z [28, 31]. However, in the current proteomic study of *E. gracilis* Z, the HK enzyme from the transcriptome could not be detected in the proteome. Moreover, HK could also not be detected from the proteome of *E. gracilis* var. *saccharophila* [25]. Instead, it appears that these strains utilise glucokinase (GK, EC 2.7.1.2) for the first step of the EMP pathway. This finding agrees with only one past study suggesting that instead of using a low-specificity HK like higher eukaryotes, *E. gracilis* exploits a range of high-specificity kinases to phosphorylate different sugars with two distinct kinases for glucose and fructose phosphorylation [32].

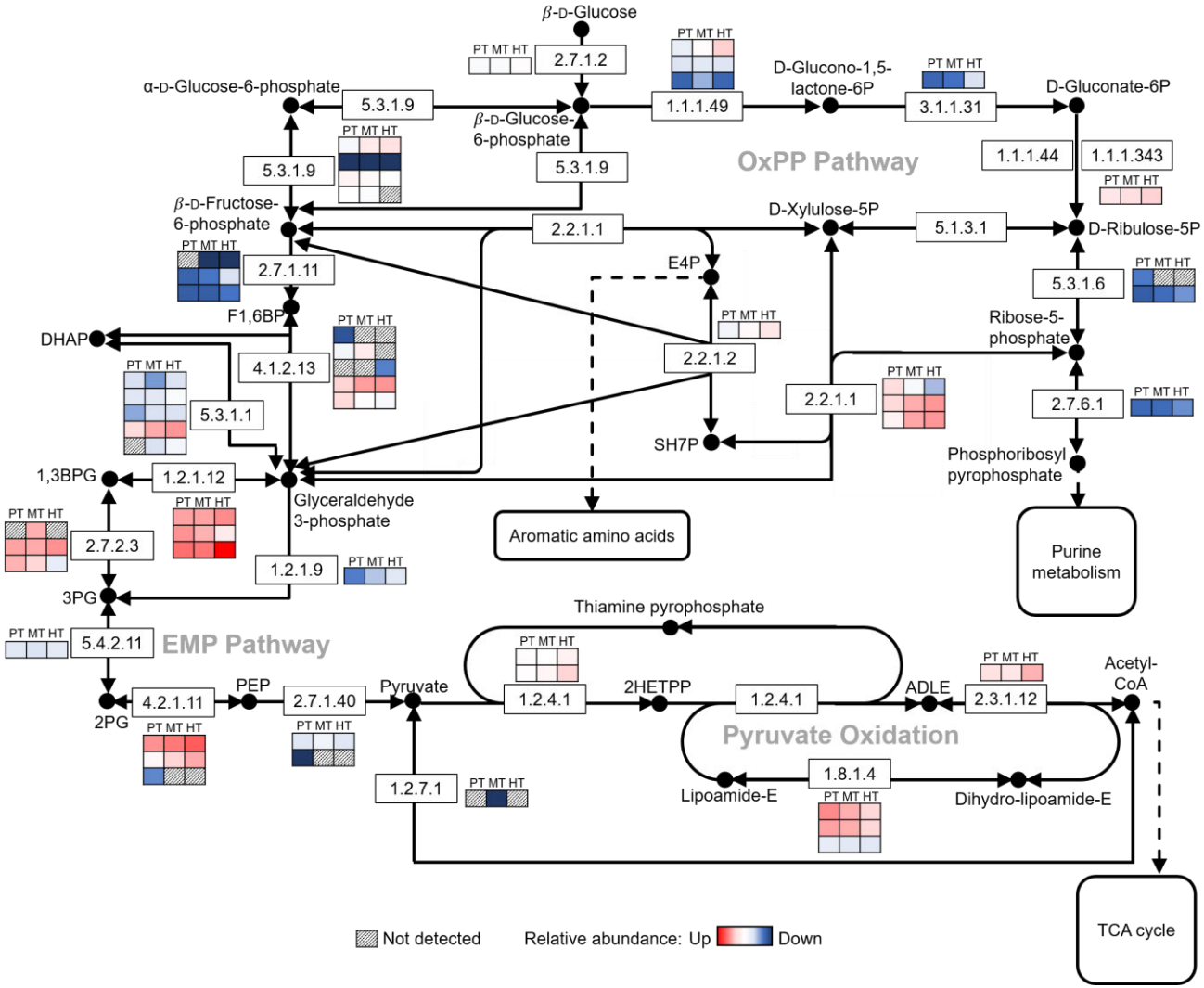

Figure 4: The Embden-Meyerhof-Parnas (EMP) and oxidative pentose phosphate (oxPP) pathways of *E. gracilis*. The heatmaps indicate enzyme expression levels of *E. gracilis* Z under photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivations, using PT cultivation as the control. Each row indicates a different isozyme for the respective enzyme class and the colour shows the relative abundance of the isozyme across all conditions. Overall, enzymes such as GAPDH (EC 1.2.1.12) showed a very high relative abundance across all conditions. Details of the expression patterns are described in Section 3.4 and Section 3.6.

Footnote: Isozymes that could not be identified from the transcriptome by amino acid sequence alone, or that were not detected at the proteome level have not been shown here. F1,6BP: β-D-Fructose-1,6-bisphosphate; DHAP: Dihydroxyacetone phosphate; 1,3BPG: 1,3-bisphosphoglycerate; 3PG: 3-phosphoglycerate; 2PG: 2-phosphoglycerate; PEP: phosphoenolpyruvate; 2HETPP: 2-Hydroxy-ethyl-TPP; ADLE: S-Acetyl DLE; CoA: Coenzyme A; X5P: D-Xylulose-5-phosphate; E4P: Erythrose-4-phosphate; SH7P: Sedoheptulose 7-phosphate; TCA: tricarboxylic acid.

3.5 Tricarboxylic acid (TCA) cycle

An overall up-regulation of the TCA cycle enzymes was observed in MT and HT cultures, especially in the latter. This could be owing to a lower rate of synthesis of these proteins and protein catabolism during growth under PT conditions. Protein catabolism is thought to provide amino acids for the synthesis of specific proteins under PT growth conditions [33]. The regulation of the TCA cycle enzymes is complex as the substrates are required for many other reactions within the cell, and the synthesis of a particular substrate can be increased or decreased based on how quickly it is removed from the cycle.

E. gracilis has been reported to use a modified TCA cycle (Figure 5), where 2-oxoglutarate is converted first to succinate-semialdehyde and then to succinate, as it lacks the eukaryotic 2-oxoglutarate dehydrogenase enzyme complex that converts 2-oxoglutarate to succinyl-CoA [34]. This modified cycle requires a 2-oxoglutarate decarboxylase enzyme (EC 4.1.1.-), which has a much lower affinity for its substrate than other TCA cycle enzymes and carries out an irreversible reaction, thus playing a critical role in the regulation of the TCA cycle [35]. The decarboxylase enzyme is suggested to be distinct from any other previously reported 2-oxoglutarate decarboxylase enzymes [35]. In the current proteomic study of *E. gracilis* Z, the candidate 2-oxoglutarate decarboxylase enzyme was up-regulated in HT cultures, but not in MT cultures (Figure 5). This was also true for the *E. gracilis* var. *saccharophila* proteome from our previous study (Supplementary Table S3). The expression of succinate-semialdehyde dehydrogenase isomers (SSDH, EC 1.2.1.16), which work in conjunction with the decarboxylase to convert succinate-semialdehyde to succinate, was also up-regulated under HT condition. This modified TCA cycle has also been reported in bacteroids and cyanobacteria that lack the 2-oxoglutarate dehydrogenase enzyme, and does not affect their ability to carry out cellular respiration [36-37].

An alternative route to bypass the 2-oxoglutarate dehydrogenase enzyme involves the use of a 4-aminobutyrate (GABA) shunt, and its existence has been confirmed in bacteroids and the mitochondrial matrix of *E. gracilis* Z [36, 38]. This route involves the conversion of 2-oxoglutarate to GABA via glutamate, which in turn is converted to succinate-semialdehyde (Figure 5). Candidates for the conversion of glutamate to GABA (glutamate decarboxylase, EC 4.1.1.15), and GABA to succinate-semialdehyde (4-aminobutyrate aminotransferase, EC 2.6.1.19) were identified in both the proteome of *E. gracilis* Z studied here, and the proteome published for *E. gracilis* var. *saccharophila* under the same growth conditions [25]. However, the relative abundance of enzymes from the route that uses 2-oxoglutarate decarboxylase, was significantly higher than the relative abundance of enzymes from the GABA shunt route (Figure 5). This indicates that while the 2-oxoglutarate decarboxylase route may have been able to compensate for the lack of a 2-oxoglutarate dehydrogenase complex, the GABA shunt route might not be sufficient, as the relative abundance of glutamate decarboxylase was too low. This can further be confirmed by reports of the activity of the SSDH isozymes that was too high to be coupled with the low activity of the GABA shunt, but could easily be coupled with the higher activity of 2-oxoglutarate decarboxylase [39].

Along with 2-oxoglutarate decarboxylase, most isozymes for aconitate hydratase (EC 4.2.1.3) and succinate dehydrogenase (EC 1.3.5.1) were also up-regulated under MT and HT cultivations compared to PT cultivation, while fumarate hydratase (EC 4.2.1.2) and malate dehydrogenase (EC 1.1.1.37) were up-regulated under HT cultivations only. These results match well with the growth pattern (Figure 1) as an increase in activity of the TCA cycle indicates higher metabolism rate, higher intermediate synthesis rate (required for various other biosynthetic pathways), increased flux throughout the pathway, and more energy production and cell growth, which were observed in the MT and HT cultures at mid-log growth phase. The lower mitochondrial activity under PT cultivation has also been previously reported in *E. gracilis* as a measure of succinate dehydrogenase activity [40].

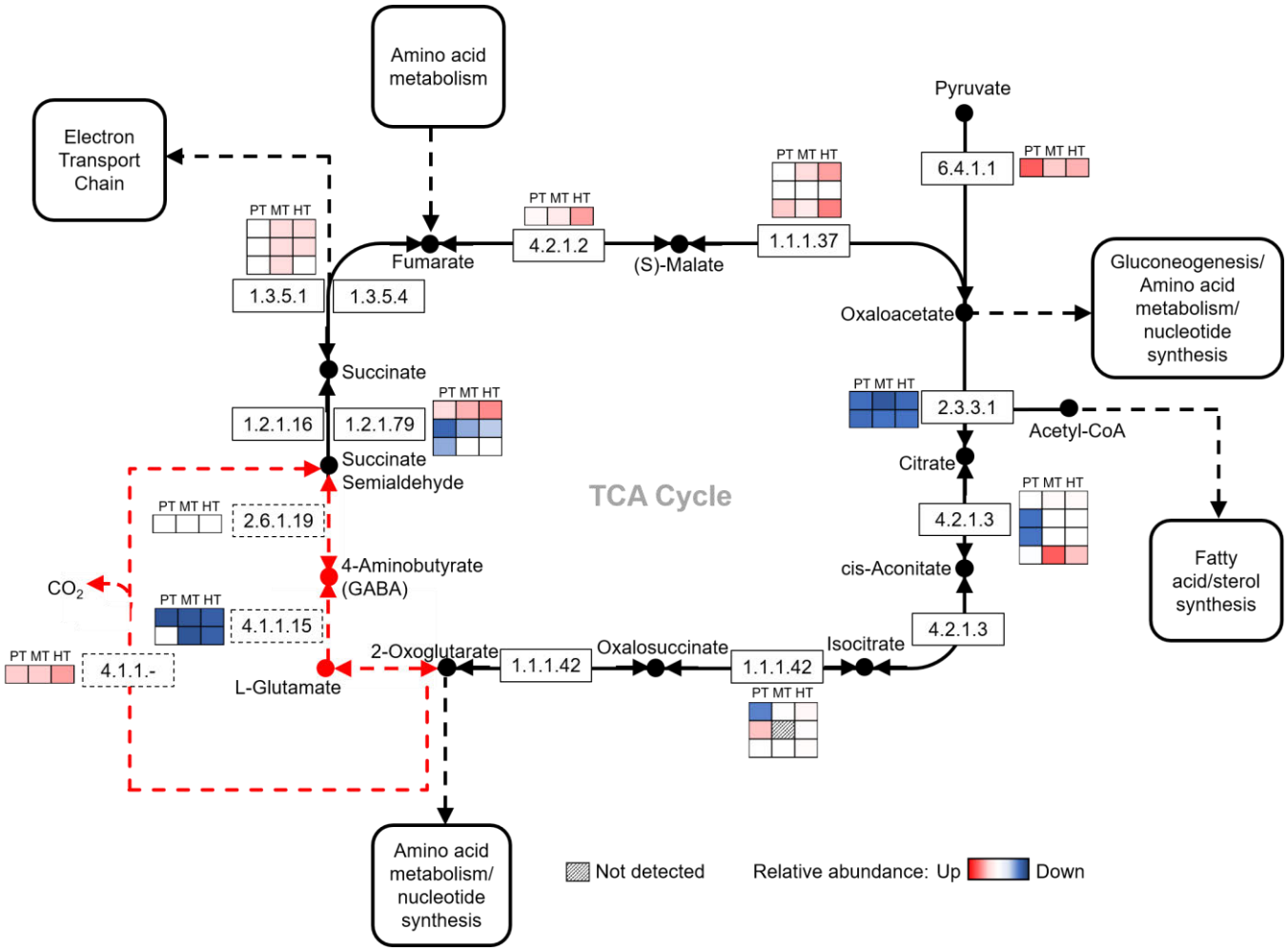

Figure 5: The alternate tricarboxylic acid (TCA) cycle of *E. gracilis*, with heatmaps indicating comparative protein expression in *E. gracilis* Z under photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivations, using PT as the control. Each row indicates a different isozyme for the respective enzyme class and the colour shows the relative abundance of the isozyme. The red dotted lines represent two alternate routes via the GABA shunt and 2-oxoglutarate decarboxylase (EC 4.1.1-), with the latter as the predominant route (Section 3.5). The TCA cycle provides intermediates for several pathways, some of which are shown in the black rounded boxes.

Footnote: Isozymes that could not be identified from the transcriptome by amino acid sequence alone, or that were not detected at the proteome level have not been shown here. CoA: Coenzyme A.

3.6 Oxidative pentose phosphate (oxPP) pathway

The oxPP pathway is an important source of metabolic intermediates for biosynthetic processes. The first three steps of the conventional pathway that involve conversion of glucose-6-phosphate (G6P) to ribulose-5-phosphate represent the irreversible oxidative phase of the oxPP pathway (Figure 4). Overall, there is an increase in the activity of the irreversible phase of the oxPP pathway during HT and MT cultivations, with some isozymes up-regulated under HT and MT cultivations and some unchanged. The G6P dehydrogenase enzyme (EC 1.1.1.49) in the first step of the oxPP pathway is the rate-limiting enzyme that determines the flux of the oxPP pathway, and the flux division between the oxPP and EMP pathways [41]. Three isozymes were detected for this enzyme, of which the first isozyme (Figure 4) was predominant, and was up-regulated under both HT and MT conditions. The overall up-regulation of this enzyme under HT and MT cultivations of *E. gracilis* Z may indicate that more G6P was dedicated to the oxPP pathway than to the EMP pathway under these growth conditions. The predominant G6P dehydrogenase isozyme was also up-regulated under HT cultivations of *E. gracilis* var. *saccharophila* [25]. The preference of the oxPP pathway over the EMP pathway in glucose metabolism during HT cultivation has been reported in other microalgae, such as *Chlorella sorokiniana* and cyanobacteria, although the EMP pathway was also functioning in these organisms [42]. On the other hand, under MT cultivation, these microalgae preferred the EMP pathway, similar to PT cultivation [42]. However, unlike them, MT cultivated cells of *E. gracilis* seemed to behave more like the HT cultivated cells preferring the oxPP pathway over the EMP pathway, as observed from the up-regulation of the irreversible reactions in MT cultures.

The increase in the irreversible reactions of the oxPP pathway during HT and MT cultivation leads to an increased generation of ribose-5-phosphate and erythrose-4-phosphate intermediates required for the biosynthesis of nucleotides, fatty acids and chorismate (precursor

to aromatic compounds) [41]. This is important for rapidly proliferating cells that tend to withdraw these intermediates from the oxPP pathway to meet the biological demands for DNA, lipid and protein syntheses. Candidates for ribose-phosphate pyrophosphokinase (EC 2.7.6.1) and glutamine phosphoribosyl amidotransferase (EC 2.4.2.14), which are involved in nucleotide synthesis, were also expressed in the *E. gracilis* Z proteome, as well as the *E. gracilis* var. *saccharophila* proteome from our previous work [25].

3.7 Carbon fixation

Both chemotrophic and photoautotrophic organisms are known to fix CO_2 from the atmosphere. There are six natural mechanisms of CO_2 fixation known [43-44]. The enzymes for the reductive acetyl-CoA pathway (Wood-Ljungdahl pathway) [44-46] were found to be absent in *E. gracilis* Z, similar to other eukaryotes and aerobic organisms. These enzymes were also absent from the *E. gracilis* var. *saccharophila* proteome studied previously [25]. Other CO_2 fixation pathways include the reductive tricarboxylic acid cycle (rTCA or Arnon-Buchanan cycle) [47] used by anaerobic bacteria/archaea or bacteria growing at very low oxygen concentrations, the 3-hydroxypropionate bicycle (3-hydroxypropionate/malyl-CoA cycle) found in green non-sulphur bacteria [48-49], and the 3-hydroxypropionate/4-hydroxybutyrate cycle and the dicarboxylate/4-hydroxybutyrate cycle used by archaea [50-51]. Since most of the enzymes from the rTCA cycle are common to the general TCA cycle, it is difficult to distinguish the involvement of these enzymes in the rTCA cycle in the current study. However, *E. gracilis* is capable of growing under low oxygen and anaerobic conditions [31] and recently, the rTCA cycle has been reported to occur in *E. gracilis* under dark, anaerobic conditions [52]. This suggests that in the current study carried out under aerobic conditions, the enzymes are probably reflective of the TCA cycle rather than the rTCA cycle. There is no evidence of CO_2 fixation via any of the other cycles mentioned above in *E. gracilis*, except for

a few enzymes detected in the *E. gracilis* Z proteome that also take part in fatty acid metabolism. However, the vast majority of *E. gracilis* proteins have yet to be investigated, and some may only be expressed under other growth conditions or in other strains not studied here. Since *E. gracilis* can adapt to both oxygenic photosynthesis and anaerobiosis, it may possess alternate mechanisms for CO_2 fixation under anaerobic condition.

The sixth mechanism of CO_2 fixation is the most prevalent pathway for organisms that perform oxygenic photosynthesis, such as plants, algae and cyanobacteria, as well as autotrophic proteobacteria, which can be anaerobic [44]. This reductive pentose phosphate pathway, or the Calvin cycle (Figure 6), is already known to exist in *E. gracilis*. Interestingly, the proteomic data here suggests that at least some isozymes of all the enzymes of the Calvin cycle are present under HT cultivation of *E. gracilis* Z, except phosphoribulokinase (EC 2.7.1.19) and sedoheptulose-bisphosphatase (EC 3.1.3.37), although most were lower in abundance than PT and MT cultures. Some of the other photosynthesis-related proteins encoded by the chloroplast genome were also present in HT cultures, such as PSII CP43 chlorophyll apoprotein, PSII cytochrome b559 subunit α, cytochrome b6, chloroplast ATPase subunits, and chloroplast ribosomal proteins. This could be because some of the isozymes play roles in other pathways active in the dark, but may also indicate that several of the photosynthesis-related proteins are regulated post-translationally. In the current study, the first isozyme of the RuBisCO large subunit (EC 4.1.1.39) shown in Figure 6 was the predominant one with a very high spectral count even in HT cultures, although the relative abundance of this isozyme was higher in PT and MT cultures than in HT cultures. This is similar to a recent report that also revealed that the expression of RuBisCO in *E. gracilis* is regulated post-translationally, at the level of enzyme complex formation, as the blocking of translation had no effect on the stability of the small subunit of RuBisCO [53].

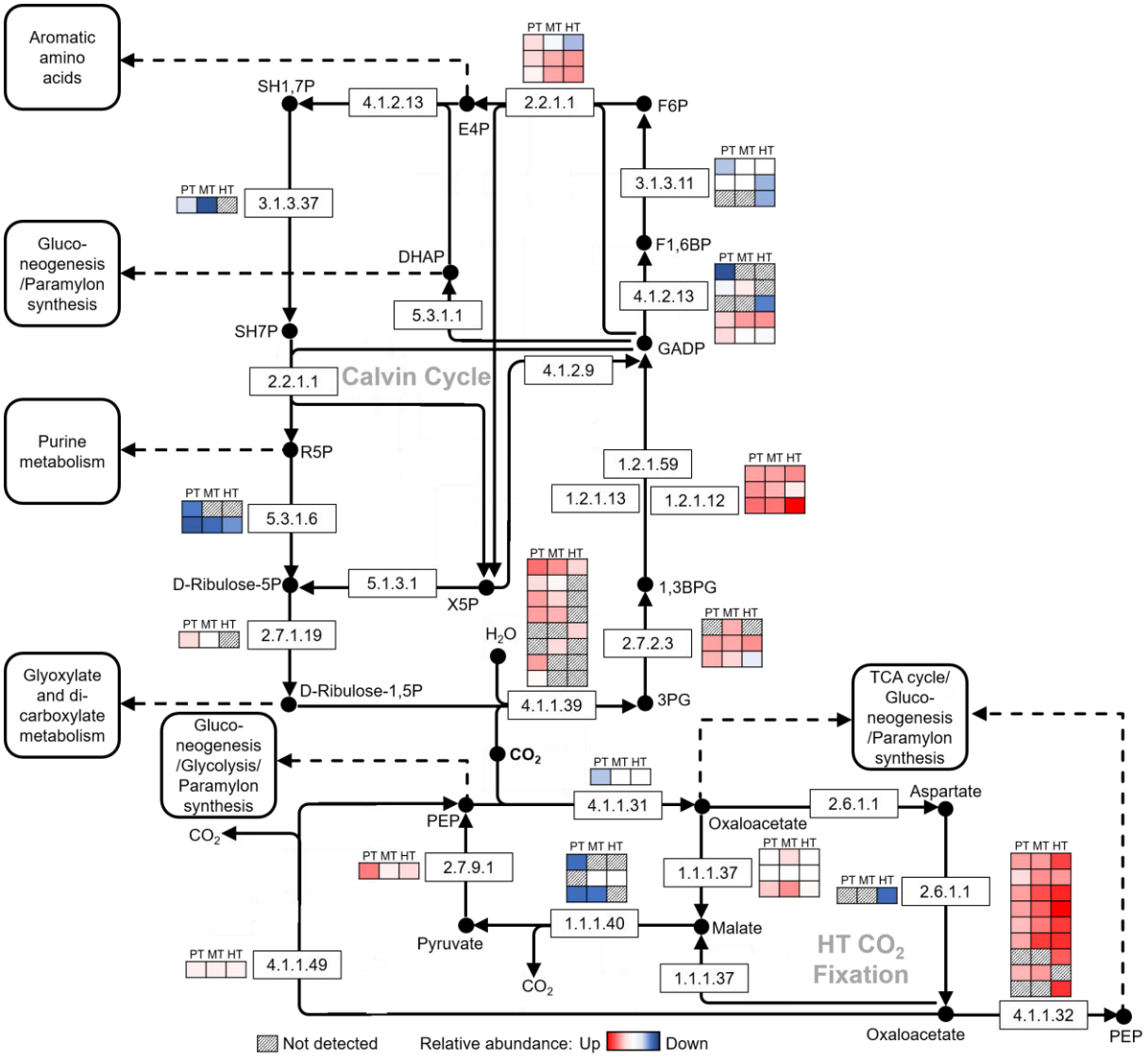

Figure 6: CO_2 fixation pathways in *E. gracilis*, using the Calvin cycle and heterotrophic CO_2 fixation. The heatmaps show enzyme expression levels of *E. gracilis* Z under photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivation, using PT cultivation as the control. Each row shows the expression of a different isozyme for the respective enzyme class, and the colour shows the relative abundance of the isozyme. Photosynthesis related-enzymes such as RuBisCO (EC 4.1.1.39) were also expressed under HT condition. The first seven rows of the heatmap for RuBisCO indicate the large subunit and the last row indicates the small subunit. HT CO_2 fixation enzymes such as PEPC (EC 4.1.1.31) and PEPCK (EC 4.1.1.32) were up-regulated during MT and HT cultivations, compared to PT cultivation.

Footnote: Isozymes that could not be identified from the transcriptome by amino acid sequence alone, or that were not detected at the proteome level have not been shown here. R5P: Ribose-5-phosphate; 3PG: 3-phosphoglycerate; 1,3BPG: 1,3-bisphosphoglycerate; GADP: Glyceraldehyde 3-phosphate; F1,6BP: β-D-Fructose-1,6-bisphosphate; F6P: β-D-Fructose-6-phosphate; E4P: Erythrose-4-phosphate; SH1,7P: Sedoheptulose-1,7-bisphosphate; SH7P: Sedoheptulose 7-phosphate; X5P: D-Xylulose-5-phosphate; DHAP: Dihydroxyacetone phosphate; PEP: Phosphoenolpyruvate.

E. gracilis has also been reported to carry out CO_2 fixation under HT condition in the dark, although the mechanism behind this fixation is unclear [54-57]. This study shows that although it is unlikely to be via one of the pathways mentioned above, all the enzymes used for Crassulacean acid metabolism (CAM) photosynthesis as well as the enzymes for C_4 carbon fixation are expressed in the proteome of *E. gracilis* Z (Supplementary Table S3). Under HT condition, *E. gracilis* appears to be using the same mechanism as these systems to assimilate CO_2 into C_4 dicarboxylic acids. The main candidate for this CO_2 fixation is phosphoenolpyruvate carboxylase (PEPC, EC 4.1.1.31) [58]. *E. gracilis* has been reported to produce PEPC, although it was previously suggested to be absent under PT condition and therefore concluded not to play an important role in CO_2 fixation [59]. The proteomic analysis of *E. gracilis* Z here, however, showed that this enzyme was present under all growth conditions, but was highly up-regulated under HT and MT conditions, and thus may contribute towards CO_2 fixation. A similar expression pattern was also observed from the proteome of the *E. gracilis* var. *saccharophila* strain (Supplementary Table S3). On the other hand, PEPC also takes part in the TCA cycle biosynthesis flux and its up-regulation under HT and MT conditions may also indicate the channelling of substrates into the TCA cycle.

A schematic pathway has been proposed previously for heterotrophic CO_2 fixation, which showed similarity to the initial steps of C_4 metabolism with the eventual synthesis of paramylon in a separate vesicle, provided lactate was present in the medium [60, 61]. In the current study with glucose in the medium, there is indication that a similar scheme was used for CO_2 fixation under HT cultivation (Figure 6). This evidence can be gathered from the increase of gluconeogenesis during HT cultivation (Figure 6) with the up-regulation of the regulatory PEP carboxykinase (PEPCK, EC 4.1.1.32) enzyme. There are also contradictory reports in the literature on PEPCK with reports of its presence and absence under PT cultivation, its presence under HT cultivation only when acetate or lactate is present in the medium, its presence in the cytosol only, its presence in the mitochondrial matrix, its

involvement in gluconeogenesis only, and its involvement in anaerobic CO_2 fixation [12-13, 59-61]. In the current proteomic study, at least nine PEPCK isozymes were detected, most of which were up-regulated under HT and MT cultivations, while some were unique to dark cultivation (HT cultivation only) and light cultivation (both PT and MT cultivations) respectively. This could mean that different isozymes were reported before in the contradictory studies. Overall, this could suggest that along with glucose uptake and conversion to paramylon, CO_2 fixation by PEPC and increase of gluconeogenesis may also contribute significantly towards paramylon synthesis under HT and MT cultivations during the mid-log phase.

3.8 Other carbon metabolism

The proteome of *E. gracilis* Z presented here, and the proteome of *E. gracilis* var. *saccharophila* from our previous work [25] show lack of evidence for the existence of methane metabolism, as well as the prokaryotic and archaeal Entner-Doudoroff pathways. Although the glyoxylate cycle is present in *E. gracilis*, a specific expression pattern could not be determined from the current study to distinguish between the different cultivation types in *E. gracilis* Z (Supplementary Table S3). Also, from the proteomic data, the enzymes of photorespiration were mostly up-regulated under PT cultivation compared to HT and MT cultivations (Supplementary Table S3). Photorespiration leads to the production of glycolate, which is converted to glyoxylate [62]. There are three fates of this glyoxylate, two of which involve the subsequent synthesis of serine via glycine and formate respectively [63]. The serine is eventually used for carbohydrate synthesis [62], and may be one of the major routes of carbohydrate biosynthesis during PT cultivation. The third fate of this glyoxylate is to regulate the TCA cycle, on which it has an inhibitory effect [64]. To ensure that inhibition is decreased, the glyoxylate is again reduced back to glycolate in the mitochondria by NADPH:glyoxylate

reductase (EC 1.1.1.79) [64]. Compared to PT condition, this enzyme was up-regulated under HT condition, but not MT condition, which may further explain why the TCA cycle was so highly up-regulated under HT cultivation (Section 3.5).

4 Conclusion

The total number of proteins detected for *E. gracilis* Z (3673 proteins) was comparable to the number of proteins detected for *E. gracilis* var. *saccharophila* (3843 proteins) [25], with about 30% proteins that could not be annotated. The proteomic analysis presented in this study helped to establish and validate some of the core carbon metabolism mechanisms in *E. gracilis*. For example, the HK sequence identified from the transcriptome was not expressed at the proteomic level, and it appears that instead *E. gracilis* uses a high-specificity GK for the first step of the EMP pathway. The relative abundance of the 2-oxoglutarate decarboxylase enzyme indicates a preference for the alternate TCA cycle route over the GABA shunt route. As in other microalgae, *E. gracilis* also showed some evidence of preference for the oxPP pathway over the EMP pathway for glucose metabolism under HT cultivation, but unlike other microalgae, oxPP pathway was also prevalent under MT cultivation. Expression of certain chloroplast-encoded proteins, such as RuBisCO under HT cultivation indicated post-translational control at the level of enzyme complex formation. Heterotrophic CO_2 fixation probably occurs mainly via PEPC, and may involve different pathways to produce C_4 and C_2 intermediates. Higher photorespiration under PT cultivation may further explain the up-regulation of the TCA cycle under HT cultivation. Several putative functional annotations have been provided in this study for proteins that have not been identified in *E. gracilis* before. The regulatory analyses presented here provide information about central carbon metabolism in *E. gracilis* that can be exploited to further improve *E. gracilis* strains for metabolite production.

Conflict of interest

The authors declare that they have no conflict of interest.

Contribution of authors

Conception and design: MTH, AS[1], HN, JT, GH, AS[2]. Collection and assembly of data: MTH. Analysis and interpretation of data: MTH, AS[1], MM. Statistical expertise: MM. Drafting of the article: MTH. Critical revision of the article for important intellectual content: HN, AS[1], JT, AS[2], GH. Obtaining of funding: HN, JT, GH.

Note: AS[1] refers to Angela Sun, and AS[2] refers to Anwar Sunna.

Acknowledgements

This work was supported by the Australian Research Council's Industrial Transformation Training Centre funding scheme (Project Number: IC130100009). The authors would like to thank Dr. Dana Pascovici for her help with bioinformatics, and acknowledge the contribution of APAF in facilitating this research, using infrastructure provided by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).

References

[1] Noor E, Eden E, Milo R and Alon U. 2010. Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. *Mol Cell*. 39(5): 809-820.

[2] Hurlbert RE and Rittenberg SC. 1962. Glucose metabolism of *Euglena gracilis* var. *bacillaris*; growth and enzymatic studies. *J Protozool*. 9(2): 170-182.

[3] Belsky MM and Schultz J. 1962. Partial characterization of hexokinase from *Euglena gracilis* var. *bacillaris*. *J Protozool*. 9(2): 195-200.

[4] Pelzer-Reith B, Wiegand S and Schnarrenberger C. 1994. Plastid class I and cytosol class II aldolase of *Euglena gracilis* (purification and characterization). *Plant Physiol*. 106(3): 1137-1144.

[5] Mo Y, Harris BG and Gracy RW. 1973. Triosephosphate isomerases and aldolases from light- and dark-grown *Euglena gracilis*. *Arch Biochem Biophys*. 157(2): 580-587.

[6] Grissom FE and Kahn JS. 1975. Glyceraldehyde-3-phosphate dehydrogenases from *Euglena gracilis*: Purification and physical and chemical characterization. *Arch Biochem Biophys*. 171(2): 444-458.

[7] Nowitzki U, Gelius-Dietrich G, Schwieger M, Henze K and Martin W. 2004. Chloroplast phosphoglycerate kinase from *Euglena gracilis*: Endosymbiotic gene replacement going against the tide. *Eur J Biochem*. 271(20): 4123-4131.

[8] Hannaert V, Brinkmann H, Nowitzki U, Lee JA, Albert MA, Sensen CW, Gaasterland T, Müller M, Michels P and Martin W. 2000. Enolase from *Trypanosoma brucei*, from the amitochondriate protist *Mastigamoeba balamuthi*, and from the chloroplast and cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. *Mol Biol Evol*. 17(7): 989-1000.

[9] Inui H, Miyatake K, Nakano Y and Kitaoka S. 1984. Occurrence of oxygen-sensitive, $NADP^+$-dependent pyruvate dehydrogenase in mitochondria of *Euglena gracilis*. *J Biochem*.

96(3): 931-934.

[10] Inui H, Ono K, Miyatake K, Nakano Y and Kitaoka S. 1987. Purification and characterization of pyruvate: $NADP^+$ oxidoreductase in *Euglena gracilis*. *J Biol Chem*. 262(19): 9130-9135.

[11] Ogawa T, Kimura A, Sakuyama H, Tamoi M, Ishikawa T and Shigeoka S. 2015. Identification and characterization of cytosolic fructose-1,6-bisphosphatase in *Euglena gracilis*. *Biosci Biotechnol Biochem*. 79(12): 1957-1964.

[12] Pönsgen-Schmidt E, Schneider T, Hammer U and Betz A. 1988. Comparison of phosphoenolpyruvate-carboxykinase from autotrophically and heterotrophically grown *Euglena* and its role during dark anaerobiosis. *Plant Physiol*. 86(2): 457-462.

[13] Miyatake K, Ito T and Kitaoka S. 1984. Subcellular location and some properties of phosphoenol-pyruvate carboxykinase (PEPCK) in *Euglena gracilis*. *Agric Biol Chem*. 48(8): 2139-2141.

[14] Buetow DE. 1989. The mitochondrion. In Ed Buetow DE, *The Biology of Euglena, vol. 4, Subcellular Biochemistry and Molecular Biology*. Academic Press, San Diego. 247-314.

[15] Zimorski V, Rauch C, van Hellemond JJ, Tielens AGM and Martin WF. 2017. The Mitochondrion of *Euglena gracilis*. In Eds. Schwartzbach SD and Shigeoka S, *Euglena: Biochemistry, Cell and Molecular Biology*. *Adv Exp Med Biol.* Springer International Publishing, Cham. 979: 19-37.

[16] Nakazawa M, Nishimura M, Inoue K, Ueda M, Inui H, Nakano Y and Miyatake K. 2011. Characterization of a bifunctional glyoxylate cycle enzyme, malate synthase/isocitrate lyase, of *Euglena gracilis*. *J Eukaryot Microbiol*. 58(2): 128-133.

[17] Vesteg M, Vacula R, Burey S, Löffelhardt W, Drahovská H, Martin W and Krajcovic J. 2009. Expression of nucleus-encoded genes for chloroplast proteins in the flagellate *Euglena gracilis*. *J Eukaryot Microbiol*. 56(2):159-166.

[18] Hoffmeister M, van der Klei A, Rotte C, van Grinsven KW, van Hellemond JJ, Henze

K, Tielens AG and Martin W. 2004. *Euglena gracilis* rhodoquinone:ubiquinone ratio and mitochondrial proteome differ under aerobic and anaerobic conditions. *J Biol Chem*. 279(21): 22422-22429.

[19] Keller M, Chan RL, Tessier LH, Weil JH and Imbault P. 1991. Post-transcriptional regulation by light of the biosynthesis of *Euglena* ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit. *Plant Mol Biol*. 17(1): 73-82.

[20] Kishore R and Schwartzbach SD. 1992. Translational control of the synthesis of the *Euglena* light harvesting chlorophyll *a*/*b* binding protein of photosystem II. *Plant Sci*. 85(1): 79-89.

[21] Madhusudhan R, Ishikawa T, Sawa Y, Shigeoka S and Shibata H. 2003. Post-transcriptional regulation of ascorbate peroxidase during light adaptation of *Euglena gracilis*. *Plant Sci*. 165(1): 233-238.

[22] Levasseur PJ, Meng Q and Bouck GB. 1994. Tubulin genes in the algal protist *Euglena gracilis*. *J Eukaryot Microbiol*. 41(5): 468-477.

[23] Saint-Guily A, Schantz ML and Schantz R. 1994. Structure and expression of a cDNA encoding a histone H2A from *Euglena gracilis*. *Plant Mol Biol.* 24(6): 941-948.

[24] Vacula R, Steiner JM, Krajcovic J, Ebringer L and Löffelhardt W. 2001. Plastid state- and light-dependent regulation of the expression of nucleus-encoded genes for chloroplast proteins in the flagellate *Euglena gracilis*. *Folia Microbiol (Praha)*. 46(5): 433-441.

[25] Hasan MT, Sun A, Mirzaei M, Te'o J, Hobba G, Sunna A and Nevalainen H. 2017. A comprehensive assessment of the biosynthetic pathways of ascorbate, α-tocopherol and free amino acids in *Euglena gracilis* var. *saccharophila*. *Algal Res*. 27: 140-151.

[26] Wu Y, Mirzaei M, Pascovici D, Chick JM, Atwell BJ and Haynes PA. 2016. Quantitative proteomic analysis of two different rice varieties reveals that drought tolerance is correlated with reduced abundance of photosynthetic machinery and increased abundance of ClpD1 protease. *J Proteomics*. 143: 73-82.

[27] Thompson EL, O'Connor W, Parker L, Ross P and Raftos DA. 2015. Differential proteomic responses of selectively bred and wild-type Sydney rock oyster populations exposed to elevated CO_2. *Mol Ecol*. 24(6): 1248-1262.

[28] O'Neill EC, Trick M, Hill L, Rejzek M, Dusi RG, Hamilton CJ, Zimba PV, Henrissat B and Field RA. 2015. The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Mol BioSyst*. 11(10): 2808-2820.

[29] Neilson KA, George IS, Emery SJ, Muralidharan S, Mirzaei M and Haynes PA. 2014. Analysis of rice proteins using SDS-PAGE shotgun proteomics. *Methods Mol Biol*. 1072: 289-302.

[30] Cramer M and Myers J. 1952. Growth and photosynthetic characteristics of *Euglena gracilis*. *Arch Microbiol*. 17(1-4): 384-402.

[31] Yoshida Y, Tomiyama T, Maruta T, Tomita M, Ishikawa T and Arakawa K. 2016. De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics*. 17: 182.

[32] Lucchini G. 1971. Control of glucose phosphorylation in *Euglena gracilis*. I. Partial characterization of a glucokinase. *Biochim Biophys Acta*. 242(2): 365-370.

[33] Schwartzbach SD. 2017. Photo and Nutritional Regulation of *Euglena* Organelle Development. In Eds. Schwartzbach SD and Shigeoka S, *Euglena: Biochemistry, Cell and Molecular Biology*. *Adv Exp Med Biol*. Springer International Publishing, Cham. 979: 159-182.

[34] Shigeoka S, Onishi T, Maeda K, Nakano Y and Kitaoka S. 1986. Occurrence of thiamin pyrophosphate-dependent 2-oxoglutarate decarboxylase in mitochondria of *Euglena gracilis*. *FEBS Lett*. 195(1-2): 43-47.

[35] Shigeoka S and Nakano Y. 1991. Characterization and molecular properties of 2-oxoglutarate decarboxylase from *Euglena gracilis*. *Arch Biochem Biophys*. 288(1): 22-28.

[36] Green LS, Li Y, Emerich DW, Bergersen FJ and Day DA. 2000. Catabolism of α-

ketoglutarate by a *sucA* mutant of *Bradyrhizobium japonicum*: evidence for an alternative tricarboxylic acid cycle. *J Bacteriol*. 182(10): 2838-2844.

[37] Zhang S and Bryant DA. 2011. The tricarboxylic acid cycle in cyanobacteria. *Science*. 334(6062): 1551-1553.

[38] Tokunaga M, Nakano Y and Kitaoka S. 1979. Subcellular localization of the GABA-shunt enzymes in *Euglena gracilis* strain Z. *J Protozool*. 26(3): 471-473.

[39] Shigeoka S, Hanaoka T, Kishi N and Nakano Y. 1992. Effect of L-glutamate on 2-oxoglutarate decarboxylase in *Euglena gracilis*. *Biochem J*. 282(Pt 2): 319-323.

[40] Brown GE and Preston JF. 1975. Changes in mitochondrial density and succinic dehydrogenase activity in *Euglena gracilis* as a function of the dependency on light for growth. *Arch Microbiol*. 104(1): 233-236.

[41] Kruger NJ and von Schaewen A. 2003. The oxidative pentose phosphate pathway: structure and organisation. *Curr Opin Plant Biol*. 6(3): 236-246.

[42] Perez-Garcia O, Escalante FME, de-Bashan LE and Bashan Y. 2011. Heterotrophic cultures of microalgae: Metabolism and potential products. *Water Res*. 45(1): 11-36.

[43] Kroth PG. 2015. The biodiversity of carbon assimilation. *J Plant Physiol*. 172: 76-81.

[44] Thauer RK. 2007. Microbiology: A fifth pathway of carbon fixation. *Science*. 318(5857): 1732-1733.

[45] Ljungdahl L and Wood HG. 1965. Incorporation of C^{14} from carbon dioxide into sugar phosphates, carboxylic acids, and amino acids by *Clostridium thermoaceticum*. *J Bacteriol*. 89(4): 1055-1064.

[46] Ljungdahl LG and Wood HG. 1969. Total synthesis of acetate from CO_2 by heterotrophic bacteria. *Annu Rev Microbiol*. 23: 515-538.

[47] Evans MCW, Buchanan BB and Arnon DI. 1966. A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proc Natl Acad Sci USA*. 55(4): 928-934.

[48] Herter S, Farfsing J, Gad'On N, Rieder C, Eisenreich W, Bacher A and Fuchs G. 2001.

Autotrophic CO_2 fixation by *Chloroflexus aurantiacus*: study of glyoxylate formation and assimilation via the 3-hydroxypropionate cycle. *J Bacteriol*. 183(14): 4305-4316.

[49] Zarzycki J, Brecht V, Müller M and Fuchs G. 2009. Identifying the missing steps of the autotrophic 3-hydroxypropionate CO_2 fixation cycle in *Chloroflexus aurantiacus*. *Proc Natl Acad Sci USA*. 106(50): 21317-21322.

[50] Berg IA, Kockelkorn D, Buckel W and Fuchs G. 2007. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science*. 318(5857): 1782-1786.

[51] Huber H, Gallenberger M, Jahn U, Eylert E, Berg IA, Kockelkorn D, Eisenreich W and Fuchs G. 2008. A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*. *Proc Natl Acad Sci USA*. 105(22): 7851-7856.

[52] Tomita Y, Yoshioka K, Iijima H, Nakashima A, Iwata O, Suzuki K, Hasunuma T, Kondo A, Hirai MY and Osanai T. 2016. Succinate and lactate production from *Euglena gracilis* during dark, anaerobic conditions. *Front Microbiol*. 7: 2050.

[53] Záhonová K, Füssy Z, Oborník M, Eliáš M and Yurchenko V. 2016. RuBisCO in non-photosynthetic alga *Euglena longa*: divergent features, transcriptomic analysis and regulation of complex formation. *PLoS One*. 11(7): e0158790.

[54] Peak JG and Peak MJ. 1980. Heterotrophic carbon dioxide fixation products of *Euglena*: effects of ammonium. *Plant Physiol*. 65(3): 566-568.

[55] Wolfovitch R and Perl M. 1972. A possible ribosomal-directed regulatory system in *Euglena gracilis*. Carbon dioxide fixation. *Biochem J*. 130(3): 819-823.

[56] Levedahl BH. 1966. Heterotrophic CO_2 fixation by a bleached *Euglena*. *Exp. Cell Res*. 44(2-3): 393-402.

[57] Peak JG and Peak MJ. 1976. Heterotrophic (dark) CO_2 fixation by *Euglena gracilis*. Possible regulation by tricarboxylic acid cycle intermediates. *J Protozool*. 23(1): 165-167.

[58] Perl M. 1974. Phosphoenolpyruvate carboxylating enzyme in dark-grown and in green *Euglena* cells. *J Biochem*. 76(5): 1095-1101.

[59] Peak JG and Peak MJ. 1981. Heterotrophic carbon dioxide fixation by *Euglena gracilis*: Function of phosphoenolpyruvate carboxylase. *Biochim Biophys Acta*. 677(3-4): 390-396.

[60] Calvayrac R, Laval-Martin D, Briand J and Farineau J. 1981. Paramylon synthesis by *Euglena gracilis* photoheterotrophically grown under low O_2 pressure: Description of a mitochloroplast complex. *Planta*. 153(1): 6-13.

[61] Briand J, Calvayrac R, Laval-Martin D and Farineau J. 1981. Evolution of carboxylating enzymes involved in paramylon synthesis (phosphoenolpyruvate carboxylase and carboxykinase) in heterotrophically grown *Euglena gracilis*. *Planta*. 151(2): 168-175.

[62] Merrett MJ and Lord JM. 1973. Glycollate formation and metabolism by algae. *New Phytol*. 72(4): 751-767.

[63] Yokota A, Komura H and Kitaoka S. 1985. Different metabolic fate of two carbons of glycolate in its conversion to serine in *Euglena gracilis* Z. *Arch Biochem Biophys*. 242(2): 498-506.

[64] Nakazawa M. 2017. C2 metabolism in *Euglena*. In Eds. Schwartzbach SD and Shigeoka S, *Euglena: Biochemistry, Cell and Molecular Biology*. *Adv Exp Med Biol.* Springer International Publishing, Cham. 979: 39-45.

CHAPTER 5

Differential response of *Euglena gracilis* enzymes involved in paramylon metabolism under photoautotrophic, mixotrophic and heterotrophic cultivations

5.1 Introduction

The uniflagellate microalga *Euglena gracilis* synthesises a large β-1,3-glucan called paramylon as its primary carbohydrate storage. Paramylon is an unbranched polymer, unique to Euglenoids, that has immune stimulatory [1], anti-tumour [2] and anti-HIV [3] properties when used directly as a nutraceutical or as a modified compound. The cellular paramylon content has been reported to be as high as 90% of the dry mass under heterotrophic conditions in a bleached mutant of *E. gracilis* [4]. Recently, the pathways for paramylon biosynthesis and degradation have been described using transcriptomic data [5-6]. The transcriptome of the *E. gracilis* Z strain revealed two candidates for the paramylon synthase enzyme [5], while that of *E. gracilis* var. *saccharophila* revealed four candidates [6]. Another recent study has shed further light into the biosynthesis of paramylon, confirming the existence of the two paramylon synthase candidates *EgGSL1* and *EgGSL2* in *E. gracilis* Z by gene knockdown studies [7]. Paramylon can be catabolised to meet the energy requirements of cells under low environmental carbon conditions. Under anaerobic conditions, *E. gracilis* uses paramylon as the starting material for wax ester synthesis [5]. It is also catabolised in response to ammonium ions and high atmospheric oxygen concentrations [8-9]. Several candidates for β-glucosidases were also identified in the transcriptome of both these strains [5-6]. However, the expressions of most of the paramylon synthesis and degradation pathway enzymes have not been confirmed at the proteomic level and their relative abundance has not been reported. Since *E. gracilis* controls the expression of a large number of its proteins at the post-transcriptional level, a proteomic study can provide better insight into the actual proteins expressed from the candidates predicted in the transcriptomes. The aims of this chapter were thus to determine the enzymes of the paramylon biosynthesis and degradation pathways using label-free shotgun proteomics, to provide functional annotations to the enzymes and to investigate the differential expression of these enzymes under PT, MT and HT growth conditions.

In summary, the preparation of a streptomycin-bleached strain *E. gracilis* ZSB and its proteomic profile are described in this chapter, with focus on the paramylon biosynthetic and degradation pathways. Additionally, the label-free shotgun proteomics data of *E. gracilis* var. *saccharophila* from Chapter 3 and that of *E. gracilis* Z from Chapter 4 are further analysed to investigate differentially expressed proteins from the paramylon metabolic pathways, in response to photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) growth conditions. The paramylon content of all three strains is also reported, along with qRT-PCR analysis of the *EgGSL1* and *EgGSL2* transcripts to verify the proteomics results. Finally, the proteomic profile of an additional three replicates of each of *E. gracilis* var. *saccharophila* and *E. gracilis* Z grown under HT cultivation are used to validate the results obtained from the HT cultures of *E. gracilis* var. *saccharophila* and *E. gracilis* Z from Chapters 3 and 4 respectively.

The results for this research are reported as a thesis chapter and will be prepared for publication in the future when additional validation studies have been performed for all the enzymes of the paramylon synthesis and degradation pathways.

5.2 Materials and Methods

5.2.1 Preparation of the streptomycin-bleached mutant *Euglena gracilis* ZSB

The streptomycin-bleached mutant, *E. gracilis* ZSB was bleached and grown in our lab for over two years before these experiments were carried out. Briefly, *E. gracilis* Z was grown in modified Hutner medium (pH 3.5) [10], supplemented with vitamins B_1 and B_{12}, 10 g l^{-1} yeast extract and 1.5% (v/v) D-(+)-glucose. The cells were subjected to permanent bleaching by the addition of 0.5 mg l^{-1} streptomycin to the culture and exposing it to white light illumination (2000 lx). The culture was grown for approximately 10 generations and then streaked onto *Euglena* maintenance (EM) medium plates. The EM medium consisted of (per litre): anhydrous sodium acetate (1 g), peptone (1 g), tryptone (2 g), yeast extract (2 g),

$CaCl_2.2H_2O$ (0.01 g) and Noble agar (15 g). Isolated white colonies were further streaked onto EM medium plates in ten successive cultures. One isolated colony from the 10^{th} plate was then grown in the modified Hutner medium mentioned above. This stock culture was grown continuously for one year with occasional medium replenishment. After one year, the streak plating technique of the stock culture was repeated on 10 successive EM medium plates and a single colony from the 10^{th} plate was isolated. The isolated colony was further grown in the modified Hutner medium for one year (stock culture) to confirm permanent bleaching and maintain the cell line. The stock culture was then inoculated into fresh modified Hutner medium in two flasks, one of which was exposed to light and the other was kept in complete darkness. The shake-flask cultures were grown for about three months and served as the stock cultures for the experiments that followed.

5.2.2 Cultivation of the *Euglena gracilis* ZSB strain

The *E. gracilis* ZSB strain was maintained as pure axenic stock cultures using the modified Hutner medium as mentioned above. *E. gracilis* ZSB was grown under mixotrophic (MT) and heterotrophic (HT) growth conditions only. Photoautotrophic (PT) cultivation could not be carried out as *E. gracilis* ZSB is an obligate heterotroph having no mature chloroplasts. Cultivation under MT and HT growth conditions has already been described in Chapter 3 (Section 2.2) and the same method was applied to the *E. gracilis* ZSB cultures. The light-adapted stock culture described above was used to inoculate the MT cultures of *E. gracilis* ZSB, while the dark-adapted stock cultures were used to inoculate the HT cultures.

5.2.3 Confirmation of paramylon synthase expression using additional HT cultures

Three additional replicates of *E. gracilis* var. *saccharophila* and *E. gracilis* Z were grown under HT condition. The replicates were grown using the same cultivation parameters of the HT cultures described in Chapters 3 (Section 2.2). The proteomic profiles from these extra replicates were used to confirm the label-free shotgun proteomics data of the paramylon synthase candidates (EgGSL1 and EgGSL2) from the HT cultures of *E. gracilis* var. *saccharophila* (Chapter 3) and *E. gracilis* Z (Chapter 4).

5.2.4 Determination of dry mass, paramylon and unused glucose in the medium

Dry mass of *E. gracilis* ZSB was measured every 24 h using pre-weighed aluminium evaporating dishes as described in Chapters 3 and 4.

Paramylon assay was also carried out every 24 h for the *E. gracilis* ZSB strain described in this chapter, as well as the *E. gracilis* var. *saccharophila* and *E. gracilis* Z strains described in Chapters 3 and 4 respectively using aliquots from the respective cultures. The paramylon assay method described previously using the phenol-sulphuric acid method was employed [10] with slight modifications. Briefly, one ml of harvested cells was washed once with MilliQ water and one ml of 9:1 (v/v) water:30% perchloric acid was added to the pellet. The cells were sonicated at 4°C for one cycle of 10 sec at 30% of the maximum amplitude of a Branson Digital Sonifier® 450 with a probe of tip diameter three mm, and centrifuged at 1100 *g* for two min. The supernatant was discarded and the pellet was resuspended in one ml of 1% (w/v) SDS by vortexing vigorously for one min. The suspension was heated in a boiling water bath for 15 min and then centrifuged at 1100 *g* for 15 min. The supernatant was discarded and the pellet was resuspended in one ml of 1 N NaOH solution. A 50 µl aliquot of this was transferred to a fresh tube and 600 µl of 5% (w/v) phenol was added to it, followed by 1250 µl of concentrated

(95-98% v/v) sulphuric acid. The resulting yellow-orange colour was allowed to develop for 20 min at 70°C. The absorbance was measured at 480 nm using a microplate reader (BMG FLUOstar Galaxy, Germany) and converted to glucose equivalents using a glucose standard calibration curve. Unused glucose left in the medium was measured in the same way as paramylon, using supernatant from the initial harvest of cells.

5.2.5 Label-free shotgun proteomics

Proteomic analysis was carried out during the mid-log phase at 36 h for MT and HT cultures of *E. gracilis* ZSB, and at 24 h for the additional HT cultures of *E. gracilis* var. *saccharophila* and *E. gracilis* Z. Protein extraction, precipitation by methanol:chloroform, quantification by BCA assay, fractionation by SDS-PAGE and in-gel tryptic digestion were carried out as described in Chapter 2 (Section 2.5.1-2.5.4). Peptide extraction and purification and nanoLC-MS/MS were carried out as described in Chapter 2 (Section 2.5.5-2.5.6), while the protein/peptide identification from the MS/MS spectra has been described in Chapter 3 (Section 2.9). The global proteome machine (GPM) was used for protein/peptide identification and the parameters used for this software have been defined in Chapter 2 (Section 2.5.7). Finally, data processing and quality control, and statistical analysis and functional annotations using BLASTP have been described in Chapters 2 (Section 2.5.8) and 3 (Sections 2.10 and 2.11) respectively.

For the two-sample unpaired *t*-test comparison the proteomic profiles of the MT and HT cultures of *E. gracilis* ZSB described in this chapter were compared to the proteomic profiles of the MT and HT cultures respectively of *E. gracilis* Z (Chapter 4). The three additional HT cultures of *E. gracilis* var. *saccharophila* and *E. gracilis* Z were only used for validation purposes to confirm the expression of the EgGSL1 and EgGSL2 proteins under HT cultivation; two-sample unpaired *t*-test comparison were not carried out for these additional

replicates. Furthermore, the proteomic profiles of *E. gracilis* var. *saccharophila* and *E. gracilis* Z (Chapters 3 and 4 respectively) were analysed to compare differential expression of proteins from the paramylon biosynthetic and degradation pathways in MT and HT cultures, compared to PT cultures.

5.2.6 Real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR)

The proteomic results for the paramylon synthase candidates (EgGSL1 and EgGSL2) were further validated using qRT-PCR. The qRT-PCR reactions were carried out for all cultures of the *E. gracilis* ZSB strain described in this chapter. Sample aliquots were also collected from all the cultures of the *E. gracilis* var. *saccharophila* strain and the *E. gracilis* Z strain described in Chapters 3 and 4 respectively for qRT-PCR. One ml of sample was collected for qRT-PCR at each of the following time points: 24, 48, 72 and 96 h for PT cultures of *E. gracilis* var. *saccharophila* and *E. gracilis* Z; 12, 24, 48, 72 and 96 h for MT and HT cultures of all strains.

5.2.6.1 RNA extraction

RNA was extracted from the samples with TRIzol™ reagent (Thermofisher Scientific, USA), as per the manufacturer's protocol for cells grown in suspension. All steps were carried out at 4°C. Harvested cells were homogenised in 1 ml TRIzol™ reagent, 0.2 ml chloroform was added, and the mixture was shaken vigorously by hand for 10 sec, allowed to stand for 2-3 min and centrifuged at 12,000 *g* for 15 min. The aqueous upper phase was transferred to a fresh tube, 0.5 ml isopropanol was added, the mixture was allowed to stand for 10 min and centrifuged at 12,000 *g* for 10 min. The supernatant was discarded and the RNA precipitate

was washed with 1 ml of 75% (v/v) ethanol and centrifuged at 7500 *g* for 5 min. The supernatant was discarded, the RNA pellet air dried for 5 min and resuspended in 20 µl of RNase-free water.

5.2.6.2 RNA quantification by Qubit® RNA BR assay

RNA quantification was carried out using the Qubit™ RNA BR Assay Kit (Thermofisher Scientific, USA) as per the manufacturer's protocol. The Qubit® working solution was prepared by diluting the Qubit® RNA BR Reagent 1:200 (v/v) in Qubit® RNA BR Buffer. Ten µl of each Qubit® standard was mixed with 190 µl of the Qubit® working solution by vortexing for 2-3 sec. Similarly, the RNA samples were mixed with the Qubit® working solution to a final volume of 200 µl and vortexed for 2-3 sec. The RNA sample was then quantified using the RNA Broad Range assay type on the Qubit® 3.0 Fluorometer (Thermofisher Scientific, USA), by reading the Qubit® standard solutions first, followed by the RNA sample.

5.2.6.3 cDNA synthesis

Any contaminating genomic DNA in the RNA sample was digested using five units of DNase I (NEB, UK) per 25 µg RNA in a reaction volume made up to 60 µl with 1x DNase Reaction Buffer (NEB, UK). This was mixed thoroughly and incubated at 37°C for 10 min. The reaction was stopped with 1 µl of 0.5 M EDTA (NEB, UK), followed by heat inactivation of the enzymes at 75°C for 10 min.

The AffinityScript QPCR cDNA Synthesis Kit (Agilent Technologies, UK/Ireland) was then used for cDNA synthesis starting with 2 µg of the RNA sample. RNase-free water to a final volume of 20 µl, 10 µl of the first strand master mix (2x), 3 µl of random primers (0.1 µg

μl^{-1}), 1 µl of the AffinityScript RT/RNase Block enzyme mixture and the RNA sample were added in that order to a microcentrifuge tube. The mixture was incubated at 25°C for 5 min (primer annealing), 42°C for 15 min (cDNA synthesis) and 95°C for 5 min (terminating cDNA synthesis). Negative controls were prepared using the same cDNA synthesis method, but adding RNase-free water instead of the AffinityScript RT/RNase Block enzyme mixture. This enabled detection of any amplification resulting from contaminating genomic DNA as these samples did not have any synthesised cDNA to amplify.

5.2.6.4 Primers and parameters for quantitative PCR (qPCR)

Three sets of primers were designed and used for the reactions, two for the paramylon synthase transcripts (*EgGSL1* and *EgGSL2*) and one for 18S rRNA as the house-keeping gene. The forward and reverse primers for *EgGSL1* were 5'-AGCGCATCCCGGTCACAATG-3' and 5'-TTGTCCCACGTGTACGCCAC-3' respectively; for *EgGSL2* they were 5'-ATGCAGTACGTGGTCTCCGTGC-3' and 5'-AGGAACCCCAATTGCCCCAACC-3' respectively; and for the 18S rRNA they were 5'-CTTAGATCGCTGCCAGATCC-3' and 5'-GGCTGTGGATTTCTCGTTGT-3' respectively.

The Brilliant II SYBR® Green QPCR Master Mix (Agilent Technologies, UK/Ireland) was used for qRT-PCR as per the manufacturer's protocol. Five µl of nuclease-free PCR-grade water, 10 µl of the 2x Brilliant II SYBR® Green QPCR master mix, 2 µl of 200 nM forward primer, 2 µl of 200 nM reverse primer and 1 µl (100 ng) of cDNA were added in that order to a well of a 384-well plate (LightCycler® 480 Multiwell Plate 384, Roche, Australia) using the Echo® 550 Liquid Handler (Labcyte, CA, USA). Three technical replicate reactions were carried out for each biological replicate from each time point. The LightCycler® 480 Instrument II (Roche, Australia) was then used to carry out the qPCR reactions. The initial denaturation

was carried out at 95°C for 10 min, followed by 40 cycles of 95°C for 30 sec (denaturation) and 60°C for 1 min for (annealing/extension).

5.2.6.5 Data analysis by the $2^{-\Delta\Delta C_T}$ method

Data were analysed using the $2^{-\Delta\Delta C_T}$ method [11]. Briefly, the 18S rRNA housekeeping gene was used to normalise systematic sample to sample variation. The threshold cycle (C_T) for each reaction was retrieved from the LightCycler® 480 Software. The average C_T for replicates was calculated and the average C_T for the 18S rRNA housekeeping gene was then subtracted from the average C_T of the target cDNA (*EgGSL1* and *EgGSL2*) at each time point to obtain ΔC_T. The PT sample collected at 24 h, and the MT and HT samples collected at 12 h were used as the controls against which the transcript fold-change was calculated for all other time points. The ΔC_T value of the control was subtracted from the ΔC_T value of each sample time point (Section 5.2.6) to obtain the $\Delta\Delta C_T$ value. The transcript fold-change was then calculated by using the formula $2^{-\Delta\Delta C_T}$.

5.3 Results and discussion

5.3.1 Dry mass of the streptomycin-bleached *Euglena gracilis* ZSB strain

The exponential growth phase of the bleached mutant *E. gracilis* ZSB occurred between 0 and 72 h under both MT and HT conditions (Figure 5.1) and was similar to the exponential growth phase of *E. gracilis* Z HT cultures. MT and HT cultures of the mutant strain yielded a higher dry mass than *E. gracilis* var. *saccharophila* and *E. gracilis* Z and the mid-log phase was about 36 h for both MT and HT cultures.

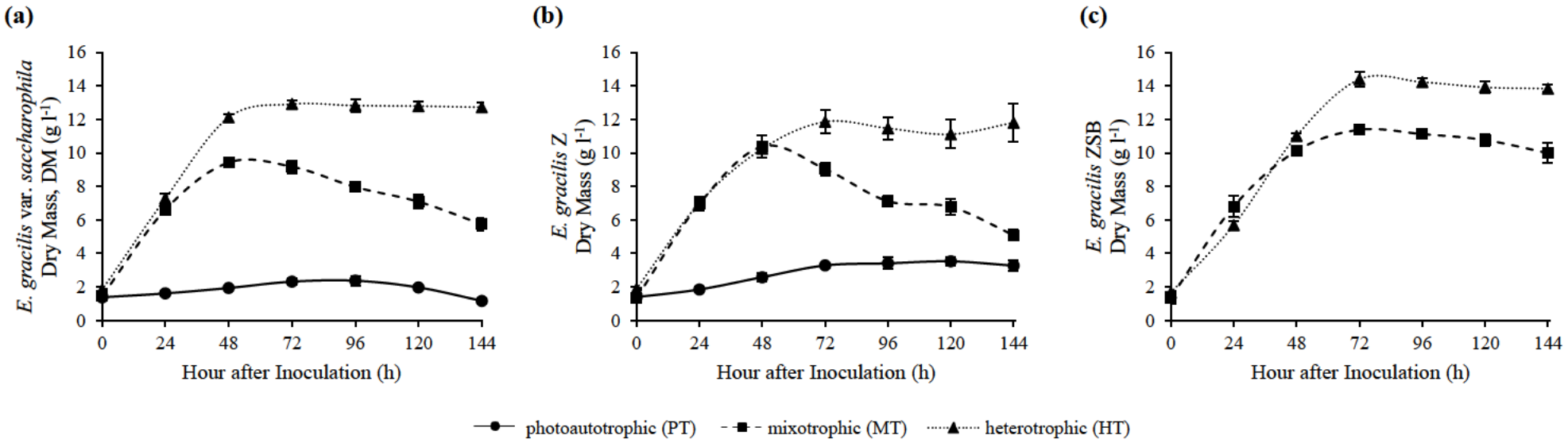

Figure 5.1: Growth patterns of (a) *E. gracilis* var. *saccharophila*, (b) *E. gracilis* Z and (c) *E. gracilis* ZSB. The growth curves of *E. gracilis* var. *saccharophila* and *E. gracilis* Z have been reproduced from Chapters 3 and 4 respectively. Exponential growth phase occurs between 0 to 48 h for *E. gracilis* var. *saccharophila* and *E. gracilis* Z under MT and HT cultivations and between 0 to 72 h for both all other cultures. HT cultivations of all strains, PT cultivation of *E. gracilis* Z and MT cultivation of *E. gracilis* ZSB showed prolonged stationary phases, while MT cultivations of *E. gracilis* var. *saccharophila* and *E. gracilis* Z and PT cultivation of *E. gracilis* var. *saccharophila* showed much shorter stationary phases.

5.3.2 Paramylon accumulation

The paramylon content of the three strains under all cultivation conditions is shown in Figure 5.2, along with the percentage of paramylon per dry mass during exponential growth. The paramylon contents of the PT cultures of *E. gracilis* var. *saccharophila* and *E. gracilis* Z were negligible, while the highest paramylon content was accumulated by the HT cultures of all strains. The bleached mutant strain accumulated significantly higher paramylon per volume of culture (over 6 g l^{-1} and 10 g l^{-1} for MT and HT cultivations respectively). Paramylon was rapidly broken down in both *E. gracilis* var. *saccharophila* and *E. gracilis* Z by late exponential growth phase under MT and HT cultivations. During MT cultivations, the paramylon content eventually returned to about the same initial baseline level of the respective inocula. However, during the HT cultivations, after an initial rapid catabolism, the paramylon content did not decrease further, remaining significantly higher than the initial baseline level of the respective inocula. In the mutant strain, paramylon was broken down very slowly under both MT and HT cultivations, remaining much higher than the initial baseline level of the respective inocula throughout the cultivation period. Paramylon is reported to be rapidly broken down in presence of light to provide energy required for chloroplast development [12]. This may explain why paramylon breaks down rapidly in *E. gracilis* var. *saccharophila* and *E. gracilis* Z under MT cultivation, but not in the bleached mutant, which does not produce chloroplasts.

The maximum percentage of paramylon per dry mass was the highest for *E. gracilis* ZSB (87.5% and 67.0% in HT and MT cultivations respectively), followed by *E. gracilis* var. *saccharophila* (83.1% and 77.6% for HT and MT cultivations respectively) and *E. gracilis* Z (63.9% and 51.8% for HT and MT cultivations respectively). The peak paramylon content per l of culture occurred at about 48 h for MT and HT cultures of *E. gracilis* Z, and at 48 h for MT cultures and 72 h for HT cultures of *E. gracilis* ZSB (Figure 5.2 b and c). However, paramylon accumulation was more rapid in *E. gracilis* var. *saccharophila*, reaching a peak within about

24 to 30 h (Figure 5.2 a). This could partly have been influenced by the rapid conversion of tryptophan to indoleacetate (IA) observed in *E. gracilis* var. *sacchrophila* (Chapter 3, Section 3.6.1), which in turn encourages faster growth and stimulates paramylon accumulation, especially during MT and HT cultivations [13]. *E. gracilis* var. *saccharophila* also took up glucose from the medium much faster than *E. gracilis* Z and ZSB, during both the HT and MT cultivation conditions (Supplementary Figure S5.1). The paramylon synthesis and breakdown patterns under MT cultivation of *E. gracilis* var. *saccharophila* and *E. gracilis* Z indicate that the culture initially behaves like a HT culture when glucose is present in the medium, accumulating large amounts of paramylon. As the glucose is depleted the cells synthesise chlorophyll, behaving more like a PT culture. This leads to the rapid breakdown of paramylon to provide energy for chloroplast development [12]. Since a large portion of the biomass comprises of paramylon, this may also explain the sudden decline in biomass of cells under MT cultivation as they lose their paramylon and shrink in size.

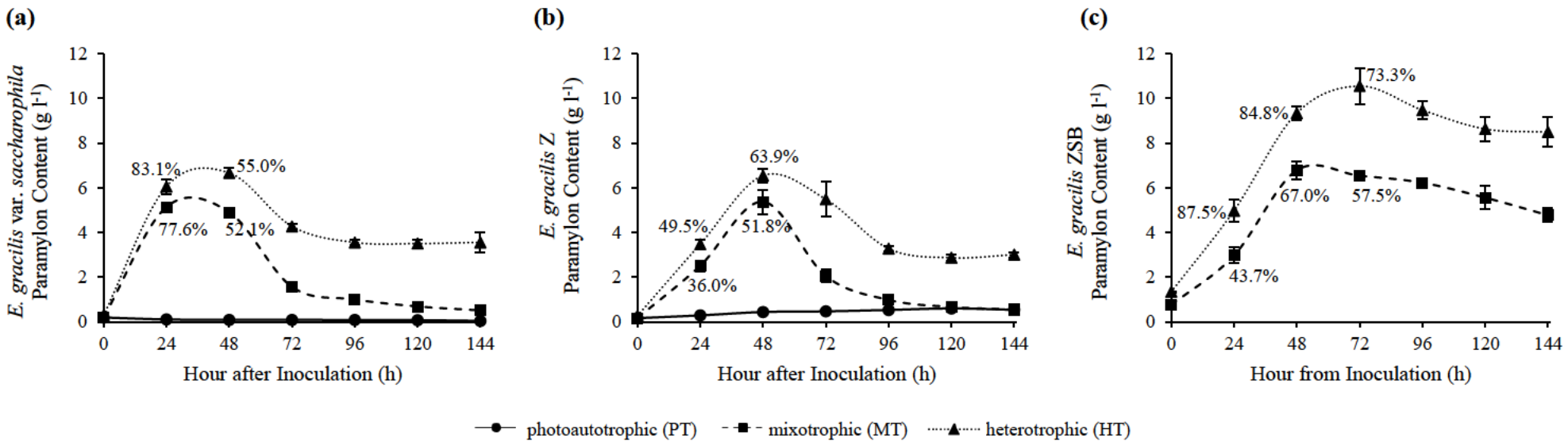

Figure 5.2: Paramylon content of (a) *E. gracilis* var. *saccharophila*, (b) *E. gracilis Z* and (c) *E. gracilis* ZSB under photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) growth conditions. The percentage indicates the amount of paramylon per dry mass. The highest paramylon content was synthesised during HT cultivation, followed by MT cultivation, while the PT cultivation had very little paramylon. *E. gracilis* var. *saccharophila* acquired its maximum paramylon content per l in a shorter period of time after inoculation (about 24-30 h) than the other two strains (about 46-54 h for all MT cultivations and HT cultivation of *E. gracilis* Z and about 72 h for HT cultivation of *E. gracilis* ZSB).

5.3.3 Analysis of label-free shotgun proteomics data

The raw data from the GPM analysis of *E. gracilis* ZSB proteins are provided in Supplementary Table S5.1. The low-quality hits from the GPM were filtered out to produce a high stringency dataset of reproducibly identified proteins present in each cultivation using the same criteria as mentioned in Chapter 3 (Section 2.10). The false discovery rates for the high-stringent datasets were lower than 1% for proteins and lower than 0.2% for peptides, and the data was not filtered further (Table 5.1). As for the other strains, about 30% of the proteins could not be annotated using a BLASTP *E*-value cut-off of $1e^{-10}$. Putative functional annotations were assigned to the proteins that could be identified by BLASTP. Only the proteins involved in paramylon metabolism are discussed below. The expression levels of the proteins for all three strains are shown by heatmaps in the metabolic pathway in Figure 5.3. The heatmaps were generated using proteins from PT cultivation as the control for proteins expressed during MT and HT cultivations of *E. gracilis* var. *saccharophila* and *E. gracilis* Z (Chapter 3 and 4 respectively), and using MT and HT cultivations of *E. gracilis* Z as the controls for the respective *E. gracilis* ZSB cultures (Section 5.2.5), to find out the comparative expression levels of paramylon metabolism-related proteins. The results from the two-sample *t*-test comparisons between the MT and HT cultures of *E. gracilis* ZSB and *E. gracilis* Z are provided in Supplementary Table S5.2.

Table 5.1: Summary of proteins identified under mixotrophic (MT) and heterotrophic (HT) cultivations for *E. gracilis* ZSB, and their corresponding FDR values.

Growth Condition[a]	No. of proteins[b]	Protein FDR[c] (%)	Peptide FDR[c] (%)
MT	4466	0.69	0.10
HT	3466	0.83	0.12

[a]MT and HT: mixotrophic and heterotrophic respectively. [b]Total number of proteins common to all three replicates. [c]FDR: false discovery rate.

5.3.4 Analysis of qRT-PCR data

The qRT-PCR reactions were carried out for *EgGSL1* and *EgGSL2* as a validation for the proteomics data obtained for the enzymes EgGSL1 and EgGSL2. The results for qRT-PCR of *EgGSL1* and *EgGSL2* (paramylon synthase transcripts), using 18S rRNA as the house-keeping gene, are shown in Table 5.2 for all three strains under the growth conditions mentioned above. The qRT-PCR analysis was carried out over 96 h of cultivation by comparing transcript levels at each time point against the initial transcript levels (controls) at 12 h for MT and HT cultivations, and 24 h for PT cultivations. Most of the negative controls did not show any amplification, while a few of them were only amplified after the 35th cycle and had a different melting temperature from the actual amplicons, which may have indicated primer dimers or non-specific binding after several cycles. Since all the negative controls were only amplified after the 35th cycle or not amplified at all, the data was considered reliable for analysis.

The least amount of variation for the *EgGSL1* transcript throughout the cultivation period was seen in the HT cultivations of all three strains with very little increase in *E. gracilis* var. *saccharophila* and *E. gracilis* Z, while there was a decrease in the transcript level for *E. gracilis* ZSB. Although there appeared to be an increase in *EgGSL1* transcript level during PT and MT cultivations, none reached even a two-fold change in the first 96 h of cultivation.

The amount of transcript for *EgGSL2* on the other hand doubled under HT cultivation within the first 24 h for *E. gracilis* var. *saccharophila* and *E. gracilis* Z, and within the first 48 h for *E. gracilis* ZSB. However, with the exception of the MT cultivation of *E. gracilis* Z at 24 h and *E. gracilis* ZSB at 48 h, none of the other light-exposed (PT and MT) cultures could double the amount of *EgGSL2* transcript within the 96 h of cultivation. Also, there was a gradual decline in the *EgGSL2* transcript level under the MT and HT cultivations of *E. gracilis* var. *saccharophila* and *E. gracilis* Z beyond 24 h, and of *E. gracilis* ZSB beyond 48 h. However, *E. gracilis* Z continued to accumulate more paramylon beyond 24 h (Figure 5.2 b) under MT and HT cultivations as mentioned in Section 5.3.2.

Table 5.2: The mRNA levels of *EgGSL1* and *EgGSL2* under photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) growth conditions in *E. gracilis* var. *saccharophila*, *E. gracilis* Z and *E. gracilis* ZSB, using 18S rRNA as the housekeeping gene to normalise systemic sample to sample variation. The expression levels were determined using the amount of transcript at 24 h as the control for PT condition and the amount of transcript at 12 h as the control for MT and HT conditions to allow sufficient time for growth of the initial inoculum. The corresponding value for each time point represents the expression level compared to the control time point. Data analysis was carried out using the $2^{-\Delta\Delta C_T}$ method [11].

mRNA		*EgGSL1*					*EgGSL2*				
Time (h)		12	24	48	72	96	12	24	48	72	96
E. gracilis var. *saccharophila*	PT		1±0	0.65±0.03	1.89±0.29	1.74±0.17		1±0	1.61±0.13	1.31±0.23	1.35±0.16
	MT	1±0	1.29±0.14	1.27±0.21	1.49±0.3	1.51±0.14	1±0	1.89±0.18	1.35±0.14	1.6±0.3	1.07±0.06
	HT	1±0	1.21±0.12	1±0.06	1.12±0.05	1.27±0.04	1±0	2.57±0.02	1.84±0.15	1.71±0.11	1.8±0.11
E. gracilis Z	PT		1±0	1.31±0.11	1.69±0.34	1.75±0.42		1±0	1.06±0.07	1.58±0.46	1.36±0.39
	MT	1±0	0.95±0.14	1.84±0.03	1.28±0.2	1.78±0.12	1±0	2.1±0.24	1.83±0.11	1.47±0.09	1.33±0.03
	HT	1±0	0.96±0.18	1.36±0.09	1.42±0.17	1.18±0.1	1±0	2.27±0.18	1.47±0.18	1.64±0.13	1.79±0.26
E. gracilis ZSB	MT	1±0	1.09±0.21	1.85±0.46	1.59±0.37	1.78±0.17	1±0	1.30±0.15	2.72±0.14	1.72±0.38	1.31±0.25
	HT	1±0	1.21±0.23	0.58±0.05	0.58±0.04	0.69±0.19	1±0	1.59±0.38	2.41±0.16	2.34±0.20	2.12±0.17

5.3.5 Differential expression of paramylon biosynthetic and degradation enzymes

The hypothetical pathway for paramylon biosynthesis and degradation has been generated using the transcriptomic data [5, 6] as well as the proteomic data obtained in this study. An isozyme of phosphoglucomutase (EC 5.4.2.2), which converts glucose-6-phosphate (G6P) from the Embden-Meyerhof-Parnas (EMP) and oxidative pentose phosphate (oxPP) pathways to glucose-1-phosphate (G1P) prior to paramylon synthesis was up-regulated under HT condition (Figure 5.3). Isoforms of UDP-glucose pyrophosphorylase (EC 2.7.7.9) that convert G1P to UDP-glucose, were found under all conditions, but were not significantly differentially expressed under the conditions studied. Paramylon synthesis is initiated with the glycosyl transfer of this UDP-glucose onto a membrane protein primer, carried out by a membrane bound enzyme complex (β-1,3-glucan synthase complex) [14]. The synthase then carries out polymerisation to form a linear β-1,3-glucan chain that produces the complex fibrillar structure of paramylon [14].

5.3.5.1 Candidate enzymes for paramylon biosynthesis

None of the sixteen glycosyltransferase (GT) family 2 candidate β-1,3-glucan synthases found in the transcriptome [6] were detected at the proteomic level under any of the conditions studied here. However, of the four sequences identified as candidate GT48 family β-1,3-glucan synthases (EC 2.4.1.34) in the transcriptome of *E. gracilis* var. *saccharophila* [6], two were detected in the proteome of all strains studied here. This has also been confirmed recently by Tanaka *et al.* (2017), who documented the existence of two glucan synthase-like proteins, EgGSL1 and EgGSL2 in *E. gracilis* Z by gene knockdown studies [7]. Compared to PT condition, both EgGSL1 and EgGSL2 were up-regulated under MT condition in *E. gracilis* Z, whereas only EgGSL2 was up-regulated under MT condition in *E. gracilis* var. *saccharophila*.

In the bleached mutant, *E. gracilis* ZSB, the MT condition supported even higher expression of EgGSL1 than the MT condition of the wild-type *E. gracilis* Z, with a fold-change of about three times, while the expression of EgGSL2 was about the same in both strains. The proteomic data show that this GT48 family isoform, EgGSL2, is the predominant isoform with higher relative abundance, which has been reported to be indispensable for paramylon synthesis in *E. gracilis* [7]. Interestingly, only EgGSL2 could be detected in the proteome of HT cultures of all strains, and this candidate enzyme was up-regulated in HT cultures compared to PT cultures of *E. gracilis* var. *saccharophila* and *E. gracilis* Z. In order to confirm the absence of EgGSL1 from the HT cultures, all the peptides identified for EgGSL1 in PT and MT cultures were searched in the raw MS/MS spectra of peptides from the HT cultures. None of these peptides could be detected in any of the HT cultures of *E. gracilis* var. *saccharophila* (Chapter 3), *E. gracilis* Z (Chapter 4) or *E. gracilis* ZSB (this chapter). The nanoLC-MS/MS was further repeated with three more biological replicates of *E. gracilis* var. *saccharophila* and *E. gracilis* Z grown under HT condition (Section 5.2.3), and EgGSL1 could still not be detected in this new set of HT replicates of either strain (Supplementary Table S5.3). These evidences confirmed that EgGSL1 was not expressed at the proteomic level under dark cultivation. Moreover, the presence of an organic carbon source (glucose) in the medium had no effect on EgGSL1 expression, as it was detected under MT cultivation but not HT cultivation, during both of which cells were exposed to glucose.

The qRT-PCR data of all strains, on the other hand, showed the presence of the *EgGSL1* transcript under HT growth condition (Table 5.2), although the expression level of the transcript did not change significantly over the period of cultivation. The presence of the transcript and absence of a translated protein probably indicates that *E. gracilis* does not regulate this protein at the transcriptional level. This has been reported for numerous other proteins in this organism, where protein expression was observed to be controlled post-transcriptionally [15-22]. An explanation for this may be that the translation of *EgGSL1*

transcript is light-induced, similar to a few other proteins transcribed from the nuclear genome of *E. gracilis* [23]. EgGSL1, therefore, must have a role in cell cultures grown under light conditions only. It is unclear what this role might be, although it may be related to negative regulation of growth [7]. Tanaka *et al.* (2017) also speculated that EgGSL1 may play a role in negative regulation of chlorophyll accumulation as knockdowns of *EgGSL1* accumulated chlorophyll at a faster rate [7]. From our proteomic analysis, however, EgGSL1 was absent in the dark when the cells lacked chlorophyll, although this does not rule out a possible role in chlorophyll accumulation. Marechal and Goldemberg (1964) reported that an inhibitor exists for paramylon synthase, especially in green cells [24]. EgGSL1 may play a role in this inhibitory regulation of EgGSL2, and this could be a possible reason why HT cells lacking EgGSL1 are able to accumulate more paramylon. The exact role of EgGSL1 for now remains unknown.

The overall increased expression of the predominant β-1,3-glucan synthase, EgGSL2, in HT cultures concurs with previous studies that showed higher specific activity of the synthase enzyme in dark condition compared to constant illumination and diurnal conditions [24]. This can also be deduced from the paramylon content in Figure 5.2, where HT cultures accumulated more paramylon than MT and PT cultures. The paramylon synthase complex consists of at least seven subunits with a total molecular mass of 670 kDa [25], and the GT48 family proteins EgGSL1 and EgGSL2 identified in this study (approximately 300 kDa each) may be part of this complex. Both EgGSL1 and EgGSL2 are much larger than any of the subunits reported [25], however, and may be proteolytically processed into the correct sized mature subunits that have been identified.

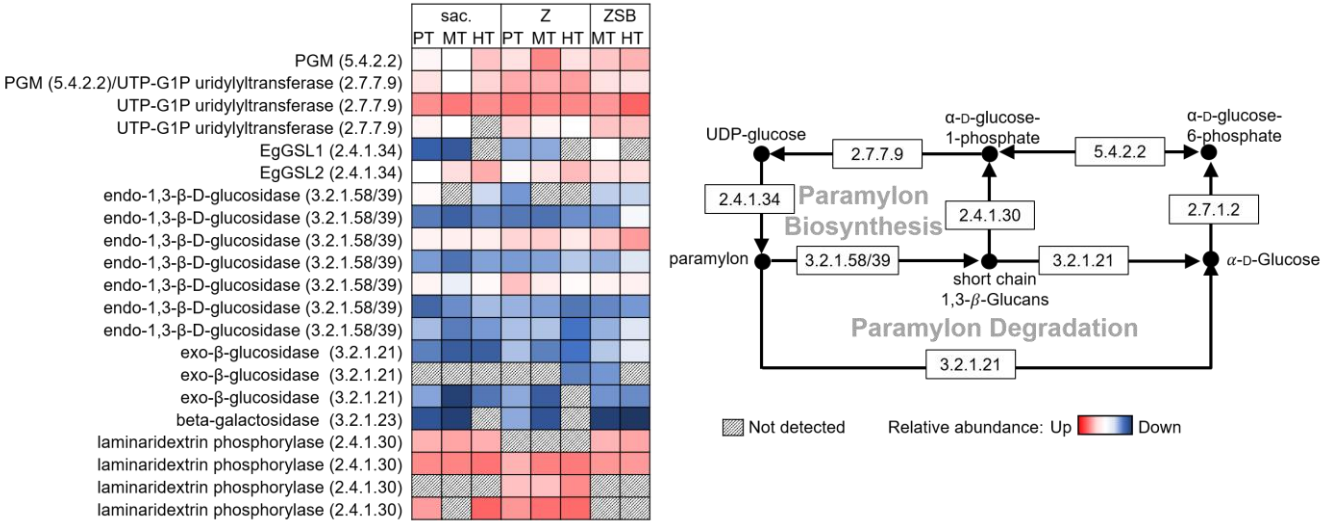

Figure 5.3: Paramylon metabolic pathway of *E. gracilis*. The heatmap shows enzyme expression under photoautotrophic (PT), mixotrophic (MT) and heterotrophic (HT) cultivations. Each row represents an isozyme for the respective enzyme class shown in brackets and the colour indicates the relative abundance of the isozyme. PT cultures were used as the control for MT and HT cultures of *E. gracilis* var. *saccharophila* (sac.) and *E. gracilis* Z (Z) from Chapters 3 and 4 respectively, while MT and HT cultures of *E. gracilis* Z (Chapter 4) were used as controls for the respective *E. gracilis* ZSB (ZSB) cultures. EgGSL1 could not be detected under HT cultivation of any of the strains. EgGSL2 on the other hand, showed high relative abundance, especially under HT cultivation. The relative abundance of EgGSL2 in *E. gracilis* ZSB was even higher than that of *E gracilis* Z under both MT and HT conditions. During mid-log phase, although several β-glucosidases (EC 3.2.1.58/39 and EC 3.2.1.31) were expressed, their overall abundance was quite low.

5.3.5.2 Candidate enzymes for paramylon degradation

Paramylon degradation in *E. gracilis* is mainly orchestrated by two groups of enzymes: glucan 1,3-β-glucosidases and glucan *endo*-1,3-β-D-glucosidases (EC 3.2.1.58 and EC 3.2.1.39), which cleave it into shorter β-1,3-glucan chains, and *exo*-β-glucosidases (EC 3.2.1.21), which further break down the chains into glucose [6]. Although a large number of both *endo*-1,3-β-D-glucosidases and *exo*-β-glucosidase candidates were also identified in the transcriptome [6], only seven of the former and four of the latter could be detected in the proteome during the mid-log phase (Figure 5.3). From the glycoside hydrolase (GH) family enzymes reported in the transcriptome, none of the GH17 family *endo*-β-glucosidases were detected in the proteome, while one member of the GH64 family and six members of the GH81 family could be detected. Only two of the seven *endo*-β-glucosidases had a higher relative abundance in *E. gracilis* ZSB, compared to *E. gracilis* Z under MT cultivation. Compared to the PT cultures, all of the *endo*-β-glucosidases were either unchanged or down-regulated in the HT cultures of *E. gracilis* Z and all but one were unchanged or down-regulated in the HT cultures of *E. gracilis* var. *saccharophila*. These results indicated that light exposure during log phase of growth increases paramylon conversion to short β-1,3-glucan chains only marginally. Several families of *exo*-β-glucosidase candidates were also predicted from the transcriptome [6], but only two proteins belonging to the GH2 family and one belonging to GH5 family could be detected in the proteome and these were either down-regulated or unchanged in presence of glucose in the medium (HT and MT conditions). The *exo*-β-glucosidase member of the GH55 family from the transcriptome could only be detected under the HT condition of the *E. gracilis* Z strain and MT condition of the *E. gracilis* ZSB strain. Similar expression patterns were seen for the *exo*-β-glucosidases as the *endo*-β-glucosidases, with most of them remaining unchanged across the growth conditions.

No phosphorylase activity has been reported in *E. gracilis* to convert paramylon directly to G1P [26]. However, paramylon can first be broken down to short chains by the *endo*-1,3-β-D-glucosidases, and the short chains can then be directly phosphorylated by laminaribiose phosphorylase (EC 2.4.1.30) to G1P. This enzyme has been purified and characterised from *E. gracilis* [27, 28], but lack of sequence similarity to laminaribiose phosphorylases from other organisms in existing databases may suggest that this enzyme belongs to a new family of laminaribiose phosphorylases [6]. The candidates for this enzyme had an overall high relative abundance across all cultures during mid-log phase, probably indicating that short glucan chains in *E. gracilis* are quickly phosphorylated to G1P, which serves as a precursor to numerous other pathways. A specific expression pattern could not be determined for the candidate laminaribiose phosphorylases from the current study, with most candidates up-regulated only in the *E. gracilis* Z strain under MT cultivation, while none of the candidates were differentially expressed in *E. gracilis* var. *saccharophila*.

The results from this study suggest that during paramylon synthesis (mid-log growth phase), some glucosidases were still expressed in the cells, but they did not affect the activity of the synthase enzymes. Marechal and Goldemberg (1964) demonstrated a similar effect by mixing snail gut juice containing β-1,3-glucanases with an extracted fraction of paramylon synthase, with no change in the synthase enzyme activity [24]. This could be owing to the inability of the glucosidases to thwart the initial binding of UDP-glucose to the membrane protein primer in the first step of paramylon synthesis, or the inability to interact with the membrane protein primer at all.

5.4 Concluding remarks

The proteomic data from this study with *E. gracilis* ZSB, as well as the previous two studies of *E. gracilis* var. *saccharophila* and *E. gracilis* Z helped to provide putative functional

annotations for enzymes involved in the paramylon biosynthetic and degradation pathways. A large pool of proteins has been identified from the transcriptome for these pathways, but this study helped to establish those that were translated at the proteomic level. *E. gracilis* ZSB produced the highest amount of paramylon per biomass under both HT (up to 87.5%) and MT (67.0%) cultivations, followed by *E. gracilis* var. *saccharophila* (up to 83.1% and 77.6% respectively), and *E. gracilis* Z (up to 63.9% and 51.8% respectively). The highest paramylon content per l was reached sooner by *E. gracilis* var. *saccharophila* (24-30 h) than the other two strains (46-54 h). The EgGSL2 enzyme, which was shown to be indispensable for paramylon synthesis in earlier gene knockdown studies of *E. gracilis* Z, was also predominant in the proteome of all strains studied here. The EgGSL1 enzyme, on the other hand, was lower in abundance in PT and MT cultures, and was not detected at all in HT cultures of any of the strains. However, qRT-PCR revealed that transcripts for both *EgGSL1* and *EgGSL2* were present under all cultivation conditions, indicating that the expression of EgGSL1 was regulated post-transcriptionally. Moreover, expression of EgGSL1 was not affected by the presence of glucose in the medium (MT cultivation) as long as light was present. This indicated that EgGSL1 expression may be induced by light, and the protein probably plays a key role in cells grown under light cultivation. The expression of some β-glucosidases during mid-log phase of growth indicated that the activity of the paramylon synthesis enzymes was not affected by β-glucosidases.

References

[1] Kataoka K, Muta T, Yamazaki S and Takeshige K. 2002. Activation of macrophages by linear (1→3)-β-D-glucans. Implications for the recognition of fungi by innate immunity. *J Biol Chem*. 277(39): 36825-36831.

[2] Chan GCF, Chan WK and Sze DMY. 2009. The effects of β-glucan on human immune and cancer cells. *J Hematol Oncol*. 2(25).

[3] Koizumi N, Sakagami H, Utsumi A, Fujinaga A, Takeda M, Asano K, Sugawara I, Ichikawa A, Kondo H, Mori S, Miyatake K, Nakano Y, Nakashima H, Murakami T, Miyano N and Yamamoto N. 1993. Anti-HIV (human immunodeficiency virus) activity of sulfated paramylon. *Antiviral Res*. 21(1): 1-14.

[4] Barsanti L, Vismara R, Passarelli V and Gualtieri P. 2001. Paramylon (β-1,3-glucan) content in wild type and WZSL mutant of *Euglena gracilis*. Effects of growth conditions. *J Appl Phycol*. 13: 59-65.

[5] Yoshida Y, Tomiyama T, Maruta T, Tomita M, Ishikawa T and Arakawa K. 2016. De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics*. 17: 182.

[6] O'Neill EC, Trick M, Hill L, Rejzek M, Dusi RG, Hamilton CJ, Zimba PV, Henrissat B and Field RA. 2015. The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Mol BioSyst*. 11(10): 2808-2820.

[7] Tanaka Y, Ogawa T, Maruta T, Yoshida Y, Arakawa K and Ishikawa T. 2017. Glucan synthase-like 2 is indispensable for paramylon synthesis in *Euglena gracilis*. *FEBS Lett*. 591(10): 1360-1370.

[8] Sumida S, Ehara T, Osafune T and Hase E. 1987. Ammonia- and light-induced degradation of paramylum in *Euglena gracilis*. *Plant Cell Physiol*. 28 (8): 1587-1592.

[9] Yokota A and Kitaoka S. 1982. Synthesis, excretion, and metabolism of glycolate under highly photorespiratory conditions in *Euglena gracilis* Z. *Plant Physiol*. 70(3): 760-764.

[10] Rodríguez-Zavala JS, Ortiz-Cruz MA, Mendoza-Hernández G and Moreno-Sánchez R. 2010. Increased synthesis of α-tocopherol, paramylon and tyrosine by *Euglena gracilis* under conditions of high biomass production. *J Appl Microbiol*. 109(6): 2160-2172.

[11] Livak KJ and Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods*. 25(4): 402-408.

[12] Dwyer MR, Smydzuk J and Smillie RM. 1970. Synthesis and breakdown of β-l,3-glucan in *Euglena gracilis* during growth and carbon depletion. *Aust J Biol Sci*. 23(4): 1005-1014.

[13] Fernandez-Valiente E and Rodriguez-Lopez M. 1970-1980. Effect of indole-acetic acid and gibberellic acid on paramylon synthesis in *Euglena gracilis*. *Microbiol Esp*. 32-33: 11–20.

[143] Tomos AD and Northcote DH. 1978. A protein-glucan intermediate during paramylon synthesis. *Biochem J*. 174(1): 283-290.

[15] Vesteg M, Vacula R, Burey S, Löffelhardt W, Drahovská H, Martin W and Krajcovic J. 2009. Expression of nucleus-encoded genes for chloroplast proteins in the flagellate *Euglena gracilis*. *J Eukaryot Microbiol*. 56(2):159-166.

[16] Hoffmeister M, van der Klei A, Rotte C, van Grinsven KW, van Hellemond JJ, Henze K, Tielens AG and Martin W. 2004. *Euglena gracilis* rhodoquinone:ubiquinone ratio and mitochondrial proteome differ under aerobic and anaerobic conditions. *J Biol Chem*. 279(21): 22422-22429.

[17] Keller M, Chan RL, Tessier LH, Weil JH and Imbault P. 1991. Post-transcriptional regulation by light of the biosynthesis of *Euglena* ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit. *Plant Mol Biol*. 17(1): 73-82.

[18] Kishore R and Schwartzbach SD. 1992. Translational control of the synthesis of the *Euglena* light harvesting chlorophyll *a*/*b* binding protein of photosystem II. *Plant Sci*. 85(1):

79-89.

[19] Madhusudhan R, Ishikawa T, Sawa Y, Shigeoka S and Shibata H. 2003. Post-transcriptional regulation of ascorbate peroxidase during light adaptation of *Euglena gracilis*. *Plant Sci*. 165(1): 233-238.

[20] Levasseur PJ, Meng Q and Bouck GB. 1994. Tubulin genes in the algal protist *Euglena gracilis*. *J Eukaryot Microbiol*. 41(5): 468-477.

[21] Saint-Guily A, Schantz ML and Schantz R. 1994. Structure and expression of a cDNA encoding a histone H2A from *Euglena gracilis*. *Plant Mol Biol.* 24(6): 941-948.

[22] Vacula R, Steiner JM, Krajcovic J, Ebringer L and Löffelhardt W. 2001. Plastid state- and light-dependent regulation of the expression of nucleus-encoded genes for chloroplast proteins in the flagellate *Euglena gracilis*. *Folia Microbiol (Praha)*. 46(5): 433-441.

[23] Schwartzbach SD. 2017. Photo and Nutritional Regulation of *Euglena* Organelle Development. In Eds. Schwartzbach SD and Shigeoka S, *Euglena: Biochemistry, Cell and Molecular Biology*. *Adv Exp Med Biol.* Springer International Publishing, Cham. 979: 159-182.

[24] Marechal LR and Goldemberg SH. 1964. Uridine diphosphate glucose-β-1,3-glucan β-3-glucosyltransferase from *Euglena gracilis*. *J Biol Chem*. 239(10): 3163-3167.

[25] Bäumer D, Preisfeld A and Ruppel HG. 2001. Isolation and characterization of paramylon synthase from *Euglena gracilis* (Euglenophyceae). *J Phycol*. 37(1): 38-46.

[26] Vogel K and Barber AA. 1968. Degradation of paramylon by *Euglena gracilis*. *J Protozool*. 15(4): 657-662.

[27] Kitaoka M, Sasaki T and Tamiguchi H. 1993. Purification and properties of laminaribiose phosphorylase (EC 2.4.1.31) from *Euglena gracilis* Z. *Arch Biochem Biophys*. 304(2): 508-514.

[28] Goldemberg SH, Maréchal LR and De Souza BC. 1966. β-1,3-oligoglucan: Orthophosphate glucosyltransferase from *Euglena gracilis*. *Biochim Biophys Acta*. 146(2): 417-430.

CHAPTER 6

Investigation of *Euglena gracilis* var. *saccharophila* as a producer of valuable metabolites during mixotrophic cultivation in a laboratory-scale bioreactor

6.1 Introduction

The freshwater microalga *Euglena gracilis* synthesises several metabolites of interest including paramylon, the vitamin antioxidants ascorbate, α-tocopherol and β-carotene and all 20 of the protein-building amino acids. The β-1,3-glucan paramylon is the primary storage carbohydrate of *E. gracilis* and has several health-enhancing properties [1-8]. The current sources of β-1,3-glucan include yeast, cell walls of fungi and plants, and algal supplements [9-11]. However, under optimal growth conditions paramylon is reported to contribute up to 90% of the biomass of *E. gracilis* [12]. This microalga can also synthesise all of the essential vitamins required by the human diet except vitamins B_1 and B_{12} [13]. Among these, ascorbate, α-tocopherol and β-carotene are the most promising for industrial production of antioxidants in *E. gracilis* to relieve oxidative stress responsible for several human diseases [14]. Only a few other organisms are known to produce such a wide range of water- and fat-soluble vitamins concurrently [15]. Additionally, this microalga synthesises all 20 of the protein-building amino acids, including eight amino acids that are essential to the human diet [16], and four amino acids that are essential for infants and growing children [17]. Free amino acids are very valuable as additives in animal feed [1]. Microalgae are also known to be good sources of proteins and under optimal growth conditions up to 60% of the biomass of *E. gracilis* can consist of proteins [1]. Scaling-up of *E. gracilis* Z in laboratory-scale bioreactors to produce paramylon, α-tocopherol and lipids has been described before [18-20]. However, simultaneous production of biomass, paramylon, ascorbate, α-tocopherol, β-carotene, protein and free amino acids, which are of specific interest in the current study has not been reported in laboratory-scale bioreactor cultures.

Laboratory-scale bioreactors are the first step towards increasing the volume of algal cultures, which is usually done by a factor of 10 per step with commercial scale bioreactors ranging from 1000 to 1 x 10^9 l as the ultimate step, depending on the purpose of the culture

[21]. Increasing the volume of algal cultures is important in that it allows large-scale production of biomass, waste-treatment and production of important metabolites to meet consumer demands [21]. However, there are many challenges associated with large-scale algal cultures. The main problems of growing algal cultures in light conditions are photoinhibition and self-shading, which result from light-induced damage to the photosystems and from dense cultures that do not allow light to penetrate through respectively [21]. The process of growing the inoculum to the correct cell density is also a critical factor in successful large-scale bioreactor cultivation [21]. Other challenges in large-scale cultures include maintaining the optimum temperature and pH, supplying CO_2 during the day to overcome the inhibition of RuBisCO by high O_2 levels from photosynthesis, cost of nutrients and their supply in optimal concentrations, negative effect on the culture of producing certain metabolites in large amounts and difficulty of recycling culture medium [21].

Before cultivating the algal cells in laboratory-scale bioreactors, testing of the strains and relevant cultivation conditions in shake-flasks have to be carried out to determine the best producer and conditions for cultivation in a bioreactor. The shake-flask cultures of *E. gracilis* var. *saccharophila*, *E. gracilis* Z and *E. gracilis* ZSB have been described in Chapters 3, 4 and 5 respectively. The ascorbate, α-tocopherol and free amino acid contents of *E. gracilis* var. *saccharophila*, and the paramylon content of all three stains have already been described for shake-flask cultures in Chapters 3 and 5 respectively. However, the analysis of the ascorbate, α-tocopherol and free amino acid contents of *E. gracilis* Z and *E. gracilis* ZSB in shake-flask cultures were not done in the previous chapters and were thus carried out here. Combined with the metabolite analysis from the shake-flask cultivations of *E. gracilis* var. *saccharophila* (Chapter 3), and the paramylon content analysis from the shake-flask cultivations of all three strains (Chapter 5), the results from the shake-flask cultivations in this chapter gave an overview of metabolite production capability of all three strains used in this project at shake-flask level.

Based on the results from the shake-flask cultures, the MT cultivation of *E. gracilis* var. *saccharophila* was chosen for bioreactor cultivation of biomass and metabolite production. This strain has the ability to accumulate more paramylon than *E. gracilis* Z (Chapter 5) and produce more antioxidants than the bleached mutant *E. gracilis* ZSB (Chapter 3 and the current chapter respectively). The MT cultivation condition was chosen as the paramylon content in PT cultures and antioxidant content in HT cultures were negligible. The aim of this part of the project was thus to investigate *E. gracilis* var. *saccharophila* as a source of metabolites of interest as food supplements.

The results for this research are reported as a thesis chapter and will be prepared for publication in the future when additional bioreactor cultivation studies have been performed to further enhance metabolite production in *E. gracilis* var. *saccharophila*.

6.2 Materials and methods

6.2.1 Sample source for shake-flask metabolite analysis

Metabolite analysis described in this chapter involved samples collected from the *E. gracilis* Z shake-flask cultures described in Chapter 4, and samples collected from the *E. gracilis* ZSB shake-flask cultures described in Chapter 5. These aliquoted samples were used to determine ascorbate, α-tocopherol and free amino acids of *E. gracilis* Z and *E. gracilis* ZSB in this chapter, as mentioned in Section 6.2.3 below. The ascorbate, α-tocopherol and free amino acid contents of *E. gracilis* var. *saccharophila* at shake-flask level have already been described in Chapter 3.

6.2.2 Bioreactor cultivation

The *E. gracilis* var. *saccharophila* strain was selected for bioreactor cultivation based on its ability to produce more paramylon than *E. gracilis* Z and more antioxidants than *E. gracilis* ZSB at the shake-flask level under MT cultivation. Batch culture was performed in GNY medium (Chapter 2, Section 2.3.1) at 23°C, with an initial pH of 3.5, in a 3 l New Brunswick BioFlo/CelliGen 115 bioreactor (working volume 2 l). The bioreactor apparatus was set up as shown in Figure 6.1. The pH was not maintained as constant but it was monitored throughout the cultivation. Two Rushton impellers were used for agitation, and non-silicone antifoam 204 (Sigma) was added as required to prevent foaming. The dissolved oxygen (DO) was set at above 30% by controlling agitation at 150-250 rpm and air flow rate at 0.2-0.8 vvm. A photoperiod of 14 h light/10 h darkness was maintained with white light illumination (2000 lx). A starved light-adapted inoculum (as described above) was used to a final concentration of 1×10^6 cell ml^{-1} of culture to achieve the same initial cell concentration used in the shake-flask cultivations. Biomass, paramylon content and unused glucose in the medium were determined at 0, 6, 12, 18, 24, 36, 48, 72, 96, 120 and 144 h after inoculation, while ascorbate, α-tocopherol and β-carotene content were determined at 0, 12, 24, 48, 72, 96, 120 and 144 h. Protein content was determined at only 24, 48 and 96 h and free amino acid content was determined at only 48 h. The experiment was repeated twice with the same strain and metabolite measurements.

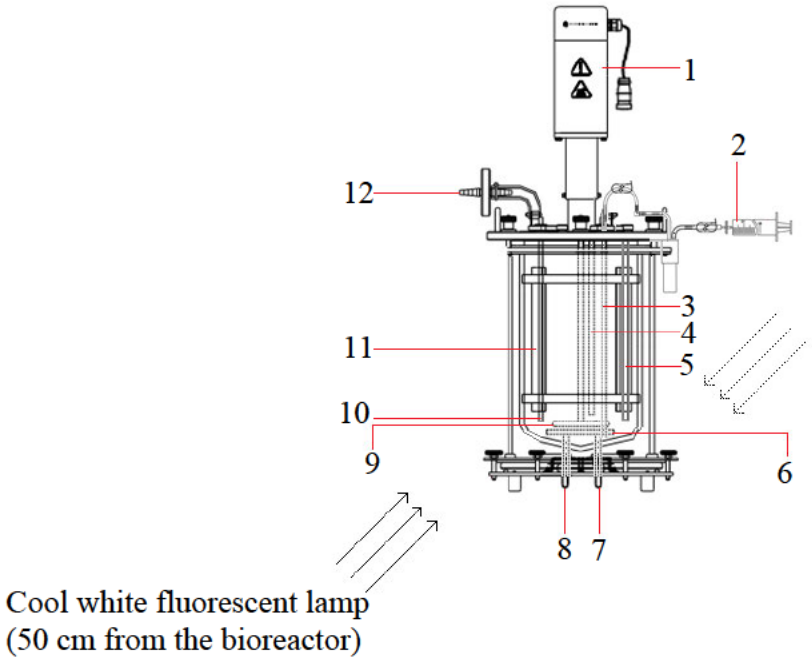

Figure 6.1: Set up of the 3 l New Brunswick BioFlo/CelliGen 115 bioreactor, showing (1) the agitation motor, (2) harvest outlet, (3) harvest tube, (4) DO probe, (5) temperature probe, (6) cooling coil, (7) cooling water inlet, (8) cooling water outlet, (9) sparger, (10) pH probe, (11) baffle and (12) exhaust. The light intensity at the surface of the bioreactor was 2000 lux.

6.2.3 Determination of dry mass, protein and metabolites

The methodologies used to determine dry mass (DM), ascorbate and free amino acid contents have already been described in detail in Chapter 3 (Sections 2.3, 2.4 and 2.6 respectively). Determination of the amount of paramylon and unused glucose in the medium have been described in Chapter 5 (Section 5.2.4). The combined extraction of α-tocopherol and β-carotene was carried out as mentioned in Chapter 2 (Section 2.4.1), and their content was measured simultaneously on the same RP-HPLC run (Chapter 3, Section 2.5). The UV absorbance of α-tocopherol was measured at 294 nm, while that of β-carotene was measured at 455 nm. The protein concentration was determined by BCA assay as described in Chapter 2 (Section 2.5.2).

6.2.4 Light microscopy

Light microscopy was used to observe cells of *E. gracilis* var. *saccharophila* at mid-log and stationary phases from the MT bioreactor cultures, and all three strains at mid-log and stationary phases from the MT shake-flask cultures (Supplementary Figure S6.1). Ten µl of a diluted and homogenised cell suspension was pipetted onto a slide and covered with a coverslip. An Olympus BX 53 bright-field compound microscope with DP26 digital camera was used to take images of the cells. The width of the cells was measured using the imaging software CellSen (V 1.14).

6.2.5 Productivity calculation

For the cultivation in the bioreactor, where mentioned, productivity was calculated using the following equation:

$$productivity = \frac{N - N_0}{T}$$

where N is the final concentration of DM or metabolite, N_0 is the initial concentration of DM or metabolite, and T is the duration in hours. The concentration of DM and paramylon was calculated in g l^{-1}, while those of protein and antioxidants were calculated in mg l^{-1} and µg l^{-1} respectively.

6.3 Results and discussion

6.3.1 Metabolites from shake-flask cultivations

The dry mass of shake-flask cultures of *E. gracilis* var. *saccharophila*, *E. gracilis* Z and *E. gracilis* ZSB has been discussed in Chapters 3, 4 and 5 respectively, and the paramylon and

unused glucose in the medium have been discussed in Chapter 5 (Figure 5.2 and Supplementary Figure S5.1 respectively). The amounts of ascorbate, α-tocopherol and free amino acids of PT, MT and HT shake-flask cultures of *E. gracilis* var. *saccharophila* have also been described in Chapter 3. The aim of this section was to describe the production of ascorbate, α-tocopherol and free amino acids in shake-flask cultures of *E. gracilis* Z and *E. gracilis* ZSB, which have not been discussed in the previous chapters, and compare them to the same metabolites produced by *E. gracilis* var. *saccharophila,* which is reported in Chapter 3.

The amounts of ascorbate and α-tocopherol in shake-flask cultures of *E. gracilis* Z and *E. gracilis* ZSB are shown in Figure 6.2. The highest amount of antioxidants per DM was found in *E. gracilis* Z under PT cultivation. However, since the DM of PT cultivated *E. gracilis* Z cells was quite low (Chapter 4, Figure 1), the antioxidant concentration produced per l of culture was lower than that in MT cultures, which yielded a much higher biomass. Comparatively, very low amounts of antioxidants were produced during HT cultivation. The ascorbate and α-tocopherol contents of *E. gracilis* Z under MT cultivation were up to 2.71 and 2.16 mg g^{-1} DM respectively, and were much higher than that of *E. gracilis* var. *saccharophila* (0.97 and 1.4 mg g^{-1} respectively, Chapter 3 Sections 3.4 and 3.5). On the other hand, the ascorbate and α-tocopherol concentrations of the bleached mutant *E. gracilis* ZSB under MT cultivation were only 0.61 and 1.01 mg g^{-1} DM respectively. Thus, in terms of antioxidant content, *E. gracilis* Z under MT cultivation performed the best among the three strains at shake-flask level.

The free amino acid contents of *E. gracilis* Z and ZSB during shake-flask cultivation are shown in Table 6.1. In terms of the concentration of total free amino acids, these were very similar to shake-flask cultures of *E. gracilis* var. *saccharophila* (Chapter 3, Table 3), with the latter producing marginally more free amino acids at mid-log phase than the other two strains. However, of the eight essential amino acids required in the human diet, *E. gracilis* var. *saccharophila* had the highest amount of tryptophan, valine, isoleucine and leucine. Of the

additional four essential amino acids required by infants and growing children, this strain had the highest amount of cysteine and arginine. The tryptophan contents of *E. gracilis* Z and ZSB were still increasing at 48 h and 72 h respectively, whereas the tryptophan content of *E. gracilis* var. *saccharophila* already began to decline by 48 h (Chapter 3, Table 3). This could have partly influenced the faster paramylon accumulation of *E. gracilis* var. *sacchrophila*, as one of the many products of tryptophan metabolism is indoleacetate (IA), which enhances growth and paramylon accumulation [22].

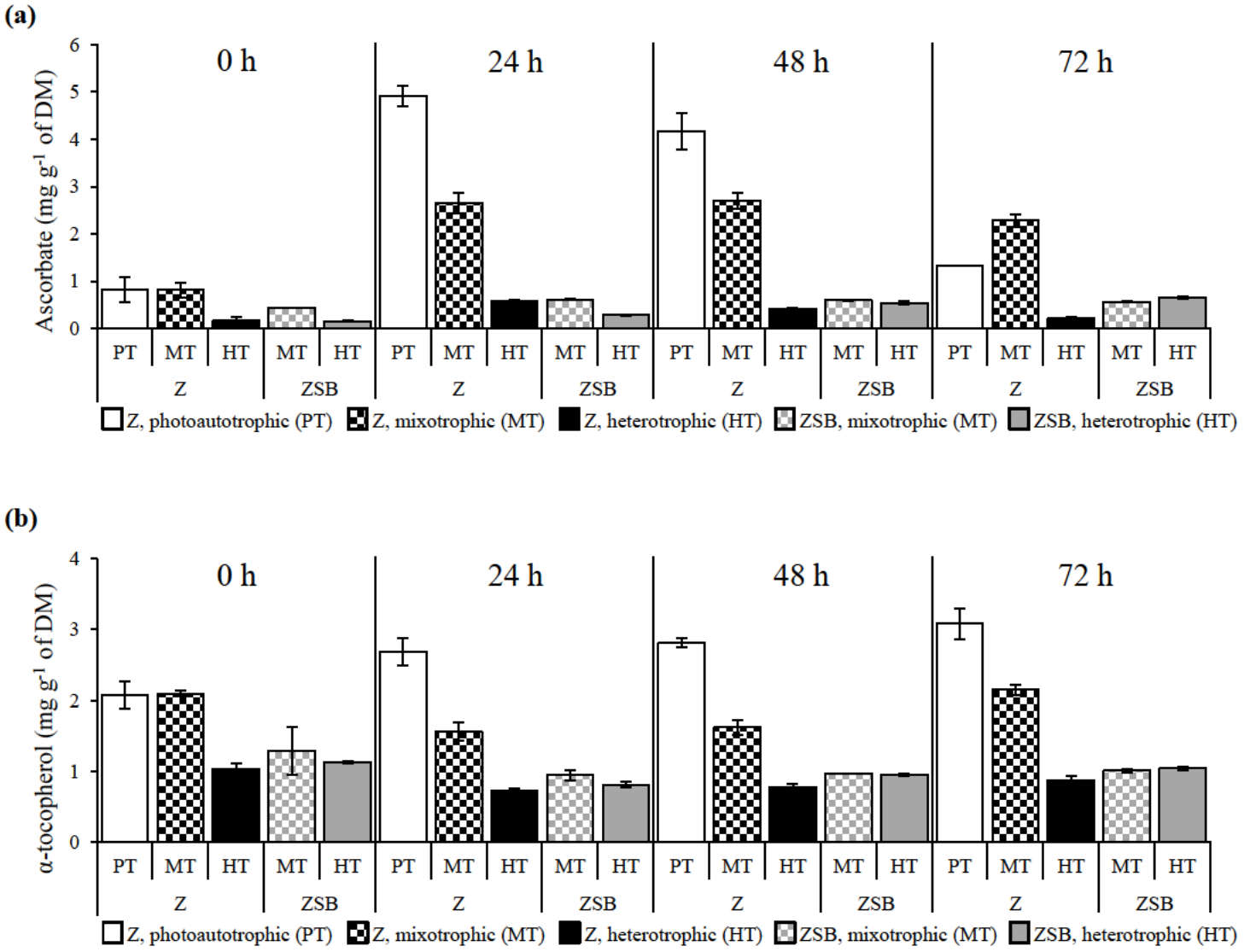

Figure 6.2: The contents of (a) ascorbate and (b) α-tocopherol in mg g^{-1} of dry mass (DM) in the shake-flask cultures of *E. gracilis* Z and *E. gracilis* ZSB. 0 h indicates the time of inoculation and 72 h indicates the beginning of the stationary phase. The error bars show standard deviation between triplicates.

Footnote: *E. gracilis* Z did not show a comprehensible stationary phase during mixotrophic (MT) cultivation, and by 72 h the MT cells were already beginning to decrease in DM as shown in Chapter 4 (Figure 1).

Amino Acid	Z (mg g^{-1} DM)						ZSB (mg g^{-1} DM)			
	PT		MT		HT		MT		HT	
	48h	72h	24h	48h	24h	48h	36h	72h	36h	72h
Trp	0.18[a]	0.18[a]	0.25[a]	0.37[a]	0.28	0.47[a]	0.51[a]	0.58[a]	0.54	0.60
Phe	0.44[a]	0.18	1.23[b]	0.73[a]	1.38[b]	0.43[a]	0.87[a]	1.26	0.89[a]	1.15
Tyr	0.27	0.94[a]	0.13[a]	0.93[a]	0.10[a]	0.09	0.28[a]	0.20	0.24	0.26
Gly	0.26[a]	0.25	0.21[a]	0.24[a]	0.20[a]	0.19	0.18	0.17	0.21[a]	0.20
Ser	0.59[a]	0.56[a]	0.29[a]	0.19[a]	0.09[a]	0.06[a]	0.17	0.19	0.09	0.09
Thr	0.56[a]	0.46	0.37[a]	0.16[a]	0.32[a]	0.05	0.18	0.18	0.18	0.14
Val	0.24[a]	0.16[a]	0.43[a]	0.22[a]	0.41[a]	0.11	0.48[a]	0.29[a]	0.47[a]	0.18
Ile	0.19[a]	0.22[a]	0.16	0.34[a]	0.13	0.38[a]	0.18[a]	0.33[a]	0.18	0.31[a]
Leu	0.18[a]	0.12[a]	0.35[a]	0.09[a]	0.38[a]	0.05	0.29[a]	0.07	0.29[a]	0.04
Met	0.54[a]	0.11[a]	0.09	0.04	0.06	0.03	0.05	0.04	0.04	0.03
Cys	0.73[a]	0.83[a]	0.54[a]	0.72[a]	0.60[a]	0.65[a]	0.70[a]	0.94[a]	0.75[a]	0.83
Asp	2.81[b]	2.79[b]	2.39[b]	5.62[b]	2.28[b]	4.77[b]	2.21[a]	3.20	2.20	3.20[a]
Asn	7.77[b]	2.52[b]	10.10[b]	0.62[b]	9.49[b]	1.91[a]	9.53[a]	3.04[a]	8.89[b]	3.29[a]
Glu	0.86[a]	0.84[a]	0.82[a]	0.85[a]	0.80[a]	0.79[a]	0.88[a]	0.86[a]	1.15	1.13
Gln	2.93[a]	0.69[a]	2.52[a]	0.95[b]	1.43[a]	0.53[a]	2.17[a]	1.28[a]	1.37[a]	1.28[a]
Ala	0.32[a]	0.39[a]	1.09[a]	2.67[b]	0.75[a]	0.92[a]	2.39[a]	5.53[b]	2.30	4.36[a]
Arg	18.40[b]	14.07[b]	14.88[b]	13.11[b]	15.65[c]	14.90[b]	16.29[b]	15.49[b]	16.18[a]	16.02
Pro	0.21[a]	0.17	0.19	0.19[a]	0.24[a]	0.24	0.20	0.18	0.26	0.36[a]
Lys	1.77[a]	2.21[a]	2.17[b]	0.63[a]	4.97[b]	1.08[a]	0.77[a]	0.75[a]	0.94[a]	0.93
His	5.21[a]	3.07[b]	4.05[b]	2.96[b]	3.78[b]	2.41[b]	3.69[a]	1.56[a]	3.15[a]	1.27[a]
Total	44.46	30.77	42.27	31.63	43.33	30.06	42.00	36.14	40.30	35.65

[a] SD < ± 0.0985. [b] SD < ± 0.996. [c] SD = ± 1.14. All other SD < ± 0.00986.

Table 6.1: Concentration of free amino acids (mg g^{-1} of DM) in shake-flask cultures of *Euglena gracilis* Z and ZSB, during the mid-log and beginning of the stationary phases. The mid-log phases were at 48 h for PT cultures and 24 h for MT and HT cultures of *E. gracilis* Z, and at 36 h for MT and HT cultures of *E. gracilis* ZSB. The beginning of the stationary phases were at 72 h for PT and HT cultures and 48 h for MT cultures of *E. gracilis* Z, and at 72 h for MT and HT cultures of *E. gracilis* ZSB.

6.3.2 Bioreactor cultivation of *E. gracilis* var. *saccharophila*

The aim of this section was to select the most suitable strain and growth condition for simultaneous production of biomass, paramylon, ascorbate, α-tocopherol and free amino acids from shake-flask cultivations. A summary of the performance of all strains at the shake-flask level is shown in Table 6.2, in respect to biomass, metabolite production, duration to reach highest dry mass, and duration to take up 90% of the glucose from the medium (applicable for MT and HT cultivations only).

Although *E. gracilis* Z produced the highest concentration of antioxidants among the strains (Figure 6.2) and *E. gracilis* ZSB produced the highest amount of paramylon among the strains (Chapter 5, Figure 5.2 c) at shake-flask level, the paramylon content of *E. gracilis* Z was the lowest among the strains (Chapter 5, Figure 5.2 b) and the antioxidant contents of *E. gracilis* ZSB were very low (Figure 6.2). Based on the results from the shake-flask cultures, *E. gracilis* var. *saccharophila* was thus selected for laboratory-scale bioreactor cultivation, as it produced high amounts of paramylon during both MT and HT cultivations (Chapter 5, Figure 5.2 a) and relatively high amounts of antioxidants during PT and MT cultivations (Chapter 3, Figures 3 and 4). This was also the fastest growing strain (Chapter 3, Figure 1) and could take up glucose from the medium more rapidly than the other two strains (Chapter 5, Supplementary Figure S5.1). The bioreactor cultivation was done under MT cultivation to offset the low antioxidant contents of HT cultures cells and the low DM and paramylon content of PT cells. Bioreactor cultivation of this *E. gracilis* strain has not been reported in the literature before. The selected strain and cultivation condition was then grown in a laboratory-scale bioreactor to further assess its performance as a producer of biomass and metabolites. In addition, the duration of culture for optimum production of biomass and each metabolite was also determined.

Table 6.2: Ranking of the strains at the shake-flask level grown under the different growth conditions in terms of dry mass, metabolite production, fastest growth rate and fastest uptake of glucose (applicable for MT and HT conditions only). The numbers in bold represent the rankings from 1 to 8, while the numbers in brackets represent the actual amount of dry mass, metabolite, duration to maximum growth and duration to take up 90% of the glucose from the medium, and the red number indicates the best candidate. The average score indicates that *E. gracilis* var. *saccharophila* MT and HT cultivations rank the best overall. The MT cultivation was chosen as negligible antioxidants were produced during HT cultivation.

	E. gracilis* var. *saccharophila			***E. gracilis* Z**			***E. gracilis* ZSB**	
	PT	MT	HT	PT	MT	HT	MT	HT
Dry mass (g l^{-1})	**8** (2.38±0.29)	**6** (9.42±0.29)	**2** (12.91±0.19)	**7** (3.53±0.26)	**5** (10.38±0.65)	**3** (11.86±0.69)	**4** (11.38±0.17)	**1** (14.39±0.43)
Metabolite:								
Paramylon (mg g^{-1})	**8** (13.86±0.52)	**3** (77.65±1.96)	**2** (83.10±3.22)	**7** (16.96±1.12)	**6** (51.77±2.07)	**5** (63.89±2.39)	**4** (66.98±3.47)	**1** (87.49±10.36)
Ascorbate (mg g^{-1})	**2** (4.13±0.59)	**4** (0.97±0.04)	**8** (0.50±0.09)	**1** (4.92±0.22)	**3** (2.71±0.16)	**7** (0.58±0.03)	**6** (0.61±0.01)	**5** (0.65±0.02)
α-Tocopherol (mg g^{-1})	**2** (2.52±0.22)	**4** (1.59±0.09)	**8** (0.64±0.01)	**1** (3.09±0.22)	**3** (2.16±0.07)	**7** (1.03±0.07)	**5** (1.29±0.33)	**6** (1.13±0.02)
Overall free amino acids (mg g^{-1})	**1** (45.64±1.23)	**4** (43.44±0.49)	**2** (44.72±3.88)	**3** (44.46±1.06)	**6** (42.27±2.57)	**5** (43.33±2.88)	**7** (42.00±0.48)	**8** (40.30±0.26)
Shortest duration to reach highest dry mass (h)	**3** (96)	**1** (48)	**1** (48)	**3** (96)	**1** (48)	**2** (72)	**2** (72)	**2** (72)
Shortest duration to uptake 90% of the glucose from the medium (h)	**NA**	**2** (48 to 72)	**1** (48)	**NA**	**5** (96)	**5** (96)	**4** (72 and 96)	**3** (72)
Average Ranking Score	**4.00**	**3.43**	**3.43**	**3.67**	**4.14**	**4.86**	**4.57**	**3.71**

6.3.2.1 Growth and paramylon production

The growth curve, paramylon content, unused glucose in the medium and pH of culture for the MT cultivation of *E. gracilis* var. *saccharophila* in a bioreactor are shown in Figure 6.3. The DM at 48 h (12.06 g l^{-1}) was higher than the DM of the corresponding MT cultures of all strains in shake-flasks. Although HT cultivations are deemed to be optimal for biomass accumulation, the DM obtained here was comparable to some HT batch cultivations reported in the literature (about 13 g l^{-1}) [18-19], but lower than other HT fed-batch cultivations and MT batch cultivations of *E. gracilis* Z supplemented with corn steep liquor and additional CO_2 [19-20].

The dry mass productivity calculated for the first 48 h was 0.22 g l^{-1} h^{-1}, but reduced beyond 48 h, as the stationary phase was not maintained by replenishing the medium with nutrients. The dry mass productivity may have been influenced by the concentration of glucose, temperature, light intensity and sparging. The concentration of glucose and temperature were selected based on experiments carried out at the shake flask level, which showed 15 g l^{-1} and 23°C to be the optimum for rapid growth (data not shown). The effect of light intensity on the culture was not determined for this study. The light intensity parameter was instead based on values obtained from the literature which showed optimum growth and antioxidant accumulation at 2000 lux [23-26]. Rapid multiplication of cells as seen here can also result in self-shading where light is prevented from penetrating through the bioreactor by the high cell density. This can lead to reduction of productivity during photosynthesis, even though the baffles move the cells around the bioreactor to ensure they receive some light. Sparging was carried out with air rather than plain carbon dioxide or a mixture of air and carbon dioxide. This may have affected the dry mass productivity as carbon dioxide may have been the limiting factor of photosynthesis for the MT cells.

The paramylon content obtained with the bioreactor cultivation was up to 9.49 g l^{-1} at 36 h (80.9% of DM), which was comparable to HT cultures of the bleached-mutant strain *E. gracilis* ZSB grown in shake-flasks (10.55 g l^{-1}, Chapter 5, Figure 5.2 c). Although more paramylon per DM was obtained at 24 h, the DM at this time point was significantly lower than that at 36 and 48 h (Figure 6.3). The paramylon levels achieved in this study were higher than reported in many other studies carried out under MT cultivations of *E. gracilis* Z [12, 27], but lower than those carried out under HT cultivations of bleached mutants, which can be up to 90% of the total DM at time points of highest DM accumulation [12]. The paramylon content here already begins to decline by the time point of highest DM accumulation (48 h).

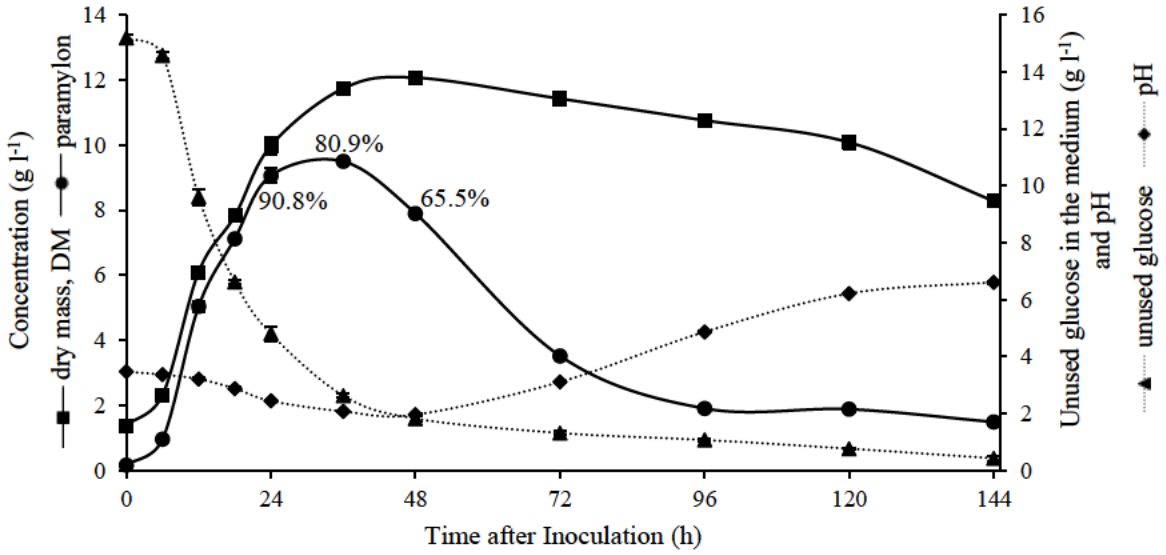

Figure 6.3: Concentration of biomass, paramylon and unused glucose in the medium, and culture pH of *E. gracilis* var. *saccharophila* grown under mixotrophic (MT) cultivation in a laboratory-scale bioreactor for 144 h.

The microscopic images comparing the cells during mid-log and stationary phases show a difference in the number of paramylon granules and chloroplasts, with very few mature chloroplasts present during mid-log phase (Figure 6.4). Microscopic images of the shake-flask cultures of all strains grown under MT cultivation during mid-log phase are shown in

Supplementary Figure S6.1. Both the shake-flask cultures and the bioreactor cultivations of *E. gracilis* var. *saccharophila* (Supplementary Figure S6.1 a and Figure 6.4 respectively) had fewer chloroplasts than *E. gracilis* Z grown in shake-flasks (Supplementary Figure S6.1 c) at mid-log phase. This may indicate why *E. gracilis* var. *saccharophila* was able to accumulate more paramylon than *E. gracilis* Z, as one of the main purposes of paramylon catabolism is to provide energy for chloroplast development [28]. Since *E. gracilis* var. *saccharophila* develops chloroplasts slowly, it is able to retain higher intracellular paramylon contents for a longer period of time during MT cultivation (Figure 6.4). The paramylon productivity was very high for the first 36 h at 0.26 g l^{-1} h^{-1}, but plummeted to just 0.16 g l^{-1} h^{-1} by 48 h, as the medium was depleted of glucose.

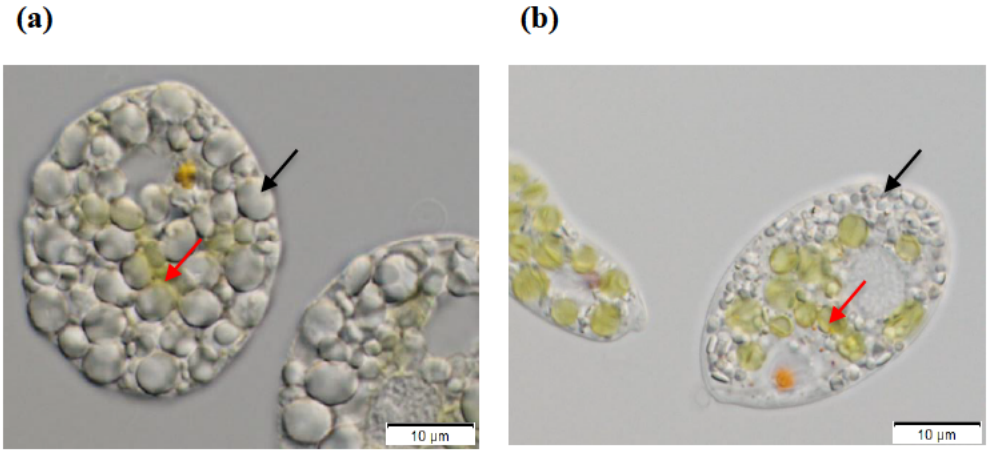

Figure 6.4: Microscopic images of *E. gracilis* var. *saccharophila* cells grown in a bioreactor under MT condition, during (a) mid-log phase at 24 h and (b) at the end of cultivation, 144 h. A large amount of paramylon was accumulated by this strain during mid-log phase (a). Even by the end of cultivation at 144 h (b), *E. gracilis* var. *saccharophila* retained a lot of its paramylon granules. Also, this strain had a very low number of mature chloroplasts during paramylon accumulation at mid-log phase (a). The black arrows indicate paramylon granules, while the red arrows indicate chloroplasts.

Glucose was consumed from the medium rapidly (Figure 6.3), and almost completely diminished within 48 h (about 88% consumed). In fed-batch cultures, it is possible to supply glucose to the medium for continued biomass and paramylon accumulation, and such a strategy has shown to enhance biomass accumulations to 48 g l^{-1} for *E. gracilis* Z [19]. In the batch cultivation carried out here, biomass could not be maintained at its highest beyond 48 h. A decrease in the pH of the medium from 3.5 to about 2 was observed in the exponential phase (during paramylon accumulation), after which it rose to above 6 in the stationary phase (during paramylon degradation) as shown in Figure 6.3. The rapid uptake of glucose from the medium and drop in pH observed in the first 36 h combined with the rapid increase in biomass and paramylon levels indicate that the MT culture was behaving like a HT culture during the first 36 h of cultivation when there was abundant glucose present in the medium. Beyond 36 h, as the glucose was depleted the MT culture behaved more like a PT culture, breaking down the paramylon to provide energy for chloroplast development and photosynthesis. The breakdown of paramylon also resulted in the cells shrinking in size (Figure 6.4). The biomass appeared to decline although this could be because the new cells that were generated were just smaller in size. Development of the chloroplast and initiation of photosynthesis also resulted in the generation of higher amounts of reactive oxygen species (ROS) and production of antioxidants as described below in Section 6.3.2.2.

6.3.2.2 Vitamin antioxidant concentrations

The ascorbate, α-tocopherol and β-carotene contents of *E. gracilis* var. *saccharophila* in the bioreactor cultivation are shown in Figure 6.5. Among the antioxidant vitamins in *E. gracilis* var. *saccharophila*, ascorbate concentration was the highest, followed by β-carotene, and α-tocopherol.

6.3.2.2.1 Ascorbate

Although higher ascorbate concentration was obtained in the bioreactor cultivation than in the shake-flask cultivation under MT condition, the amount of ascorbate per g of DM was still lower than that of shake-flask PT cultures of *E. gracilis* Z (Figure 6.2) and *E. gracilis* var. *saccharophila* (Chapter 3, Figure 3). The ascorbate concentration increased at 46.2 $\mu g\ l^{-1}\ h^{-1}$ as the glucose in the medium was depleted and the cells began to rely more on photosynthesis, which in turn led to the production of more reactive oxygen species (ROS). There is no direct link known to exist between photosynthesis and chloroplast development, and the regulation of ascorbate biosynthesis in *E. gracilis* [13]. However, the production of ROS during photosynthesis may be the reason for the enhanced synthesis of ascorbate during the stationary phase. Beyond 72 h, the ascorbate productivity declined as the DM started to decrease significantly. The amount of ascorbate in *E. gracilis* produced here is much lower than that found in commercial food sources such as fruits, fruit juices and spinach [29-32], but higher than edible seaweeds [33-34].

6.3.2.2.2 Alpha-tocopherol

The α-tocopherol concentration of *E. gracilis* var. *saccharophila* in the bioreactor cultivation was higher than that in the shake-flask MT cultures and comparable to that in the shake-flask PT cultures (Chapter 3, Figure 4). However, the concentration was lower than that of shake-flask PT cultures of *E. gracilis* Z (Figure 6.2). The amount of α-tocopherol continued to increase throughout the 144 h of cultivation. This was not as high as the α-tocopherol per g of DM reported in the literature for *E. gracilis* Z and bleached mutants of *E. gracilis* grown under PT and MT conditions [1, 27, 35-36], but was higher per l of culture as the biomass per l acquired in this study for MT cultivation was higher. Moreover, the α-tocopherol

concentration here was also much higher than that of 56 other genera of microorganisms reported including several microalgae [35]. The content was also higher than in the commercially consumed sources of α-tocopherol including the most popular source, vegetable oils (0.497, 0.268, 0.125 and 0.075 mg g^{-1} for sunflower, corn, canola and soybean oils respectively) [37]. Addition of ethanol to the medium is reported to increase the concentration of α-tocopherol by inducing oxidative stress [19], but this was not explored in the current study. A positive correlation between chloroplast development and tocopherols has been suggested in the literature [13]. This may be the reason for the increase in α-tocopherol concentration as the cultivation progressed since chloroplast development in this strain is slow (Figure 6.5). This positive correlation can be further deduced from the increase in α-tocopherol productivity occurring only after 72 h of cultivation (Figure 6.5).

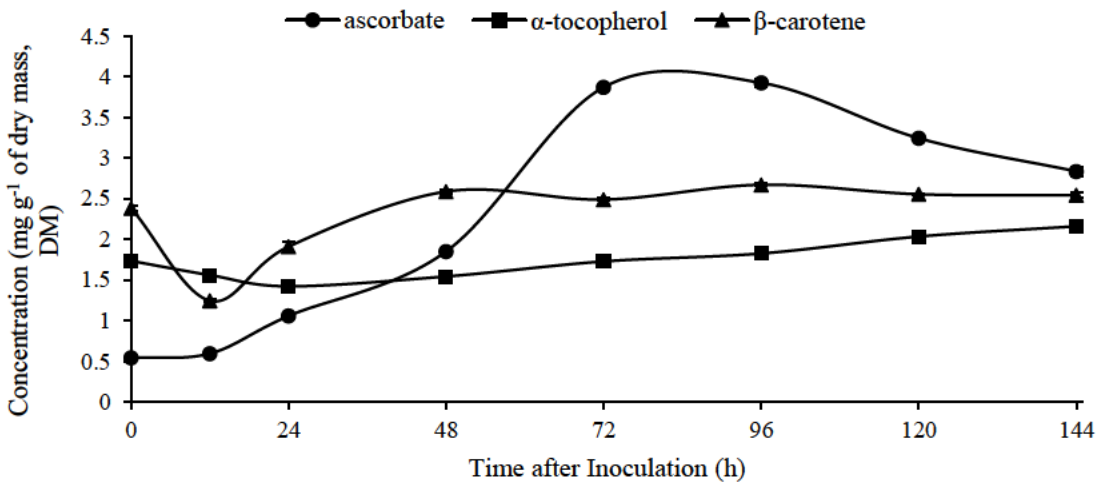

Figure 6.5: Concentration of ascorbate, α-tocopherol and β-carotene in mg g^{-1} of dry mass (DM) in *E. gracilis* var. *saccharophila* grown under mixotrophic (MT) cultivation in a bioreactor. Ascorbate productivity was the highest among the vitamin antioxidants, followed by β-carotene, while that of α-tocopherol only started to increase after 72 h.

6.3.2.2.3 Beta-carotene

The β-carotene concentration was not measured from the shake-flask cultures but was measured during the bioreactor cultivation (Figure 6.5). Beta-carotene takes part in photosynthesis by binding peptides to form pigment-protein complexes in chloroplasts [13]. This may explain why the β-carotene content declined during the first hours of cultivation as the cells consumed glucose and accumulated paramylon but had not developed many mature chloroplasts yet (Figures 6.3 and 6.4). Beta-carotene is reported to be the predominant provitamin A compound under MT cultivation of *E. gracilis* Z [38]. A wide range of β-carotene concentrations have been reported with up to 0.706 mg g^{-1} for *E. gracilis* Z under MT cultivation [38], 2.90 mg g^{-1} for *E. gracilis* var. *bacillaris* under MT cultivation with ethanol supplementation [39], and 3.24 mg g^{-1} for *E. gracilis* Z in a two-step batch culture involving MT followed by PT cultivation [15]. Although the content reported here (up to 2.67 mg g^{-1}) is not the highest, it is comparable to these reports.

6.3.2.3 Protein and free amino acid contents

The protein content was measured only at 24, 48 and 96 h as shown in Table 6.3. As the cellular paramylon content decreased (Figure 6.3), the protein content increased from 8.05 to 59.33%. Microalgae are considered high-protein food supplements, but the paramylon content of *E. gracilis* var. *saccharophila* was so high during mid-log and at the beginning of the stationary phases that the concentration of proteins was rendered low as a direct consequence, reaching only about 0.59 g g^{-1} of DM by the end of 96 h. However, by 96 h the paramylon content was only about 0.18 g g^{-1} of DM (Figure 6.3).

Table 6.3: Total protein content of *E. gracilis* var. *saccharophila* cultivated under MT condition in a bioreactor.

Hour	Protein (g l^{-1})	Protein (g g^{-1} DM)
24	0.803	0.081
48	3.542	0.294
96	6.381	0.593

The free amino acid content of the *E. gracilis* var. *saccharophila* bioreactor culture at 48 h (time point of highest DM) are shown in Table 6.4, along with those reported in the literature for a spontaneously bleached mutant of *E. gracilis* Z, beef, soybean, and the commercially available microalga *Spirulina* [1, 40-42]. Of the eight essential amino acids in the human diet, only tryptophan, phenylalanine and lysine amounts were higher in *E. gracilis* var. *saccharophila* than in the other commercial sources. However, the amounts of all the essential amino acids were lower than those reported in the literature for the spontaneously bleached mutant of *E. gracilis* Z [1]. Of the four additional amino acids required by infants and growing children, the amounts of all except tyrosine were higher in *E. gracilis* var. *saccharophila* in this study than the commercial sources reported in the literature [40-42], with a two-fold higher arginine content to that reported for the spontaneously bleached mutant of *E. gracilis* Z [1]. Although the concentrations of free amino acids in the current study are not as high as those obtained from enhanced strains of bacteria [43-47], all 20 of the protein-building amino acids can be obtained from this microalga, rendering it a single source of multiple amino acids. Adding ethanol to the medium to further improve the yield of free amino acids essential to both adults and children [1] was not tried in the current study.

Table 6.4: Free amino acid content (mg g^{-1} of DM) of *E. gracilis* var. *saccharophila* (current study) at 48 h, compared to some sources reported in the literature. The amino acids in bold indicate free amino acids that are essential to adults and children.

	Free amino acids (mg g^{-1} of DM)				
Amino Acid	*E. gracilis* var. *saccharophila* (current study)	*E. gracilis* Z* [1]	beef ** [40]	soybean [41]	*Spirulina platensis* [42]
Trp	**0.32[a]**	**0.70[a]**	**0.07**	**NR**	**ND**
Phe	**1.04[a]**	**1.20[b]**	**0.35**	**0.05[a]**	**0.53[a]**
Tyr	**0.28**	**23.00[c]**	**0.24**	**0.03[a]**	**1.50[b]**
Gly	0.22	2.20[b]	0.46[a]	0.06[a]	0.07[a]
Ser	0.35[a]	0.50[b]	0.37[a]	0.22[a]	0.24[b]
Thr	**0.34[a]**	**3.00[c]**	**0.39[a]**	**ND**	**0.58[a]**
Val	**0.43**	**3.60[c]**	**0.28**	**ND**	**0.70[a]**
Iso	**0.15[a]**	**1.10[b]**	**0.24**	**ND**	**0.79[a]**
Leu	**0.29**	**0.80[b]**	**0.40[a]**	**0.02[a]**	**0.66[a]**
Met	**0.07**	**NR**	**0.18**	**ND**	**ND**
Cys	**1.30[a]**	**9.10[a]**	**0.36[a]**	**NR**	**ND**
Asp	6.39[b]	1.00[b]	0.07	0.16[a]	3.30[b]
Asn	3.30[b]	6.00[c]	0.21	NR	NR
Glu	1.08[a]	6.00[c]	0.80[a]	0.19[a]	1.90[b]
Gln	1.62[a]	3.00[c]	5.07[b]	NR	NR
Ala	2.61[a]	45.00[c]	1.25[a]	0.09[a]	0.33[a]
Arg	**18.03[a]**	**9.00[b]**	**0.17**	**0.12[a]**	**1.00[b]**
Pro	0.24	4.00[b]	0.54[a]	ND	2.40[b]
Lys	**1.20**	**1.70[b]**	**0.35[a]**	**ND**	**0.69[b]**
His	**3.37[a]**	**6.40[c]**	**0.34[a]**	**0.05[a]**	**ND**

*spontaneously bleached mutant [1]
** beef loin cut, which had the highest free amino acid content among beef cuts studied [40]
[a]SD < ± 0.09; [b]SD < ± 0.8; [c]SD < ± 11; All other SD < ± 0.009
ND = not detected; NR = not reported

6.4 Concluding remarks

E. gracilis var. *saccharophila* was chosen for cultivation in a laboratory-scale bioreactor from the strains tested for metabolite production at the shake-flask level. MT cultivation condition was selected as HT cultivations showed very low accumulation of antioxidants at the shake-flask level, while PT cultivations did not acquire enough biomass and paramylon. In having both characteristics of HT and PT cultures, the MT cultures were able to support the production of large amounts of paramylon, and concurrently accumulate antioxidants by producing ROS via photosynthesis. Although *E. gracilis* var. *saccharophila* did not produce the highest of any of the metabolites studied compared to the literature, it still showed great promise as a simultaneous producer of several compounds, when cultivated under MT condition, and may provide an alternative means of producing metabolites of importance as nutraceuticals and animal feed.

References

[1] Rodríguez-Zavala JS, Ortiz-Cruz MA, Mendoza-Hernández G and Moreno-Sánchez R. 2010. Increased synthesis of α-tocopherol, paramylon and tyrosine by *Euglena gracilis* under conditions of high biomass production. *J Appl Microbiol*. 109(6): 2160-2172.

[2] Kataoka K, Muta T, Yamazaki S and Takeshige K. 2002. Activation of macrophages by linear (1→3)-β-D-glucans. Implications for the recognition of fungi by innate immunity. *J Biol Chem*. 277(39): 36825-36831.

[3] Kondo Y, Kato A, Hojo H, Nozoe S, Takeuchi M and Ochi K. 1992. Cytokine-related immunopotentiating activities of paramylon, a β-(1-->3)-D-glucan from *Euglena gracilis*. *J Pharmacobiodyn*. 15(11): 617-621.

[4] Xiao Z, Trincado CA and Murtaugh MP. 2004. β-glucan enhancement of T cell IFNγ response in swine. *Vet Immunol Immunopathol*. 102(3): 315-320.

[5] Koizumi N, Sakagami H, Utsumi A, Fujinaga A, Takeda M, Asano K, Sugawara I, Ichikawa A, Kondo H, Mori S, Miyatake K, Nakano Y, Nakashima H, Murakami T, Miyano N and Yamamoto N. 1993. Anti-HIV (human immunodeficiency virus) activity of sulfated paramylon. *Antiviral Res*. 21(1): 1-14.

[6] Chan GCF, Chan WK and Sze DMY. 2009. The effects of β-glucan on human immune and cancer cells. *J Hematol Oncol*. 2(25).

[7] Wang L, Behr, SR, Newman RK and Newman CW. 1997. Comparative cholesterol-lowering effects of barley β-glucan and barley oil in golden Syrian hamsters. *Nutr Res*. 17(1): 77-88.

[8] Wood PJ. 1994. Evaluation of oat bran as a soluble fibre source. Characterization of oat β-glucan and its effects on glycaemic response. *Carbohydr Polym*. 25(4): 331-336.

[9] Chen J and Seviour R. 2007. Medicinal importance of fungal β-(1→3), (1→6)-glucans. *Mycol Res*. 111(Pt 6): 635-652.

[10] Kudrenko B, Snape N and Barnes AC. 2009. Linear and branched β(1-3) D-glucans activate but do not prime teleost macrophages *in vitro* and are inactivated by dilute acid: Implications for dietary immunostimulation. *Fish Shellfish Immunol*. 26(3): 443-450.

[11] Spolaore P, Joannis-Cassan C, Duran E and Isambert A. 2006. Commercial applications of microalgae. *J Biosci Bioeng*. 101(2): 87-96.

[12] Barsanti L, Vismara R, Passarelli V and Gualtieri P. 2001. Paramylon (β-1,3-glucan) content in wild type and WZSL mutant of *Euglena gracilis*. Effects of growth conditions. *J Appl Phycol*. 13(1): 59-65.

[13] Watanabe F, Yoshimura K and Shigeoka S. 2017. Biochemistry and physiology of vitamins in *Euglena*. In Eds. Schwartzbach SD and Shigeoka S, *Euglena: Biochemistry, Cell and Molecular Biology*. *Adv Exp Med Biol.* Springer International Publishing, Cham. 979: 65-90.

[14] Padayatty SJ, Katz A, Wang Y, Eck P, Kwon O, Lee JH, Chen S, Corpe C, Dutta A, Dutta SK and Levine M. 2003. Vitamin C as an antioxidant: evaluation of its role in disease prevention. *J Am Coll Nutr*. 22(1): 18-35.

[15] Takeyama H, Kanamaru A, Yoshino Y, Kakuta H, Kawamura Y and Matsunaga T. 1997. Production of antioxidant vitamins, beta-carotene, vitamin C, and vitamin E, by two-step culture of *Euglena gracilis* Z. *Biotechnol Bioeng*. 53(2):185-190.

[16] Young VR. 1994. Adult amino acid requirements: the case for a major revision in current recommendations. *J Nutr*. 124(8 Suppl): 1517S-1523S.

[17] Imura K and Okada A. 1998. Amino acid metabolism in pediatric patients. *Nutrition*. 14(1): 143-148.

[18] Šantek B, Felski M, Friehs K, Lotz M and Flaschel E. 2010. Production of paramylon, a β-1,3-glucan, by heterotrophic cultivation of *Euglena gracilis* on potato liquor. *Eng Life Sci*. 10(2): 165-170.

[19] Ogbonna JC, Tomiyama S and Tanaka H. 1998. Heterotrophic cultivation of *Euglena*

gracilis Z for efficient production of α-tocopherol. *J Appl Phycol*. 10(1): 67-74.

[20] Rezić T, Filipović J and Šantek B. 2013. Photo-mixotrophic cultivation of algae *Euglena gracilis* for lipid production. *Agric Conspec Sci*. 78(1): 65-69.

[21] Borowitzka MA and Vonshak A. 2017. Scaling up microalgal cultures to commercial scale. *Eur J Phycol*. 52(4): 407-418.

[22] Fernandez-Valiente E and Rodriguez-Lopez M. 1970-1980. Effect of indole-acetic acid and gibberellic acid on paramylon synthesis in *Euglena gracilis*. *Microbiol Esp*. 32-33: 11–20.

[23] Madhusudhan R, Ishikawa T, Sawa Y, Shigeoka S and Shibata H. 2003. Post-transcriptional regulation of ascorbate peroxidase during light adaptation of *Euglena gracilis*. *Plant Sci*. 165(1): 233-238.

[24] Ishikawa T, Nishikawa H, Gao Y, Sawa Y, Shibata H, Yabuta Y, Maruta T and Shigeoka S. 2008. The pathway via D-galacturonate/L-galactonate is significant for ascorbate biosynthesis in *Euglena gracilis*: Identification and functional characterization of aldonolactonase. *J Biol Chem*. 283(45): 31133-31141.

[25] Ishikawa T, Tajima N, Nishikawa H, Gao Y, Rapolu M, Shibata H, Sawa Y and Shigeoka S. 2010. *Euglena gracilis* ascorbate peroxidase forms an intramolecular dimeric structure: Its unique molecular characterization. *Biochem J*. 426(2): 125-134.

[26] Shigeoka S, Ishiko H, Nakano Y and Mitsunaga T. 1992. Isolation and properties of gamma-tocopherol methyltransferase in *Euglena gracilis*. *Biochim Biophys Acta*. 1128(2-3): 220-6.

[27] Grimm P, Risse JM, Cholewa D, Müller JM, Beshay U, Friehs K and Flaschel E. 2015. Applicability of *Euglena gracilis* for biorefineries demonstrated by the production of α-tocopherol and paramylon followed by anaerobic digestion. *J Biotechnol*. 215: 72-9.

[28] Dwyer MR, Smydzuk J and Smillie RM. 1970. Synthesis and breakdown of β-l,3-glucan in *Euglena gracilis* during growth and carbon depletion. *Aust J Biol Sci*. 23(4): 1005-1014.

[29] Pisoschi AM, Danet AF and Kalinowski S. 2008. Ascorbic acid determination in commercial fruit juice samples by cyclic voltammetry. *J Autom Methods Manag Chem*. 2008: Article ID 937651.

[30] Njoku PC, Ayuk AA and Okoye CV. 2011. Temperature effects on vitamin C content in citrus fruits. *Pakistan J Nutr*. 10(12): 1168-1169.

[31] Rekha C, Poornima G, Manasa M, Abhipsa V, Devi JP, Kumar HTV and Kekuda TRP. 2012. Ascorbic acid, total phenol content and antioxidant activity of fresh juices of four ripe and unripe citrus fruits. *Chem Sci Trans*. 1(2): 303-310.

[32] Foyer C, Rowell J and Walker D. 1983. Measurement of the ascorbate content of spinach leaf protoplasts and chloroplasts during illumination. *Planta*. 157(3): 239-244.

[33] Yuan Y V and Walsh NA. 2006. Antioxidant and antiproliferative activities of extracts from a variety of edible seaweeds. *Food Chem Toxicol*. 44(7): 1144-1150.

[34] Burtin P. 2003. Nutritional value of seaweeds. *Electron J Environ Agric Food Chem*. 2(4): 498-503.

[35] Tani Y and Tsumura H. 1989. Screening for tocopherol-producing microorganisms and α-tocopherol production by *Euglena gracilis* Z. *Agric Biol Chem*. 53(2): 305-312.

[36] Ogbonna JC. 2009. Microbiological production of tocopherols: Current state and prospects. *Appl Microbiol Biotechnol*. 84(2): 217-225.

[37] Grilo EC, Costa PN, Gurgel CSS, de Lima Beserra AF, de Souza Almeida FN and Dimenstein R. 2014. Alpha-tocopherol and gamma-tocopherol concentration in vegetable oils. *Food Sci Technol*. 34(2): 379-385.

[38] Hosotani K and Kitaoka S. 1984. Determination of provitamin A in *Euglena gracilis* Z by high performance liquid chromatography and changes of the contents under various culture conditions. *J Jpn Soc Nutr Food Sci*. 37(6): 519-524.

[39] Mokrosnop VM, Polishchuk AV and Zolotareva EK. 2016. Accumulation of α-tocopherol and β-carotene in *Euglena gracilis* cells under autotrophic and mixotrophic culture

conditions. *Appl Biochem Microbiol*. 52(2): 216-221.

[40] Wu G, Cross HR, Gehring KB, Savell JW, Arnold AN and McNeill SH. 2016. Composition of free and peptide-bound amino acids in beef chuck, loin, and round cuts. *J Anim Sci*. 94(6): 2603-2613.

[41] Ali NM, Yeap SK, Yusof HM, Beh BK, Ho WY, Koh SP, Abdullah MP, Alitheen NB and Long K. 2016. Comparison of free amino acids, antioxidants, soluble phenolic acids, cytotoxicity and immunomodulation of fermented mung bean and soybean. *J Sci Food Agric*. 96(5): 1648-1658.

[42] Miranda MS, Cintra RG, Barros SB and Mancini Filho J. 1998. Antioxidant activity of the microalga *Spirulina maxima*. *Braz J Med Biol Res*. 31(8): 1075-1079.

[43] Patnaik R, Zolandz RR, Green DA and Kraynie DF. 2008. L-tyrosine production by recombinant *Escherichia coli*: fermentation optimization and recovery. *Biotechnol Bioeng*. 99(4): 741-752.

[44] Wada M and Takagi H. 2006. Metabolic pathways and biotechnological production of L-cysteine. *Appl Microbiol Biotechnol*. 73(1): 48-54.

[45] Utagawa T. 2004. Arginine metabolism: Enzymology, nutrition, and clinical significance. *J Nutr*. 134: 2854S-2857S.

[46] Zhang X, Jantama K, Moore JC, Shanmugam KT and Ingram LO. 2007. Production of L-alanine by metabolically engineered *Escherichia coli*. *Appl Microbiol Biotechnol*. 77(2): 355-366.

[47] Liu Q, Zhang J, Wei XX, Ouyang SP, Wu Q and Chen GQ. 2008. Microbial production of L-glutamate and L-glutamine by recombinant *Corynebacterium glutamicum* harboring *Vitreoscilla* hemoglobin gene *vgb*. *Appl Microbiol Biotechnol*. 77(6): 1297-1304.

CHAPTER 7

Conclusive summary and future research directions

7.1 Summary of the work

The work described here is the first attempt to use comprehensive proteomic profiling in *Euglena gracilis* to analyse central carbon metabolic pathways, and the biosynthetic pathways of the metabolites paramylon, ascorbate, α-tocopherol, free amino acids. Focus was placed on the effect of growth conditions (photoautotrophic, PT, mixotrophic, MT, and heterotrophic, HT, conditions) on these pathways. Bioreactor cultivation and production of metabolites in the lesser known sugar-loving strain *E. gracilis* var. *saccharophila*, which has not been reported before for this strain is also described. This study thus contributes to the existing knowledge of metabolic pathways deduced from the transcriptome and purified enzymes of *E. gracilis*, and helps to establish the enzymes that are actually expressed in the proteome from the large pool of isoenzymes found in the transcriptome.

Chapter 3 describes the proteomic response to different growth conditions (PT, MT and HT) on the ascorbate, α-tocopherol and free amino acid pathways in *E. gracilis* var. *saccharophila*. This strain showed a preference for growth under HT cultivation, during which it rapidly accumulated biomass. However, the highest amount of the vitamin antioxidants produced in this strain was during PT cultivation per biomass and during MT cultivation per volume of culture. All 20 of the protein-building amino acids could be detected in this strain, with very high yields of arginine, an essential amino acid for infants and growing children. A 2.5-fold higher cysteine content was detected under the MT cultivation condition, which may have been related to the slower chloroplast development in this strain of *E. gracilis*.

From the proteomic profiling, 3843 non-redundant proteins were identified across all three growth conditions, with significantly less proteins expressed under HT cultivation, compared to PT and MT cultivations. However, the number of expressed proteins was about nine times lower than the number of proteins predicted from the transcriptome [1], and further confirmed that *E. gracilis* actively synthesises mRNA and mainly controls protein expression

at the post-transcriptional level. A large pool of isozymes was reported in the transcriptome for most of the enzyme classes, but only a few of these isozymes were detected in the proteome. For example, only two of the six isozymes of ribose-phosphate diphosphokinase from histidine biosynthesis were expressed in the proteome. Putative functional annotations were also provided for many enzymes that were not reported previously in *E. gracilis*. Only the galacturonate/L-galactonate pathway [2] for ascorbate biosynthesis was dominant in the proteome and was highly up-regulated under PT cultivation, owing to the increased oxidative stress from photosynthesis. The tocopherol cyclase and tocopherol *O*-methyltransferase enzymes from the α-tocopherol biosynthetic pathway, on the other hand, were absent in the proteome and may suggest an alternate pathway in this strain. Most of the biosynthetic pathways matched well with the free amino acid contents, except a few amino acids like glutamate, aspartate and lysine, which have multiple catabolic fates that need to be investigated to truly understand the discrepancies. Amino acid metabolic pathways were similar to a diverse range of eukaryotes such as plants and fungi, with some pathways unique to *Euglena*. Like higher plants, differential expression was detected for serine biosynthesis, with the glycine decarboxylase-serine hydroxymethyltransferase system dominant under PT and MT cultivations, and the conversion of glycerate-3P from glycolysis dominant under HT cultivation. On the other hand, the lysine biosynthetic pathway was similar to higher fungi [3].

Chapter 4 gives an overview of central carbon metabolism in *E. gracilis* Z using label-free shotgun proteomics under PT, MT and HT growth conditions, and compares the results with those obtained for *E. gracilis* var. *saccharophila* in Chapter 3. The total number of non-redundant proteins detected for *E. gracilis* Z (3673) was similar to that of *E. gracilis* var. *saccharophila* under the same growth conditions. The expression patterns of the central carbon metabolic pathways of the two strains also showed similarity. Among the growth conditions, differential expression of proteins was detected between isozymes of several proteins. For instance, an isozyme of pyruvate kinase was expressed under PT condition of both strains only

(may be plastidic), while an isozyme of triosephosphate isomerase was expressed under MT and HT conditions of *E. gracilis* Z only.

A high-specificity glucokinase was detected, which may conduct the first step of the EMP pathway, instead of hexokinase. Use of high-specificity kinases by *E. gracilis* has only been reported in one other study by Lucchini (1971) [4]. Like other microalgae, the EMP pathway appeared to be dominant for glucose metabolism under PT cultivation, while the oxidative pentose phosphate pathway appeared dominant under MT and HT cultivations [5]. An alternative tricarboxylic acid (TCA) cycle has been reported in *E. gracilis* using 2-oxoglutarate decarboxylase enzyme [6]. Another alternative route was detected in this study via a 4-aminobutyrate (GABA) shunt, which is common in bacteroids [7]. Based on the relative abundance of 2-oxoglutarate decarboxylase enzyme and the GABA shunt, the former route was likely to be predominant. Even though the Calvin pathway is meant to be inactive under HT cultivation, enzymes of the pathway such as RuBisCO were still detected suggesting post-translational regulation at the level of protein complex formation. Heterotrophic CO_2 fixation occurred under HT cultivation, even when glucose was present in the medium. The mechanism of heterotrophic CO_2 fixation in *E. gracilis* is reported to be unclear [8-11], and from the proteomic study appeared to be similar to the initial steps of Crassulacean acid metabolism (CAM) and C_4 carbon fixation in higher plants [12].

Chapter 5 further investigates the proteomic profiles of *E. gracilis* Z and *E. gracilis* var. *saccharophila* from Chapters 3 and 4, as well as the proteomic profile of a streptomycin-bleached mutant *E. gracilis* ZSB in order to determine the effect of PT, MT and HT cultivation on the paramylon biosynthesis and degradation pathways. Four candidates of paramylon synthase belonging to the GT48 family have been reported in the transcriptome of *E. gracilis* var. *saccharophila* [1], and two candidates reported in *E. gracilis* Z [13]. Gene knockdown analysis in *E. gracilis* Z revealed that one of the two candidate enzymes (EgGSL2) was indispensable for paramylon synthesis [14]. Our proteomic results also showed that only two

of the four *E. gracilis* var. *saccharophila* candidate enzymes was expressed in the proteome, and confirmed that EgGSL2 had a much higher relative abundance than the other candidate enzyme (EgGSL1), across all cultivation conditions and strains. Furthermore, EgGSL1 was not detected in any of the three strains under HT cultivation, while it was detected under PT and MT cultivations of all strains respectively. However, both *EgGSL1* and *EgGSL2* transcripts were detected by qRT-PCR. The results suggested that the expression of EgGSL1 may be light-induced, and regulated post-transcriptionally. Presence of glucosidases during paramylon synthesis at mid-log phase indicated that the activity of the synthase enzymes was not affected by glucosidases. This is similar to a study by Marechal and Goldemberg (1964), where snail gut juice, containing carbohydrases, had no effect on paramylon synthesis [15].

Chapter 6 describes the ascorbate, α-tocopherol and free amino acid contents of *E. gracilis* Z and ZSB in shake-flask cultures, which were not described in the previous chapters. This chapter also describes the cultivation of *E. gracilis* var. *saccharophila* under MT cultivation, in a laboratory-scale bioreactor, for production of optimum amounts of the dietary supplements paramylon, vitamin antioxidants, proteins and free amino acids. The strain and cultivation condition were chosen for bioreactor cultivation based on the ability of *E. gracilis* var. *saccharophila* to produce more paramylon than *E. gracilis* Z (Chapter 5) and more antioxidants than *E. gracilis* ZSB under MT cultivation. Bioreactor cultivation of *E. gracilis* var. *saccharophila* resulted in a higher biomass accumulation comparable to HT cultivations at the shake-flask level, with high productivities of paramylon and the vitamin antioxidants ascorbate, α-tocopherol and β-carotene. Although a high amount of free amino acids was produced, the protein content was quite low during high paramylon productivity. The concentration of some of the metabolites achieved during the bioreactor cultivation of *E. gracilis* var. *saccharophila* was not as high as most of the commercially available sources of these metabolites. However, *E. gracilis* var. *saccharophila* is still an attractive candidate for

metabolite production as it simultaneously synthesises all of these metabolites, as well as other metabolites of interest.

In summary, the use of label-free shotgun proteomics in this work unveiled several metabolic differences between growth conditions and proves the high throughput ability of this method to identify a very large number of differentially expressed proteins concurrently. When compared to the sheer number of proteins predicted in the transcriptome, the proteomic profiling led to the detection of only a fraction of these proteins. Since *E. gracilis* is known for not controlling the expression of most of its proteins at the transcriptional level, proteomic profiling proved to be a powerful tool that enabled identification of only those enzymes that were translated at the proteomic level. Moreover, by using quantitative analysis of the proteomic profiles under different growth conditions, differentially expressed proteins could be identified in this study. The work presented here can thus serve as a first step into the high throughput analysis of *E. gracilis* proteomics to truly understand the metabolic capability of this organism and to reengineer its strains.

7.2 Future directions

The current work is the first comprehensive proteomic-based study of *Euglena gracilis* to describe the biosynthetic and degradation pathways of paramylon, and the biosynthetic pathways of ascorbate, α-tocopherol and free amino acids. This study thus represents a start for further research to detect and understand novel enzymes and pathways in *E. gracilis*, and genetically improve it as a source of nutraceuticals. The work described here has provided preliminary insight into proteins involved in metabolite production in *E. gracilis*. However, further investigations are required to fully understand the functions of these proteins and identify other proteins to optimise *E. gracilis* as an industrial production host.

Production of the metabolites reported for *E. gracilis* var. *saccharophila* in Chapter 6

can further be enhanced through medium composition optimisation and strain improvement, as little is known about this strain and its capabilities. The effect of other supplements that can be added to medium for this enhancement, such as glutamate and ethanol also needs to be investigated in the future as these supplements have been shown to increase metabolite production in other strains [16-17]. Metabolomic profiling of *E. gracilis* var. *saccharophila* can be used to truly assess its ability as a producer of supplements.

Mapping of the proteins against the genome is necessary to decipher the genes coding for the proteins of interest. The complete annotated genomic sequence of *E. gracilis* has not been published yet, although the description of strategies for a draft genome assembly with some initial features of the genome has just been made available [18]. Identification of the corresponding genes encoding proteins of interest is valuable for creating gene knockouts and metabolic engineering of strains, as well as overexpressing proteins of interest. The information from this mapping can also give access to gene promoters that can be used for inducing expression of genes of interest. For instance, in the current study, more antioxidants were produced during the stationary phase of growth, while more paramylon was produced during the exponential phase of growth. Genetically enhanced strains with condition-specific promoters may be built to produce both antioxidants and paramylon in large quantities simultaneously. Currently, genetic manipulation of *E. gracilis* is limited by the lack of molecular tools, which are under development. For example, chloroplast transformation has been described for *E. gracilis* [19, 20], whereas nuclear transformation would be required for enhancing the production of metabolites and engineering entire pathways. In the future, genetic engineering of *E. gracilis* may lead to an increase in biomass production and increased yield of metabolites of interest. For instance, genetic engineering of the rate-limiting reaction in the α-tocopherol biosynthesis pathway in the cyanobacterium *Synechocystis* sp. led to a five-fold increase in α-tocopherol content [21]. Higher yields of α-tocopherol may also be achieved in

E. gracilis in the future by reengineering the rate-limiting reaction for α-tocopherol biosynthesis.

Purification and characterisation of proteins from the current study is crucial to further understanding their roles and optimum functional conditions. These can then be used to choose the ideal isozyme for a given growth condition, or a combination of isozymes that can enhance the production of metabolites. For instance, the EMP pathway enzyme hexokinase was missing from the proteomic study indicating the use of a high-specificity glucokinase instead. Other high-specificity kinases may be present in *E. gracilis* to phosphorylate other organic carbon sources, which need to be identified to explore the mechanism behind the ability of *E. gracilis* to use such a wide variety of carbon sources [17]. Another example is the paramylon synthase candidate EgGSL1, which was not expressed under dark conditions in this study, although *EgGSL1* transcripts were present. This enzyme can be further investigated by determining the amount of *EgGSL1* mRNA associated with polysomes under dark and light cultivations. Light exposure is known to increase the number of polysomes in *E. gracilis*, probably by altering the efficiency with which individual mRNAs are translated [22]. This method has previously been used to deduce regulation of translation of the pLHCPII protein by photoinduction [22], and can be used for the EgGSL1 protein as well. Moreover, the function of EgGSL1 is unclear so far, and the protein needs to be isolated, purified and functionally characterised.

RNA silencing studies, such as the one carried out by Tanaka *et al.* (2017) on paramylon synthase from *E. gracilis* Z [14], can help to functionally characterise proteins, as well as predict and validate predominant isozymes expressed under different conditions. Knockdowns of genes corresponding to the proteins identified from the current study will help to elucidate the importance of individual isozymes in the production of the given metabolite. For example, nine isozymes of PEPCK involved in heterotrophic CO_2 fixation were detected in this study. Several of these had similar spectral counts, and comparative studies of knockdowns can help to identify the dominant PEPCK isozyme for heterotrophic CO_2 fixation.

Many other enzymes from the current study related to metabolite production have also not been characterised and the conditions for their optimal activity are unknown. Having the knowledge of optimal working conditions of an enzyme can allow development of *in vitro* synthetic pathways for metabolite production using a combination of enzymes from *E. gracilis* and other organisms that are best suited to synthesise the metabolite. These can be chosen based on their ability to operate in similar optimal conditions, and then selecting the correct substrate. For example, Schwander *et al.* (2016) developed an *in vitro* optimised synthetic pathway for CO_2 fixation by using seventeen enzymes that operated under similar conditions of interest from nine different organisms [23].

Further proteomic studies will need to be carried out on *E. gracilis* to broaden the protein range identified. The predicted number of non-redundant proteins from the transcriptome of *E. gracilis* was 32,128 [1], of which only a fraction was detected in the proteome under the conditions studied here. By exposing the cells to different conditions such as N-starvation, temperature stress, high light stress, different carbon sources, heavy metal contamination etc, more of the stress-related proteins may be expressed in the proteome. This can lead to the identification of novel enzymes to combat abiotic stresses, or detection of enzymes that have only previously been described in other organisms. The metabolic capability of *E. gracilis* remains largely untapped and the vast majority of the predicted proteins unknown. About 30% of the predicted proteins cannot be matched to any protein sequences currently available [1], and this indicates the immense opportunity of detecting and characterising novel proteins from this organism through further proteomic studies. Moreover, labelled-shotgun proteomics approaches such as TMT and iTRAQ labelling, which are more high-throughput and reduce variability among samples being compared, can be used to carry out such proteomic studies. These proteomic studies can be combined with transcriptomic, metabolomic, biochemical and physiological studies in order to produce a comprehensive network for metabolite production in *E. gracilis*.

References

[1] O'Neill EC, Trick M, Hill L, Rejzek M, Dusi RG, Hamilton CJ, Zimba PV, Henrissat B and Field RA. 2015. The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Mol BioSyst*. 11(10): 2808-2820.

[2] Wheeler G, Ishikawa T, Pornsaksit V and Smirnoff N. 2015. Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *Elife*. 4: e06369.

[3] Zabriskie TM and Jackson MD. 2000. Lysine biosynthesis and metabolism in fungi. *Nat Prod Rep*. 17(1): 85-97.

[4] Lucchini G. 1971. Control of glucose phosphorylation in *Euglena gracilis*. I. Partial characterization of a glucokinase. *Biochim Biophys Acta*. 242(2): 365-370.

[5] Perez-Garcia O, Escalante FME, de-Bashan LE and Bashan Y. 2011. Heterotrophic cultures of microalgae: Metabolism and potential products. *Water Res*. 45(1): 11-36.

[6] Shigeoka S, Onishi T, Maeda K, Nakano Y and Kitaoka S. 1986. Occurrence of thiamin pyrophosphate-dependent 2-oxoglutarate decarboxylase in mitochondria of *Euglena gracilis*. *FEBS Lett*. 195(1-2): 43-47.

[7] Green LS, Li Y, Emerich DW, Bergersen FJ and Day DA. 2000. Catabolism of α-ketoglutarate by a *sucA* mutant of *Bradyrhizobium japonicum*: evidence for an alternative tricarboxylic acid cycle. *J Bacteriol*. 182(10): 2838-2844.

[8] Peak JG and Peak MJ. 1980. Heterotrophic carbon dioxide fixation products of *Euglena*: effects of ammonium. *Plant Physiol*. 65(3): 566-568.

[9] Wolfovitch R and Perl M. 1972. A possible ribosomal-directed regulatory system in *Euglena gracilis*. Carbon dioxide fixation. *Biochem J*. 130(3): 819-823.

[10] Levedahl BH. 1966. Heterotrophic CO_2 fixation by a bleached *Euglena*. *Exp. Cell Res*.

44(2-3): 393-402.

[11] Peak JG and Peak MJ. 1976. Heterotrophic (dark) CO_2 fixation by *Euglena gracilis*. Possible regulation by tricarboxylic acid cycle intermediates. *J Protozool*. 23(1): 165-167.

[12] Bakrim N, Brulfert J, Vidal J and Chollet R. 2001. Phosphoenolpyruvate carboxylase kinase is controlled by a similar signaling cascade in CAM and C_4 plants. *Biochem Biophys Res Commun*. 286(5): 1158-1162.

[13] Yoshida Y, Tomiyama T, Maruta T, Tomita M, Ishikawa T and Arakawa K. 2016. De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics*. 17: 182.

[14] Tanaka Y, Ogawa T, Maruta T, Yoshida Y, Arakawa K and Ishikawa T. 2017. Glucan synthase-like 2 is indispensable for paramylon synthesis in *Euglena gracilis*. *FEBS Lett*. 591(10): 1360-1370.

[15] Marechal LR and Goldemberg SH. 1964. Uridine diphosphate glucose-β-1,3-glucan β-3-glucosyltransferase from *Euglena gracilis*. *J Biol Chem*. 239(10): 3163-3167.

[16] Ogbonna JC, Tomiyama S and Tanaka H. 1998. Heterotrophic cultivation of *Euglena gracilis Z* for efficient production of α-tocopherol. *J Appl Phycol*. 10(1): 67-74.

[17] Rodríguez-Zavala JS, Ortiz-Cruz MA, Mendoza-Hernández G and Moreno-Sánchez R. 2010. Increased synthesis of α-tocopherol, paramylon and tyrosine by *Euglena gracilis* under conditions of high biomass production. *J Appl Microbiol*. 109(6): 2160-2172.

[18] Ebenezer TE, Carrington M, Lebert M, Kelly S and Field MC. 2017. *Euglena gracilis* genome and transcriptome: Organelles, nuclear genome assembly strategies and initial features. In Eds. Schwartzbach SD and Shigeoka S, *Euglena: Biochemistry, Cell and Molecular Biology*. *Adv Exp Med Biol*. Springer International Publishing, Cham. 979: 125-140.

[19] Ogawa T, Tamoi M, Kimura A, Mine A, Sakuyama H, Yoshida E, Maruta T, Suzuki K, Ishikawa T and Shigeoka S. 2015. Enhancement of photosynthetic capacity in Euglena gracilis by expression of cyanobacterial fructose-1,6-/sedoheptulose-1,7-bisphosphatase leads

to increases in biomass and wax ester production. Biotechnol Biofuels. 8: 80.

[20] Doetsch NA, Favreau MR, Kuscuoglu N, Thompson MD and Hallick RB. 2001. Chloroplast transformation in Euglena gracilis: splicing of a group III twintron transcribed from a transgenic psbK operon. Curr Genet. 39(1): 49-60.

[21] Qi Q, Hao M, Ng W, Slater SC, Baszis SR, Weiss JD and Valentin HE. 2005. Application of the *Synechococcus nirA* promoter to establish an inducible expression system for Eegineering the *Synechocystis* tocopherol pathway. *Appl Environ Microbiol*. 71(10): 5678-5684.

[22] Schwartzbach SD. 2017. Photo and Nutritional Regulation of *Euglena* Organelle Development. In Eds. Schwartzbach SD and Shigeoka S, *Euglena: Biochemistry, Cell and Molecular Biology*. *Adv Exp Med Biol*. Springer International Publishing, Cham. 979: 159-182.

[23] Schwander T, von Borzyskowski LS, Burgener S, Cortina NS and Erb TJ. 2016. A synthetic pathway for the fixation of carbon dioxide in vitro. *Science*. 354(6314): 900-904.

Supporting information

Table 1: Supplementary data files associated with this research.

File Name	File Type	Description	Source Chapter
Chapter 3_Supplementary Table S1	.xlsx	Protein identifications of *E. gracilis* var. *saccharophila* from nanoLC-MS/MS spectra, using the global proteome machine (GPM), with log(e) values less than – 1.	3
Chapter 3_Supplementary Table S2	.xlsx	Analysis of variance (ANOVA) and two-sample unpaired t-tests of *E. gracilis* var. *saccharophila*, showing differential expression of proteins, based on normalised spectral abundance factor (NSAF) values, using photoautotrophic (PT) cultivation as the control.	3
Chapter 3_Supplementary Table S3	.xlsx	Fold changes of *E. gracilis* var. *saccharophila* proteins from the ascorbate, α-tocopherol and free amino acid biosynthesis pathways.	3
Chapter 4_Supplementary Table S1	.xlsx	Protein identifications of *E. gracilis* Z from nanoLC-MS/MS spectra, using the GPM, with log(e) values less than – 1.	4
Chapter 4_Supplementary Table S2	.xlsx	ANOVA and two-sample unpaired t-tests of *E. gracilis* Z, showing differential expression of proteins, based on NSAF values, using PT cultivation as the control.	4
Chapter 4_Supplementary Table S3	.xlsx	Fold changes of *E. gracilis* Z proteins from the central carbon and paramylon metabolism pathways.	4

Chapter 5_Supplementary Figure S5.1	.pdf	Comparison of unused glucose in the medium for *E. gracilis* var. *saccharophila*, *E. gracilis* Z and *E. gracilis* ZSB strains under mixotrophic (MT) and heterotrophic (HT) cultivations.	5
Chapter 5_Supplementary Table S5.1	.xlsx	Protein identifications of *E. gracilis* ZSB from nanoLC-MS/MS spectra, using the GPM, with log(e) values less than – 1.	5
Chapter 5_Supplementary Table S5.2	.xlsx	Two-sample unpaired t-tests of *E. gracilis* ZSB, showing differential expression of proteins, based on NSAF values. MT cultivations were compared against MT cultivations of *E. gracilis* Z, while HT cultivations were compared against HT cultivations of *E. gracilis* Z.	5
Chapter 5_Supplementary Table S5.3	.xlsx	Protein identifications of three extra HT replicates of each of *E. gracilis* Z and *E. gracilis* var. *saccharophila* from nanoLC-MS/MS spectra, using the GPM, with log(e) values less than – 1, to validate the EgGSL1 enzyme (ID: light_m.13249) to be missing during HT cultivation.	5
Chapter 6_Supplementary Figure S6.1	.pdf	Microscopic images of cells from shake-flask cultures of *E. gracilis* var. *saccharophila*, *E. gracilis* Z and *E. gracilis* ZSB under MT cultivation, showing paramylon granules and chloroplasts at exponential phase and stationary phase of growth.	6

Appendix

Biosafety approval

MACQUARIE UNIVERSITY

Office of the Deputy Vice-Chancellor (Research)
Research Office

C5C Research HUB East, Level 3, Room 324

MACQUARIE UNIVERSITY NSW 2109 AUSTRALIA

Phone +61 (0)2 9850 4063
Fax +61 (0)2 9850 4465

Email biosafety@mq.edu.au

13th May 2016

Professor Helena Nevalainen
Faculty of Science and Engineering
Building E8C
Macquarie University

Dear Helena,

Re: "Genetic toolbox for metabolic engineering of fresh water protist *Euglena gracilis*" [5201600395]

NOTIFICATION OF A NOTIFIABLE LOW RISK DEALING (NLRD)

The above application has been reviewed by the Chair of the Institutional Biosafety Committee (IBC) and has been approved as an NLRD, effective 13th May, 2016

This approval is subject to the following standard conditions:

- The NLRD is conducted by persons with appropriate training and experience, within a facility certified to either Physical Containment level 1 (PC1) or PC2.
- Work requiring Quarantine Containment level 2 (QC2) does not commence until the facility has been certified
- Only persons who have been *assessed by the IBC* as having appropriate training and experience may conduct the dealing. This includes persons involved in all parts of the dealing e.g. researchers, couriers and waste contractors. A copy of the IBC's record of assessment must be retained by the project supervisor.
- NLRDs classified under Part 1 of Schedule 3 of the Regulations must be conducted in a facility (or class of facilities) that is certified to at least a PC1 level **and** that is mentioned in the IBC record of assessment as being appropriate for the dealing;
- NLRDs classified under Part 2.1 of Schedule 3 of the Regulations must be conducted in a facility (or class of facilities) that is certified to at least a PC2 level **and** that is mentioned in the IBC record of assessment as being appropriate for the dealing;
- Any transport of the GMO must be conducted in accordance with the Regulator's *Guidelines for the Transport of GMOs* available at: http://www.ogtr.gov.au/internet/ogtr/publishing.nsf/Content/transport-guide-1
- A copy of the IBC's record of assessment has been attached to this approval.
- The record of assessment must be kept by the person or organisation for 8 years after the date of assessment by the IBC (regulation 13 C of the *Gene Technology Regulations 2001*).
- All NLRDs undertaken by Macquarie University will be reported to the OGTR at the end of every financial year.
- If the dealing involves organisms that may produce disease in humans, the NLRDs must be conducted in accordance with the vaccination requirements set out in the Australian Standard AS/NZS 2243.3:2010

1

- The Chief Investigator must inform the Biosafety Committee if the work with GMOs is completed or abandoned.

Please note the following standard requirements of approval:
1. Approval will be for a period of 5 years subject to the provision of annual reports. If, at the end of this period the project has been completed, abandoned, discontinued or not commenced for any reason, you are required to submit a Final Report. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. Please contact the Committee Secretary at biosafety@mq.edu.au for a copy of the annual report.

A Progress/Final Report for this study will be due on: 13th May 2017

2. If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have final approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of approval to an external organisation as evidence that you have approval, please do not hesitate to contact the Committee Secretary at the address below.

Please retain a copy of this email as this is your formal notification of final Biosafety approval.

Yours Sincerely

V

Associate Professor Subramanyam Vemulpad
Chair, Macquarie University Institutional Biosafety Committee

Encl. Copy of record submitted by Macquarie University to the OGTR.

Biosafety Secretariat
Research Office
Level 3, Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia

T: +61 2 9850 4063
F: +61 2 9850 4465

http://www.mq.edu.au/research

MACQUARIE UNIVERSITY
SYDNEY ~ AUSTRALIA
CRICOS Provider Number 00002J

Record of Proposed Notifiable Low Risk Dealing(s) (NLRDs) - Version 9

Regulation 13C(1)(a) - *Gene Technology Regulations 2001* (effective 1 Sept. 2011)

INSTRUCTIONS: Under r.13C(2)(a) of the *Gene Technology Regulations 2001* (effective 1 Sept. 2011), an accredited organisation must provide a copy of this form to the Regulator in the annual report for the financial year to be given by the organisation to the Regulator.

Under r.13C(2)(b) any other person (including non-accredited organisations) must provide a copy of this form to the Regulator in a report for the financial year to be given by the person to the Regulator.

NOTE: This form also captures information required by the Gene Technology Regulator to meet the requirements of r. 39 for NLRDs in maintaining a record of GMO and GM product dealings.

IBC NLRD Identifier *(number)*	5201600395
Date of IBC Assessment	13 May 2016
Name of IBC	Macquarie University Institutional Biosafety Committee
Name of Organisation Notifying Dealing *	Macquarie University
Project Title * (please do not include confidential information)	Genetic toolbox for metabolic engineering of fresh water protist Euglena gracilis
GMO Details	*Euglena gracilis*
PC1 NLRD TYPE *(a) – (c) **	
PC2 NLRD TYPE *(a) – (m) **	c
PC3 NLRD *	
Has the IBC assessed the proposed dealing to be an NLRD?	Yes

** Fields that will be included in the public list of NLRDs on the OGTR website*