
Genome Scale Metabolic Modelling of
Phaeodactylum tricornutum

Dipali Singh

A thesis submitted in partial fulfilment of the requirements of
Oxford Brookes University
for the award of the degree Doctor of Philosophy

This project is part of
Marie Curie Initial Training Network,
AccliPhot
funded by the European Commission

July 2017

Abstract

Diatoms are photoautotrophic unicellular algae and are among the most abundant, adaptable and diverse marine phytoplankton. Their ability to synthesise lipid as a storage compound (20%-50% dry cell weight) makes them a potential sources of biofuel and high-value commodities such as ω fatty acids. However, diatoms have unique features in their biochemistry as compared to higher plants and hence, there is a prior need to understand diatom metabolism to enable physiological and genetic manipulation, and improve their strains. The present work involves construction and analysis of genome scale metabolic models (GSMs) of *Phaeodactylum tricornutum*, a model diatom, the characterisation of physiological properties and the identification of the potential strategies to optimise the lipid production.

GSMs were constructed based on a previously published model and metabolic databases, and were analysed using structural modelling techniques to understand the metabolic responses at different environmental and physiological conditions. The model results suggest change in metabolic responses, mainly associated with the Calvin cycle, reductant transfer, photorespiration, TCA cycle, glyoxylate cycle, lipid metabolism, carbohydrate metabolism and energy dissipation mechanisms under changing environmental and physiological conditions. Carbon fixation and triose-phosphate production can take place solely in the chloroplast, despite of differences in the localisation and regulation of the Calvin cycle enzymes as compared to higher plants. Further, model analysis suggests that lipid production in *P. tricornutum* increases both when exposed to high light and with the availability of glycerol. The potential metabolic routes for lipid production involves phosphoketolase pathway, threonine metabolism, recycling of glycolate and HCO_3^- fixation.

Based on the model analysis, experiments were designed where cultures were exposed to high light, supplemented with HCO_3^- , under phototrophic and mixotrophic conditions. This resulted to an increase in biomass and lipid productivity. In addition, by revealing the potential metabolic routes involved with lipid production, our work also suggests possible targets for metabolic engineering that could divert carbon towards lipid production.

Acknowledgements

First and foremost I would like to express my sincere gratitude to my supervisors Dr. Mark Poolman and Prof. David Fell for their continuous motivation and support throughout my Phd journey. I am enormously grateful to them for sharing their immense knowledge and encouraging me to grow as a research scientist. Their advice have been tremendously educational for my professional and personal development.

I am thankful to all the past and present members of the Cell Systems Modelling Group: Dr. Hassan Hartman, Kailash Adhikari, Noah Mesfin and Ela Al-Saidi for providing wonderful working atmosphere with healthy and fun filled discussions, and for being excellent teammates and travel-mates. My special thanks to Dr. Hassan Hartman for providing crucial practical support during the early days of my Phd.

I am also very grateful to the Accliphot consortium, Marie-Curie Initial Training Network funded by European Commission, for the financial support and for providing an excellent platform to develop interdisciplinary skills. In particular, I would like to thank Dr. Valeria Villanova and Dr. Adeline Le Monnier for helping me to setup experiments during my secondment and providing additional experimental data. I am also very grateful for the discussion and support from Dr. Giovanni Finazzi and Prof. Oliver Ebenhöh. I would also like to sincerely thank Prof. Ross Carlson, Montana State University, for providing GSM of *P. tricornutum* which served as the basis for my early work.

I thank the Oxford Brookes University and the staff of the School of Life Sciences for the administrative support. In particular, I would like to acknowledge Dr. Verena Kriechbaumer, Jill Organ, Philip Voysey, Prof. Chris Hawes, Prof. Susan Brookes and Prof. David Evans.

I would also like to thank all my family members who incited me to strive towards my goal. I am grateful beyond words for the patience and support from my parents, Neelam Singh and Uday Prasad Singh, and for allowing me to realise my own potential.

Contents

Abstract	i
Acknowledgements	ii
Contents	iii
Nomenclature	vii
URLs	vii
Abbreviations	viii
1 Introduction	1
1.1 Project Background	1
1.2 Motivation	1
1.3 Aims and Structure	3
1.4 Diatoms and Biochemistry	4
1.4.1 Origin of Diatoms	6
1.4.2 Central Carbon Metabolism of Plants and Algae	7
1.4.3 Central Carbon Metabolism of Diatoms	15
1.4.4 <i>Phaeodactylum tricornutum</i>	21
1.5 Conclusion	23
2 Mathematical Modelling of Metabolism	24
2.1 Introduction	24
2.2 Structural Metabolic Modelling	25
2.2.1 The Steady State Assumption	27
2.2.2 Null-Space Analysis	27

2.2.3	Elementary Modes Analysis	31
2.2.4	Flux Balance Analysis	33
2.3	Construction of Genome Scale Metabolic Models	35
2.3.1	Model Construction and Curation	37
2.4	Metabolic Modelling Tools	39
2.4.1	The Python Programming Language	39
2.4.2	ScrumPy: Metabolic Modelling with Python	40
3	Construction of <i>P. tricornutum</i> genome scale models	45
3.1	Introduction	45
3.2	Construction and Analysis of GSMs	46
3.2.1	Model Structure	46
3.2.2	Software and Tools	48
3.2.3	Model Construction and Curation	48
3.2.4	Estimation of Experimental Parameters	53
3.2.5	Model Analysis	54
3.3	Results	55
3.3.1	General Model Properties	55
3.3.2	Metabolic Responses	55
3.4	Discussion	61
3.4.1	Model Construction and Curation	61
3.4.2	Comparison of GSMs	62
3.4.3	General Model Properties	64
3.4.4	Estimation of Maintenance Cost	65
3.4.5	Minimum Photon Flux Requirement	65
3.4.6	Metabolic Responses	66
3.5	Conclusion	69
4	Reaction Response to Variation Analysis	71
4.1	Introduction	71
4.2	Method	72
4.2.1	Photon Flux Variation Analysis	73
4.2.2	Lipid Demand Variation Analysis	73

4.2.3	Glycerol Uptake Variation Analysis	74
4.3	Results	75
4.3.1	Reaction Responses to Photon Flux Variation	75
4.3.2	Reaction Responses to Lipid Demand Variation	83
4.3.3	Reaction Response to Glycerol Uptake Variation	86
4.4	Discussion	91
4.4.1	Reductant Transfer	91
4.4.2	Nitrogen Assimilation	93
4.4.3	Energy Dissipation Mechanisms	93
4.4.4	Photorespiration	95
4.4.5	Lipid Production	97
4.4.6	Effect of Glycerol on Metabolism	99
4.5	Conclusion	100
5	Experimental Setup and Results	101
5.1	Introduction	101
5.2	Method	102
5.2.1	Experimental Setup	102
5.2.2	Nitrogen and Phosphate Concentration	102
5.2.3	Measurement of Growth	105
5.2.4	Nile Red Staining and Microscopy	105
5.2.5	Biomass Composition Measurement	105
5.3	Results	106
5.3.1	Effect of bicarbonate on growth rate	106
5.3.2	Effect of bicarbonate as a function of absorbed light	107
5.3.3	Effect of light intensity and bicarbonate availability	108
5.4	Discussion	111
5.5	Conclusion	114
6	General Discussion	115
6.1	Overview of Results	115
6.2	Limitations and Future Works	118
6.2.1	Alternative Objective Functions	118

6.2.2	Integration of Expression Data	119
6.2.3	Integration of Metabolomics Data	121
6.3	Conclusions	121
Appendix A Models and Additional materials		150
A.1	Model Construction	150
A.1.1	Deleted reactions and metabolites	150
A.1.2	Corrected reactions and metabolites	151
A.2	Python Modules for Model Analysis	151
A.2.1	ModelProperty	151
A.2.2	LightScan	151
A.2.3	LipidScan	151
A.2.4	Mixoscan	151
Appendix B Experimental Data		152
B.1	Electron Transfer Rate	152
B.2	Estimation of Growth Rate	152
B.3	Estimation of Light Absorbed and Glycerol Consumed	153
Appendix C Biomass Composition		155
Appendix D Published Papers		157

URLs

BioCyc	http://biocyc.org/
DiatomCyc	http://diatomcyc.org/
BRENDA	http://www.brenda-enzymes.org/
GLPK	http://www.gnu.org/software/glpk/
Gnuplot	http://www.gnuplot.info/
NumPy	http://www.numpy.org/
Python	http://docs.python.org/
SciPy	http://www.scipy.org/
ScrumPy	http://mudshark.brookes.ac.uk/ScrumPy

Abbreviations

Table 1: Summary of abbreviations of reactions and metabolites.

<i>Metabolites</i>	<i>Common Name</i>
Acet:	acetate
AcAld:	acetate-aldehyde
Ac-P:	acetyl-phosphate
AcCoA:	acetyl-CoA
ATP:	adenosine-triphosphate
ADP:	adenosine-diphosphate
ALA:	alanine
Ax:	antheraxanthin
ASP:	aspartate
BPGA:	1,3-diphosphateglycerate
Carbamoyl-P:	carbamoyl phosphate
Cit:	citrate
CYS:	cysteine
Cyt red:	Cytochromes-C-Reduced
Cyt ox:	Cytochromes-C-Oxidized
DAGs:	diacylglycerols
Ddx:	diadinoxanthin
Dtx:	diatoxanthin
DHAP:	dihydroxyacetone phosphate
EPA:	eicosapentaenoic acid
E4P:	erythrose-4-phosphate
FA:	fatty acid
FAAE:	fatty acid alkyl esters
FBP:	fructose bisphosphate
F6P:	fructose-6-phosphate
Fum:	fumarate

<i>Metabolites</i>	<i>Common Name</i>
G6P:	glucose-6-phosphate
G1P:	glucose-1-phosphate
GLT:	glutamate
GLN:	glutamine
GAP:	glyceraldehyde 3-phosphate
GLY:	glycine
Glyox:	glyoxylate
IsoCit:	isocitrate
2K3DPG:	2-dehydro-3-deoxy-D-gluconate-6-phosphate
2KG:	2-ketoglutarate
Mal:	malate
MAGs:	monoacylglycerols
NAD ⁺ :	nicotinamide adenine dinucleotide
NADP ⁺ :	nicotinamide adenine dinucleotide phosphate
OAA:	oxaloacetic acid
OH-Pyr:	3-hydroxypyruvate
Pi:	phosphate
2PG:	2-phospho-D-glycerate
PGA:	3-phospho-D-glycerate
PEP:	phosphoenol pyruvate
6PG:	6-phosphogluconate
PUFAs:	polyunsaturated fatty acids
Pyr:	pyruvate
Q:	ubiquinone
QH2:	ubiquinol
R5P:	ribose-5-phosphate
RuBP:	ribulose-1,5-bisphosphate
SBP:	sedoheptulose-1,7-bisphosphate
S7P:	sedoheptulose-7-phosphate
SER:	serine
Suc:	succinate

<i>Metabolites</i>	<i>Common Name</i>
SucCoA:	succinyl-S-CoA
SucAld:	succinic semialdehyde
THR:	threonine
TAGs:	triacylglycerols
TP:	triose phosphate
UDP:	uridine-diphosphate
UTP:	uridine-triphosphate
UDPG:	UDP-glucose
Vx:	violaxanthin
X5P:	xylulose-5-phosphate
Zx:	zeaxanthin
<i>Enzymatic Reactions</i>	<i>Common Name</i>
AK:	adenylate kinase
FBPaldo:	FBP aldolase
FBPase:	fructose 1,6-bisphosphatase
FTL:	formate-tetrahydrofolate ligase
GAPdh:	GAP dehydrogenase
G6Pdh:	G6P dehydrogenase
GLTdh:	glutamate dehydrogenase
GS:	GLN synthetase
GOGAT:	GLN oxoglutarate aminotransferase
Glycerol-3P:	glycerol-3-phosphate
G6Piso:	G6P isomerase
GDH:	glycolate dehydrogenase
GK:	glycerate kinase
GOX:	glycolate oxidase
KDPGaldo:	2K3DPG aldolase
LDH:	lactate dehydrogenase
Maldh:	Mal dehydrogenase
PEPC:	PEP carboxylase
PEPck:	PEP carboxykinase

<i>Enzymatic Reactions</i>	<i>Common Name</i>
PFK:	phosphofructokinase
PGM:	phosphoglucomutase
6PGdh:	6PG dehydrogenase
PGD:	phosphoglycerate dehydratase
PGK:	phosphoglycerate kinase
PK:	pyruvate kinase
PRK:	phosphoribulokinase
PPDK:	pyruvate-orthophosphate dikinase
Pyrdh:	pyruvate dehydrogenase
rubisco:	RuBP carboxylase/oxygenase
Ru5PK:	ribulose-5-phosphokinase
SBPase:	sedoheptulose-1,7-bisphosphatase
SBPaldo:	SBP aldolase
TAL:	transaldolase
TK:	transketolase
TPI:	triosephosphate isomerase
<i>Network Diagrams</i>	<i>Common Name</i>
r1:	rubisco carboxylase
r2, r41, r50:	PGK
r3, r39, r40 r49:	GAPdh
r4, r38:	TPI
r5, r37:	FBPaldo
r6, r36:	FBPase
r7, r35:	PFK
r8, r20:	G6Piso
r9, r11, r29, r30:	TK
r10, r31:	TAL
r12, r34:	ribulose-phosphate 3-epimerase
r13:	R5P isomerase
r14:	Ru5PK
r15, r51:	PGM

<i>Network Diagrams</i>	<i>Common Name</i>
r16, r52:	PGD
r17, r42, r53:	PK
r18:	PPDK
r19, r64, r100:	Pyr carboxylase
r21:	phosphoglucomutase
r22:	UTPG uridylyltransferase
r23:	Chrysolaminaran Synthesis
r27:	SBPaldo
r28:	SBPase
r32:	G6Pdh
r33:	6PGdh
r43, r66:	PEPC
r45:	phosphoketolase
r46:	phosphate acetyltransferase
r47:	phosphogluconate dehydratase
r48:	KDPGaldo
r54:	Pyrdh
r55:	Cit synthase
r56:	aconitate hydratase
r57:	IsoCit dehydrogenase
r58:	2KG dehydrogenase
r59:	SucCoA ligase
r60:	Suc dehydrogenase
r61	Fum hydratase
r62	Maldh
r63:	ME
r69, r91:	GLTdh
r70:	rubisco oxygenase
r71:	glycolate dehydrogenase
r72:	GLY aminotransferase
r73:	glycine decarboxylase complex

<i>Network Diagrams</i>	<i>Common Name</i>
r74:	serine hydroxymethyltransferase
r78:	ASP transaminase
r79:	THR biosynthesis
r80:	THR aldolase
r81:	aldehyde dehydrogenase
r82	GOX
r83	catalase
r84	Mal synthase
r85:	2KG decarboxylase
r86	SucAld dehydrogenase
r87:	serine-glyoxylate aminotransferase
r88:	glycerate dehydrogenase
r89:	GK
r90:	ALA aminotransferase
r92:	LDH
r93:	glycerol kinase
r94:	glycerol-3P dehydrogenase
r95:	GDH
r96:	SER acetyltransferase
r97:	CYS synthase
r98:	CYS desulfidase
r99:	SER synthesis from PGA
r101:	Ornithine carbamoyltransferase
r102:	argininosuccinate synthetase
r103:	argininosuccinate lyase
r104:	N-acetyltransferase
r105:	acetylglutamate kinase
r106:	N-acetyl--glutamyl-phosphate reductase
r107:	acetylorithine aminotransferase
r108:	GLT N-acetyltransferase
r109:	GLT synthase

<i>Network Diagrams</i>	<i>Common Name</i>
r110, r112:	GLN synthase
r111:	carbamoyl-P synthetase
L1:	Light Non-Cyclic reaction
L2:	Light Cyclic reaction
I:	Complex I
II:	Complex II
III:	Complex III
IV:	Complex IV
V:	Complex V
AOX:	Alternative Oxidase
PSI:	Photosystem I
PSII:	Photosystem II
<i>Others</i>	
ED:	Entner-Doudoroff pathway
EMP:	Embden-Meyerhoff-Parnass pathway
EST:	Expressed Sequence Tags
ETC:	Electron transport chain
FBA:	Flux Balance Analysis
GSM:	Genome scale metabolic model
NPQ:	Non-photochemical quenching
OPPP:	Oxidative pentose phosphate pathway
OUC:	Ornithine urea cycle
PSI:	Photosystem I
PSII:	Photosystem II
TCA:	Tricarboxylic acid cycle
XC:	Xanthophyll cycle

Chapter 1

Introduction

1.1 Project Background

This project is part of the AccliPhot consortium <http://www.accliphot.eu/>, Marie Curie Initial Training Network funded by the European Commission which is composed of eleven theoretical, experimental and industrial partners. It aims to investigate and understand acclimation processes and photosynthetic mechanism of three organisms: *Arabidopsis thaliana*, a plant, *Chlamydomonas reinhardtii*, a green algae, and *Phaeodactylum tricornerutum*, a diatom. A specific aim is to optimise and upscale *P. tricornerutum* for the production of biofuel and high-value commodities, using experimental and theoretical methods.

In particular, the project aims to understand the metabolic behaviour of *P. tricornerutum* in different environmental and experimental conditions in order to optimise and identify potential routes of lipid production using mathematical modelling techniques.

1.2 Motivation

Nature World News reported that global carbon emissions rose to a record high in 2013, 36 billion metric tons, contributing 2.1% rise in projected emissions. That put 2013 global emissions from burning fossil fuels, at 61% above

1990 levels, the baseline year for the Kyoto Protocol [1]. Above that, British Petroleum has predicted that demand for energy will increase by 36% between 2011 and 2030 [2]. With increasing demand for energy and declining fossil fuels, there is a need to find alternative sources of energy in order to meet the future energy demand and to protect the environment from pollution and climate change.

Biofuels, derived from biomass such as crops, agricultural waste, algae, are a sustainable and renewable alternative source of energy, and have a number of advantages over fossil fuels: they have lower carbon emission rates, can be renewed in a short time, and raw materials for biofuel production are available widely across the globe. Biofuels have been classified into 4 generations, based on the carbon source from which they are derived. First generation biofuels are obtained from food crops (ethanol from corn and sugarcane, biodiesel from rapeseed, soybeans, and palm) but have the disadvantage that they compete for scarce cropland, fresh water and fertilizers, and have been blamed for rises in food prices [3, 4]. Second generation biofuels are made from non-food feedstock using cellulose or plant cell wall, however feedstock costs remain high due to processing and transportation [3, 4]. Third generation biofuels are made from nonfood feedstock, primarily algae. Algal biofuel has potential advantages over biofuels made from land plants, as they do not compete for cultivatable land [5]. Fourth generation biofuels are also made from algae and are based on the improvement of the algal strains, through metabolic engineering, for the effective and economical production of biofuel [6].

In 1978, the Aquatic Species Program (ASP) made significant contribution towards exploring biofuel production from lipid-accumulating algae. This study reported 50 algal strains that were promising for biofuel production and among which 60% were diatoms [7, 8].

Diatoms are photoautotrophic unicellular algae and contribute 25% of the total primary production on earth and 40% of the total marine primary production [9, 10]. The major storage molecules of diatoms are polysaccharides and lipids. Storage lipids are mainly in the form of triacylglycerols (TAGs), which is three fatty acids (FAs) (predominately saturated) bound to a glycerol backbone.

erol backbone, and can be transesterified to produce biodiesel. On the other hand, structural lipids in diatoms have a high content of polyunsaturated fatty acids (PUFAs) making them source of high-value commodities, such as ω fatty acids, essential for humans [11].

Biodiesel

Biodiesel is produced through the process of transesterification which involves reaction between TAG and a short-chain monohydric alcohol (e.g. methanol), in the presence of a catalyst such as sodium/potassium hydroxide, methoxide to form fatty acid alkyl esters (FAAE), commonly known as biodiesel. In the first step of transesterification, alcohol replaces one of the glycerol in TAG to produce FAAE and diacylglycerols (DAGs). The latter molecule further react with alcohol to liberate another molecule of FAAE and generate monoacylglycerols (MAGs). Lastly, MAG undergoes alcoholysis to yield glycerol and FAAE. During complete conversion, every molecule of TAG produces one molecule of glycerol and three molecules of biodiesel [12, 13].

1.3 Aims and Structure

The aim of this project is to construct metabolic models of *P. triornutum* and analyse them for metabolic responses under different environment. This investigation is an effort to investigate and understand the metabolism and identify the potential strategies to optimise the lipid production in order to produce better strains of *P. triornutum*. It also aims to design experiments, based on model analysis, in collaboration with the industrial partner of AccliPhot, Fermentalg, in order to improve the lipid production in this organism.

This thesis has following structure:

Chapter 1 gives the project background and introduces the biology of diatoms.

Chapter 2 describes the mathematical modelling techniques and model construction method that will be used in subsequent chapters.

Chapter 3 describes the construction and analysis of a genome scale metabolic models (GSMs) of *P. tricornutum*.

Chapter 4 describes the analysis of the GSM, constructed in Chapter 3, using mathematical techniques described in Chapter 2.

Chapter 5 describes the design of experiments, based on the model analysis, and the experimental results.

Chapter 6 is a general discussion of results presented in this thesis.

1.4 Diatoms and Biochemistry

Diatoms are a major component of phytoplankton communities and have over 250 genera, with approximately 100,000 species, and found in almost all aquatic habitats. They are well known for their unique geometrical shapes which are maintained by the large proportion of silica in their cell walls. The silica shell of the diatom, the ‘frustule’; has a bipartite structure. Each half-wall is known as a valve, and has many narrow bands which together form the girdle, as shown in figure 1.1. Valves vary in shape, symmetry as well as the arrangement of girdles and serve as the basis for the classification. Round and Crawford [16], divided diatoms into three classes, based on morphology and cytology: Coscinodiscophyceae, Fragilariophyceae and Bacillariophyceae. The Coscinodiscophyceae have a radially symmetrical orientation on their valves; the Fragilariophyceae have a bilaterally symmetrical pattern whilst the Bacillariophyceae have a similar pattern to that of the Fragilariophyceae, except that the valve has two slits in each half. However, this classification has been disputed from a phylogenetic point of view, suggesting that a centric valve structure occurred among all diatoms and the pennates evolved later from ancestors with centric structure [17]. Medlin and Kaczmarska [18] proposed a replacement for the Round and Crawford [16] scheme, based on interpretations of molecular data, in which the diatoms

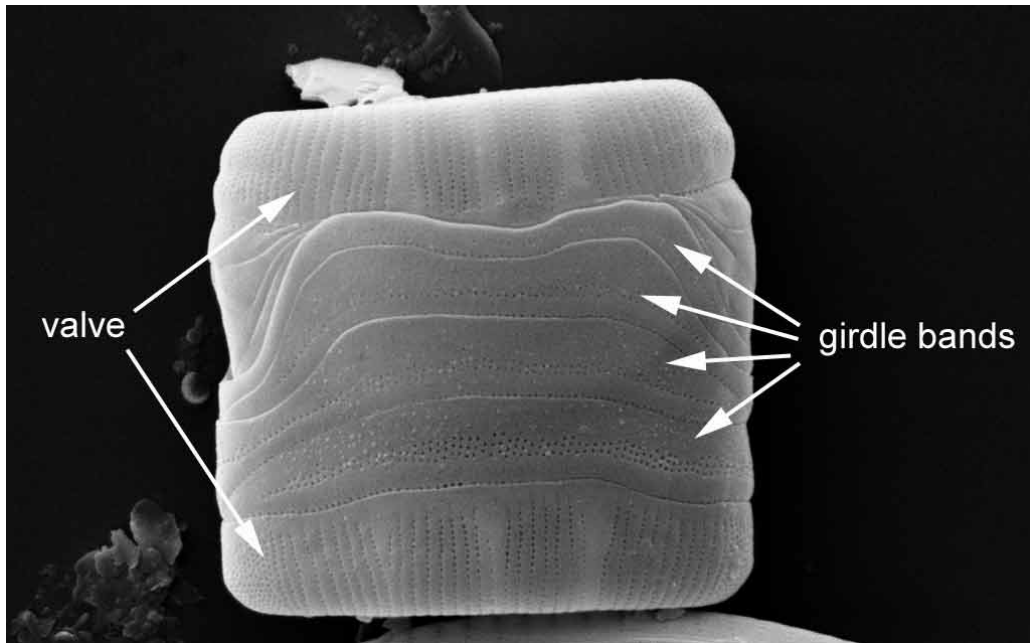


Figure 1.1: Frustule composed of valves, held together by deposition of band of silica girdle. Source: Kelly et al. [14].

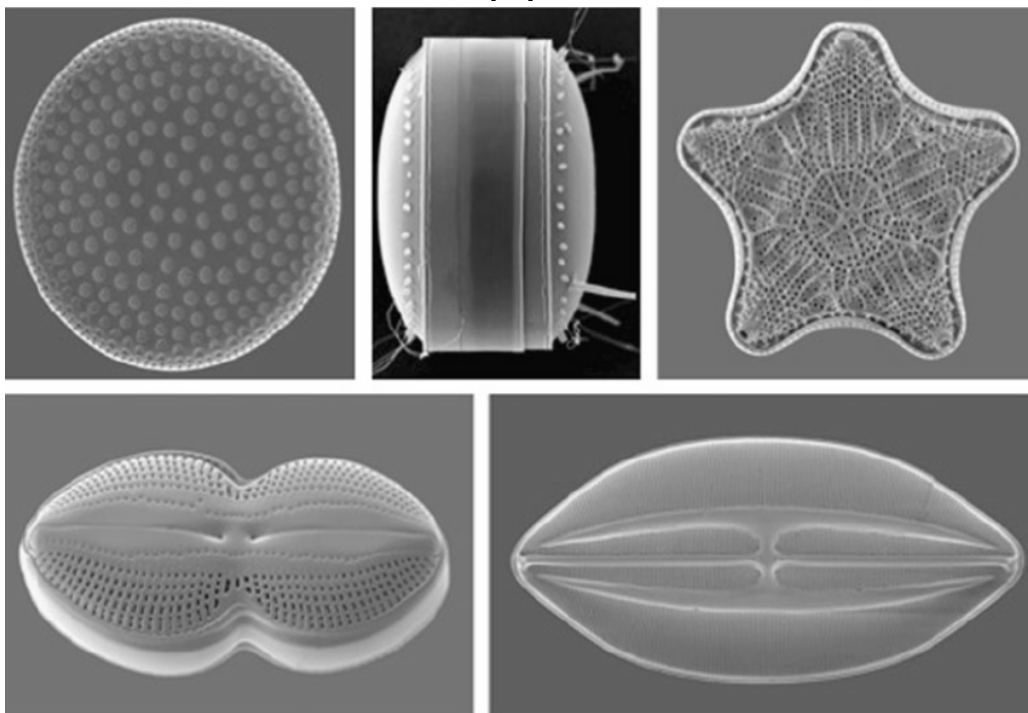


Figure 1.2: Top row: Centric diatoms with radially symmetrical orientation on their valves. Bottom row: Pennate diatoms with bilaterally symmetrical pattern. Source: Brayner et al. [15].

were divided into two subphyla: Coscinodiscophytina, which included the radial centrics pattern in valves, and Bacillariophytina, which included polar centrics as well as pennates. The latter was in turn divided into two classes, the Mediophyceae, containing the polar centrics, and Bacillariophyceae, containing the pennates. However, this classification was heavily criticised by Williams and Kociolek [19] due to the non-reproducible data. Regardless of the controversial classification system of diatoms, they are informally divided into two broad categories [20]: the centrics and the pennates, as shown in figure 1.2.

Early culture of diatoms dates back to 1880s; however Edgar and Edward [21] were the first to succeed in maintaining a stable diatom culture. For many years, *P. tricornutum* was the most frequently cultured species which was later followed by *Thalassiosira pseudonana* and *Chaetoceros calcitrans*. The early diatom research focused on cell structure and permeability [22]. The concept of using algae, including diatoms, for energy production is comparatively new (around 50 years) [23]. Diatoms have gained attention from researchers and industrialists in recent years, as potential sources of third and fourth generation biofuel.

1.4.1 Origin of Diatoms

The fossil record of diatoms was established based on recovery of their siliceous frustule and dates back to the early Jurassic age, about 120 million years ago [24]. However, molecular data suggests that they might have originated around 200 million years ago [17] and this at time might have been unsilicified [20]. The term ‘Diatomaceen’ was used first by O.F.Muller in 1786, and C. Nageli, in 1849, was the first to recognise that diatoms were single cell organisms. The question of diatoms belonging to the animal or plant kingdoms was discussed till 1853 but they were later confirmed to belong to the plant kingdom [22].

They arose as a result of secondary endosymbiosis between a photoautotrophic eukaryotic cell, close to red alga, and a heterotrophic eukaryote [25, 26] as illustrated in figure 1.3. During this process, the unicellular pho-

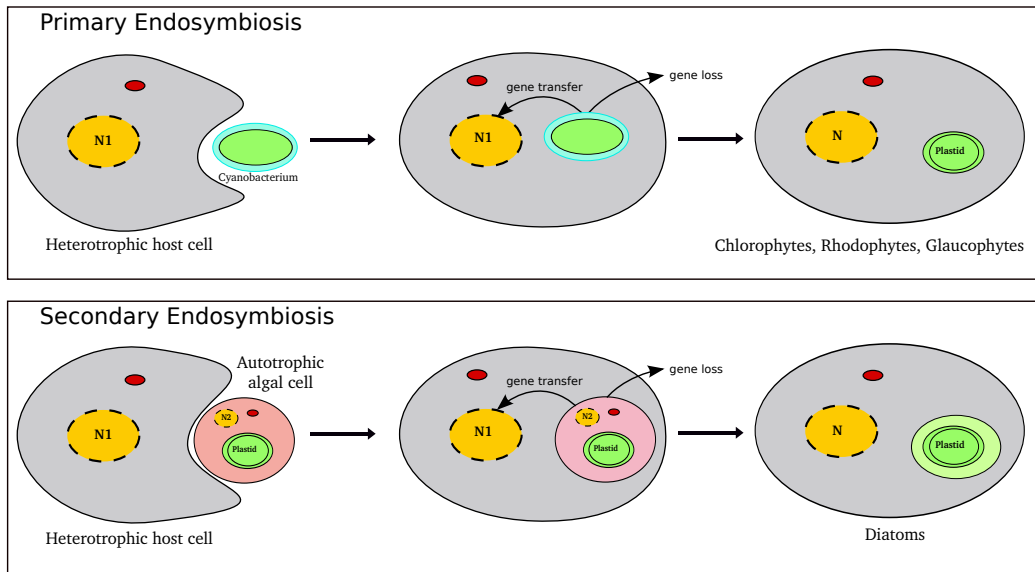


Figure 1.3: Schematic cartoon describing the putative origin of diatoms from secondary endosymbiosis and the difference between primary and secondary endosymbiosis. The upper panel shows the conventional primary endosymbiosis, the fusion of a heterotrophic ancestor and a cyanobacterium, that led to the origin of green algae, red algae, land plants and Glaucophytes. The lower panel shows the process of secondary endosymbiosis in which a heterotrophic host cell combined with green or/and red algae and subsequently lost the engulfed algal nucleus and mitochondrion leading to the origin of diatoms. Adapted from Wilhelm et al. [25].

toautotroph was reduced to a plastid which is evident from the additional membranes surrounding the plastid in modern diatoms. Further, this resulted in a secondary gene transfer and translocation of metabolic pathways. As a consequence, diatoms have unique features in their biochemistry and physiology that set them apart from higher plants.

1.4.2 Central Carbon Metabolism of Plants and Algae

Photosynthesis is the process by which plants, algae and some bacteria use light energy to fix inorganic carbon for the production of biomass. This process occurs in the chloroplast which is comprised of membrane-bound

structures known as thylakoids, where light is harvested to produce ATP and NADPH through the process of photophosphorylation, and a fluid filled space known as stroma, that contains enzymes of the Calvin cycle to fix the inorganic carbon utilising energy produced during photophosphorylation.

Photophosphorylation

The thylakoid membrane embeds set of integral membrane proteins: Photosystems (PSII and PSI), Cytochrome b_6f complex, and ATP synthase. PSI absorbs light in the wavelength range of 700 nm, and PSII absorbs light in, the wavelength range of 680 nm. Though PSI and PSII have different structures, both systems have a core complex involved in electron transport, and a peripheral antenna system responsible for harvesting light.

Light energy excites the electrons and promotes to higher energy state which is then captured by an electron acceptor, that differs between PSII and PSI. The electron from PSII travels through the electron transport chain (ETC), through cytochrome b_6f , to fill the 'hole' left by the loss of the excited electron of PSI, which is used to reduce first ferredoxin and then NADP^+ . The 'hole' left by the loss of the excited electron of PSII is filled with the electron from splitting of water. The proton released from splitting of water and additional proton pumped from the stroma to the lumen by cytochrome b_6f generate the proton gradient across the thylakoid membrane which drives the synthesis of ATP by chloroplast ATP synthase, as shown in figure 1.4. This complete process is also referred as non-cyclic electron flow and during this process two molecules of water are oxidised into one molecule of O_2 , 4 electrons and 4 protons. Alternatively, in cyclic electron flow light energy is utilised by PSI, without the involvement of PSII. The excited electron from PSI travels back to cytochrome b_6f which uses the energy of these electrons to pump protons from the stroma to the lumen generating proton gradient which drives the synthesis of ATP. During this process neither O_2 nor NADPH is generated, as shown in figure 1.4.

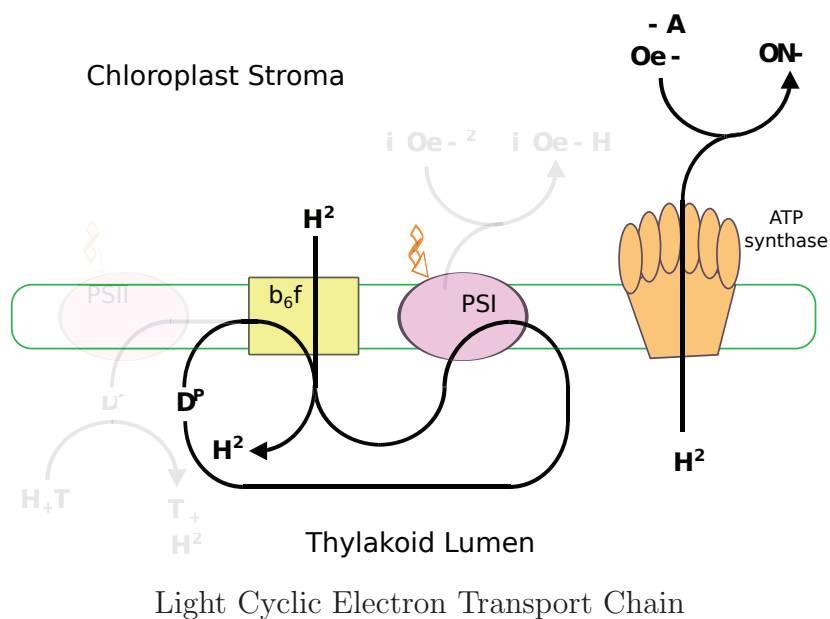
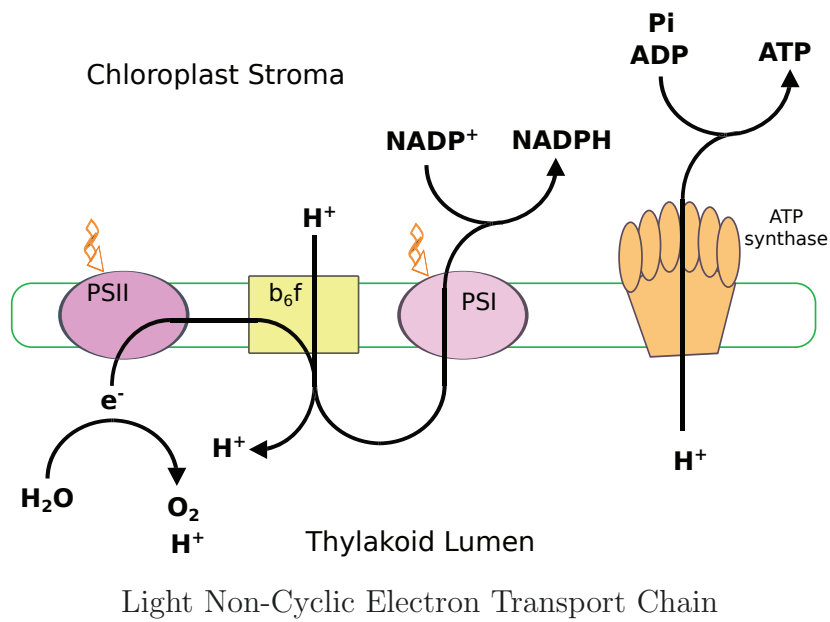


Figure 1.4: Photosynthetic electron transport chain of the thylakoid membrane. Light Non-Cyclic Electron Transport Chain: Light captured is used by PSII to oxidise water to molecular O_2 , electrons and protons. Electrons are transferred through cytochrome b_6f to PSI where $NADP^+$ is reduced to NADPH. The proton gradient across the thylakoid membrane is used to drive ATP synthesis. Light Cyclic Electron Transport Chain: Light energy is captured by PSI and the excited electron from PSI travels back to cytochrome b_6f generating proton gradient which drives the synthesis of ATP. PSI: Photosystem I, PSII: Photosystem II, b_6f : Cytochrome b_6f , Pi: Phosphate

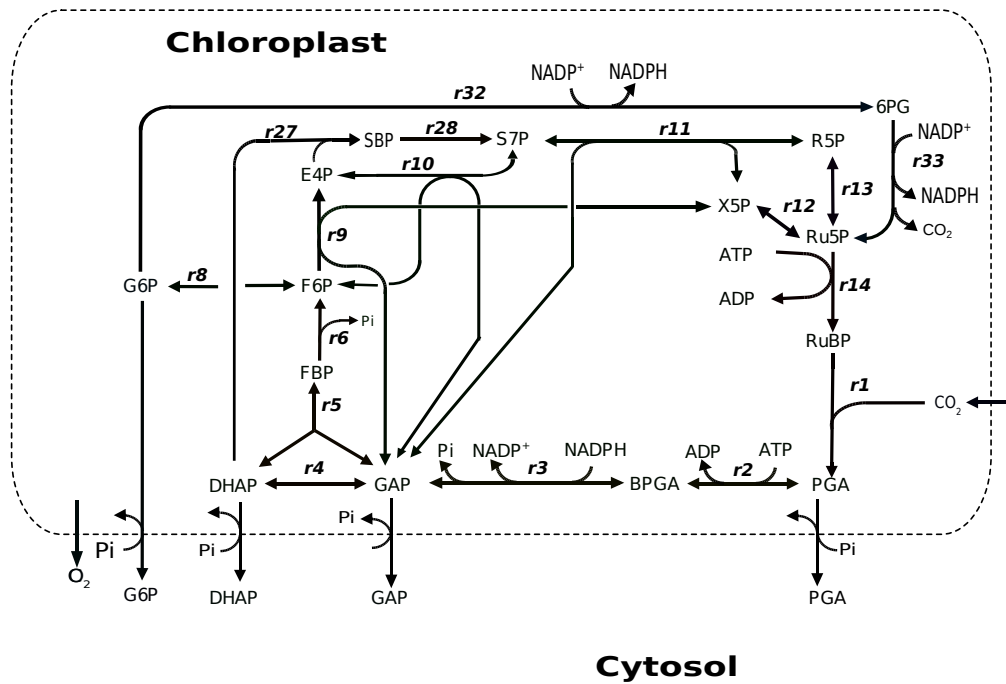


Figure 1.5: Chloroplast localisation of the Calvin cycle and OPPP enzymes in higher plants. Ru5PK (r14), GAPdh (r3), FBPase (r6), SBPase (r28) are activated by the thioredoxin system in the light while OPPP enzymes G6Pdh (r32), 6PGdh (r33) and TAL (r10) are inactivated. Reactions r1-r3, r6, r14, r28: Unique to the Calvin cycle, r32-r33, r10: Unique to OPPP, r4-r5, r8-r9, r11-r13: Shared reactions between the Calvin cycle and OPPP.

The Calvin Cycle and Oxidative Pentose Phosphate Pathway

The enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) (r1 in figure 1.5) catalyses the initial fixation of CO_2 in the Calvin cycle, by carboxylating one molecule of ribulose-1,5-bisphosphate (RuBP) to produce two molecules of 3-phospho-D-glycerate (PGA). Carboxylation is followed by the reduction phase, producing triose phosphate (TP), and the regenerative phase, resulting in the production of RuBP, thus completing the cycle.

Apart from the Calvin cycle, plants and algae also have a functional oxidative pentose phosphate pathway (OPPP) which shares many reactions and metabolites with the Calvin cycle, as shown in figure 1.5. The oxidative limb of this pathway produces ribose-5-phosphate (R5P) and CO_2 from

glucose-6-phosphate (G6P) while reducing NADP^+ to NADPH. This conversion is catalysed by series of reactions G6P dehydrogenase (G6Pdh) (r32), 6-phosphoglucono-lactonase (6PGlac) and 6-phosphogluconate dehydrogenase (6PGdh) (r33). The oxidative limb is followed by the reversible limb which consists of transaldolase (TAL) (r10) and reactions involved with the regenerative limb of the Calvin cycle. In higher plants, enzymes for both the Calvin cycle and OPPP are located within the chloroplast stroma, as shown in figure 1.5 and to avoid the futile cycle between the two processes, enzymes are regulated under dark and light conditions by the thioredoxin system [27]. The Calvin cycle enzymes, ribulose-5-phosphokinase (Ru5PK) (r14), glyceraldehyde 3-phosphate dehydrogenase (GAPdh) (r3), fructose biphosphatase (FBPase) (r6) and sedoheptulose-1,7-biphosphatase (SBPase) (r28) are activated in the light while the OPPP enzymes, G6Pdh(r32), 6PGdh(r33) and TAL(r10) are activated in dark [28–30].

Photorespiration

Rubisco is a bifunctional enzyme and, apart from carboxylation, catalyses the reaction between RuBP and molecular O_2 , resulting in the production of PGA and the two carbon compound, glycolate. The latter compound cannot be utilised directly by metabolism in higher plants and is ultimately recycled via glycine and serine to generate glycerate in the process known as the photorespiratory cycle. As shown in figure 1.6, the first step of glycolate metabolism involves its oxidation to glyoxlate, by glycolate oxidase (GOX) and/or by glycolate dehydrogenase (GDH)(pr₂), which is then metabolised to glycerate. Glycerate kinase (pr₉) catalyses the last reaction of the photorespiratory cycle and converts glycerate into PGA which can then enter the Calvin cycle. Glycerate kinase is energy dependent and the refixation of CO_2 and NH_4^+ produced during this metabolic process requires further energy, thus making photorespiration known as a wasteful process. However, this process provides metabolites such as glycine, serine, for other metabolic processes, and at high light and O_2 concentration, serves as an energy sink preventing the over-reduction of the photosynthetic components [31].

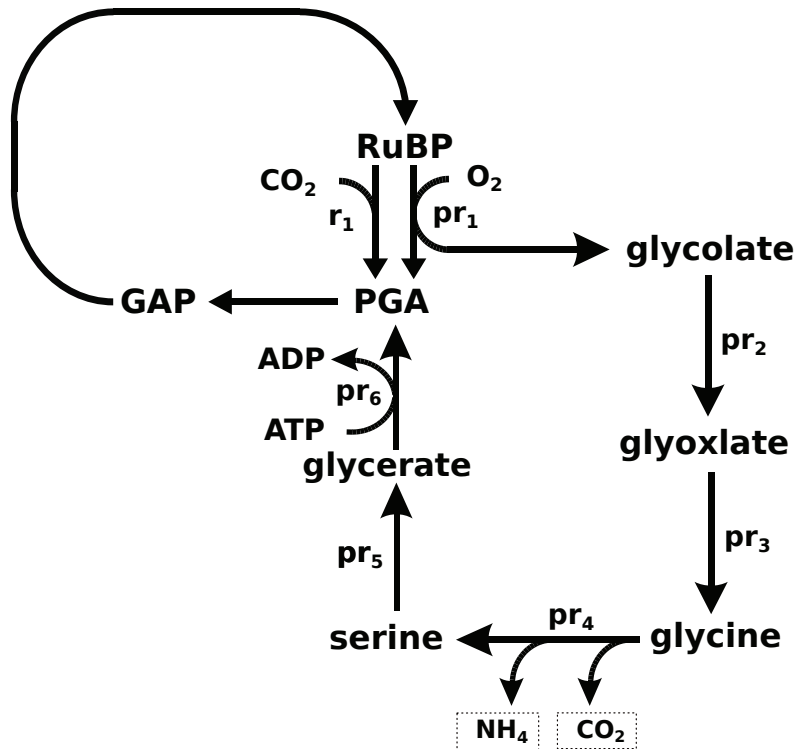


Figure 1.6: The process of photorespiration in higher plants. RuBP oxygenase (pr_1) catalyses the reaction between molecular O_2 and RuBP to produce glycolate, which is metabolised to glycerate through glycine and serine. Glycerate is metabolised to PGA, at the expense of ATP, which then enters the Calvin cycle. CO_2 and NH_4^+ produced during the process re-enters the metabolism. r_1 : RuBP carboxylase, pr_1 : RuBP oxygenase, pr_2 : glycolate oxidase (GOX) and/or glycolate dehydrogenase (GDH), pr_3 : glycine aminotransferase/alanine-glyoxylate aminotransferase, pr_4 : glycine decarboxylase and glycine hydroxymethyltransferase, pr_5 : serine-glyoxylate aminotransferase and glycerate dehydrogenase, pr_6 : glycerate kinase. Note that the process of photorespiration occurs across various compartments, which has not been considered in this schematic diagram.

Energy Dissipation Mechanisms

Photosynthetic organisms are exposed to varying light conditions and under steady-state conditions of excess light photophosphorylation delivers more NADPH than that consumed by the Calvin cycle for inorganic carbon fixation, leading to the over-reduction of the NADPH pool. In order to prevent the oxidative damage caused by the high energy, plants have developed a photo-protective mechanism known as non-photochemical quenching (NPQ) which is induced by activation of the xanthophyll cycles. Though, various xanthophyll pigments have been reported, the most common types, in plants, are violaxanthin (Vx), antheraxanthin (Ax) and zeaxanthin (Zx). The reversible conversion between these pigments leads to dissipation of excess energy as heat. Another xanthophyll cycle, mostly found in algal groups, and known as the diadinoxanthin cycle, comprises interconversion between diatoxanthin (Dtx) and diadinoxanthin (Ddx). Though, algae were long regarded to contain only the diadinoxanthin cycle, *P. tricornutum* possesses both violaxanthin and diadinoxanthin cycles [32]. All xanthophyll cycles transform the epoxidised xanthophylls to de-epoxidized xanthophylls in high light, facilitating energy dissipation, and their reversion to epoxidized xanthophylls in low light. However, this cycle is able to re-oxidise very small portion of reductant pool, about 0.05% of the total photosynthetic electron flux under light saturation for green algae [33–36].

Other well known mechanism to cope with the absorption of excessive light is through increased nitrogen assimilation. Plants and micro-organisms are capable of assimilating both NH_4^+ and NO_3^- as nitrogen sources. In plants, as well as diatoms, NO_3^- is reduced to nitrite by nitrate reductase [37] which is further reduced to NH_4^+ by nitrite reductase [38]. NH_4^+ is assimilated through glutamine synthetase (GS) and glutamate synthase (also known as glutamine oxoglutarate aminotransferase or GOGAT). In this metabolic pathway GS catalyses the amidation of glutamate (GLT) to generate glutamine (GLN). The second enzyme, GOGAT is responsible for the reductive transfer of nitrogen to 2-ketoglutarate (2KG) for the generation of two molecules of GLT, one of which is recycled for GLN biosynthesis [39]. The net stoichiom-

etry is production of 1 mole of GLT from 1 mole each of NH_4^+ , 2KG, ATP and NADPH. The other route of NH_4^+ assimilation is via glutamate dehydrogenase (GLTdh). GLTdh is a reversible reaction which inter-converts NH_4^+ and 2KG to GLT, utilising reductants. It has been known to use NAD^+ as a co-factor for the reaction in the direction of GLT deamination, while it uses NADPH as a co-factor for the reaction, in the direction of GLT synthesis [40]. The latter route of NH_4^+ assimilation is energetically less expensive; however, GLTdh has a low affinity for NH_4^+ and is activated only at high NH_4^+ concentration. In contrast, the GS-GOGAT cycle has a higher affinity for NH_4^+ but has a higher energy cost and has been suggested as an energy dissipation mechanism [41, 42].

Storage Products

The main storage products of plants and algae are starch and lipids. The process of starch synthesis starts in plastid with the formation of nucleotide activated glucose, by the ADP-glucose pyrophosphorylase. This is used as substrate by starch synthase to add glucose units to the end of the polymer chain to build up a starch molecule. There are two main types of glucose polymer: amylose where glucose monomers are linked by α -(1-4)-glucosidic bonds resulting in a linear, helical polymer that aggregates to form insoluble starch granules and amylopectin where α -(1-4)-glucosidic linked backbone has α -(1-6)-glucosidic linked branches forming more complex and branched structure [43].

The *de novo* biosynthesis of fatty acids occurs in the plastids. Plastidial acetyl-CoA acts as a precursor for fatty acid biosynthesis and can be synthesised from the conversion of pyruvate by the pyruvate dehydrogenase (PyrDh) complex and/or from the conversion of acetate by acetyl-CoA synthetase. In addition to these routes, acetyl-CoA can be transported across compartments via the carnitine shuttle pathway. This system is comprised of membrane proteins that transport L-carnitine, and carnitine acyltransferase enzymes that catalyse a freely reversible equilibrium between an acyl-CoA and an O-acyl-L-carnitine, to transport acyl groups between cellular com-

partments [44]. Triacylglycerols (TAGs), in general, can be produced either by incorporation of freshly synthesised fatty acid or by the conversion of pre-existing polar glycerolipids. TAGs generated by the first route contain high levels of C16:0 and C16:1 fatty acids, conversely, TAGs obtained from recycling of membrane lipids contain long chain C20:5 fatty acids [45].

1.4.3 Central Carbon Metabolism of Diatoms

The Calvin Cycle and OPPP

In contrast to higher plants, diatoms possess different localisation and regulation of the enzymes associated with the Calvin cycle and OPPP. As illustrated in figure 1.7, enzymes known to be associated with the Calvin cycle are distributed between chloroplast and cytosol. In plants, small subunits of rubisco is nuclear encoded and large subunit is chloroplast encoded. In diatoms both small and large subunits of the rubisco are encoded on the plastid genome, while all other Calvin cycle enzymes are encoded in the nuclear genome of *P. tricornutum*. These enzymes are later targeted to different compartments. Two genes encoding a cytosolic and a plastidial fructose biphosphate aldolase (FBPaldo) (reactions r5, r37 in figure 1.7) have been identified while four plastidial and one cytosolic FBPase genes (r6, r36) have been identified. Sedoheptulose-1,7-bisphosphate aldolase (SBPaldo) (r27) and SBPase (r28) are located in the cytosol rather than the plastid, as is the oxidative limb of the OPPP. Ru5PK (r14), GAPdh (r3), TAL (r10) which are known to be regulated by the thioredoxin system in higher plants, are not regulated in diatoms. The only plastidial enzyme that is known to be under redox regulation in diatoms is FBPase (r6) [25, 46].

Photorespiration

Diatom possesses genes for two enzymes that oxidises glycolate to glyoxylate, GOX, generally found in plants, and GDH, generally found in bacteria, although a gene for glycerate kinase was not identified in *P. tricornutum* [46]. Later, Fabris et al. [47] predicted the presence of this gene in *P. tri-*

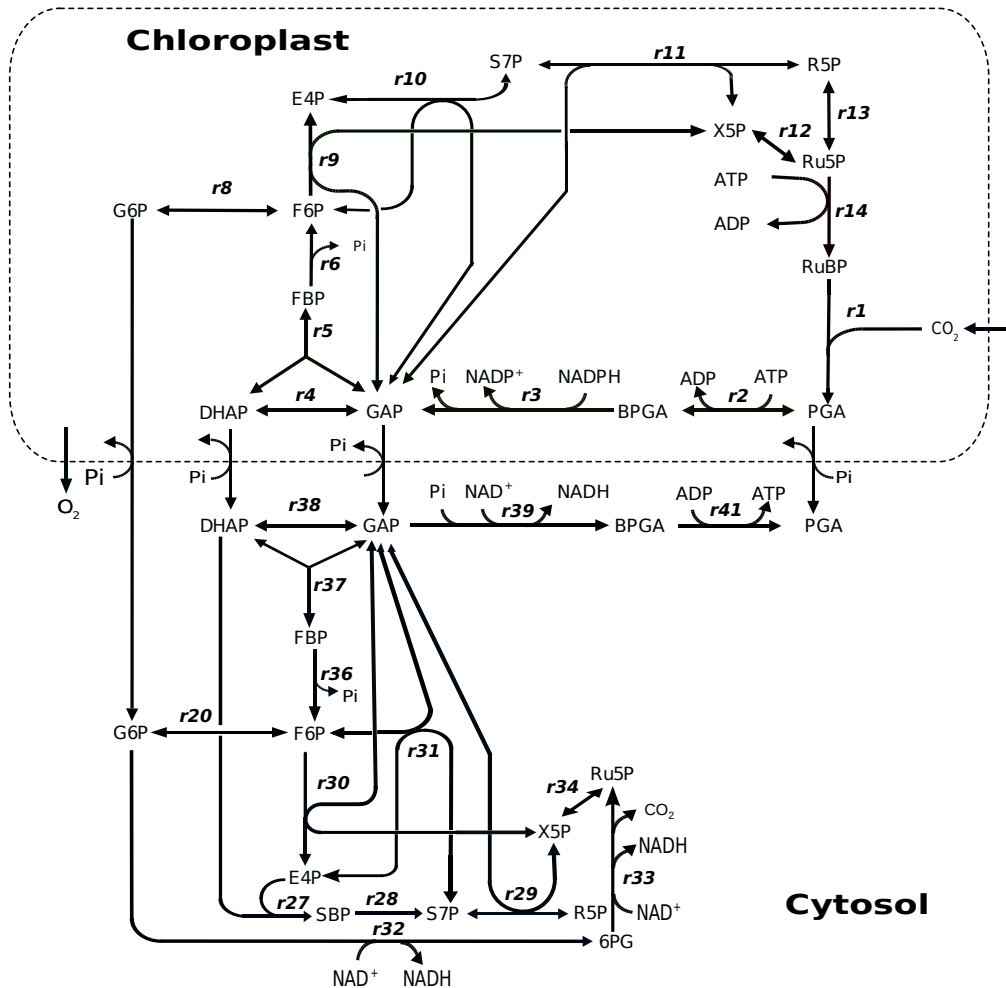


Figure 1.7: Localisation of the enzymes associated with the Calvin cycle and the OPPP in diatoms. Contrary to higher plants (shown in figure 1.5), G6Pdh (r32) and 6PGdh (r33), SBPase (r28) and SBPaldo (r27), are located in the cytosol. TAL (r10, r31) and TK (r9, r11, r29, r30) are present both in cytosol and chloroplast. The only enzyme known to be under thioredoxin regulation is plastidial FBPase (r6).

cornutum but the fate of glycolate in diatoms is not well known and there is no conclusive evidence for the photorespiratory pathways of higher plants existing in diatoms. Glycolate is known to be directly excreted under some circumstances such as exposure to high light [48, 49]. Investigations based on ^{14}C reveal glycolate being recycled to glycine and serine [50, 51]. Bromke [52], Zheng et al. [53] and Huang et al. [54] also reported that a large proportion of serine and glycine was synthesised from glycolate through photorespiratory reactions, at least under mixotrophic conditions. However, further metabolism of serine either through conventional photorespiratory route (i.e. via glycerate to PGA) or any other routes has not been reported.

The EMP, ED and Phosphoketolase Pathways

The Embden–Meyerhoff–Parnass (EMP) pathway, more commonly known as glycolysis, is an energy-generating pathway divided into two main phases. The upper phase of glycolysis (GlcK to TPI) involves conversion of glucose to TP with an initial investment of ATP in order to ultimately generate ATP and NADH in the lower phase (GAPdh to PK) by converting TP to pyruvate. Only a partial cytosolic glycolysis pathway is conserved in diatoms [55]. In *P. tricornutum*, glycolysis in the cytosol is not complete as enolase, which converts 2PG to PEP, is missing [56]. However, the lower half of glycolysis (TPI to PK) is conserved between diatoms and is predicted to be targeted to the mitochondrion [46].

Recently, the Entner–Doudoroff (ED) pathway, and phosphoketolase pathways, commonly found in prokaryotes, have been reported to be present in *P. tricornutum* [47]. The ED pathway is a more primitive glycolytic pathway than the EMP pathway, and is predominantly restricted to prokaryotic lineages. It was first discovered in 1952 in *Pseudomonas saccharophila* [57] and utilises 6-phosphogluconate dehydratase (6PGdht) and 2-keto-3-deoxy-6-phospho-D-gluconate aldolase (KDPGaldo) to convert glucose to pyruvate and GAP. The latter compound is further metabolised to pyruvate via lower glycolysis. It has a net yield of one molecule each of ATP, NADH and NADPH in contrast to conventional glycolysis (EMP pathway), which has

a net yield of two molecules of each ATP and NADH for every molecule of glucose consumed [58]. KDPGaldo in *P. tricornutum* was found to be of prokaryotic origin; however, orthologs have been identified in some eukaryotes such as *Cyanidioschyzon merolae*, *Ostreococcus tauri* and *Ostreococcus lucimarinus* and the associated genes were also identified in the centric diatom, *T. pseudonana* [47].

The phosphoketolase pathway is a catabolic variant of the pentose phosphate pathway usually present in lactic acid bacteria and bifidobacteria [59]. It converts xylulose-5-phosphate (X5P) to acetyl-P and GAP and/or converts fructose-6-phosphate (F6P) to acetyl-P and erythrose-4-phosphate (E4P) [60]. Eukaryotic phosphoketolases have been identified in a few fungal species [61]. Interestingly, the phosphoketolase enzyme in *P. tricornutum* was orthologous to those of Cyanobacteria and Lactobacillus, but no orthologs were found in *T. pseudonana* [47].

The Ornithine-Urea Cycle

The genome sequences of diatoms *T. pseudonana* [62] and *P. tricornutum* [63] revealed the presence of genes associated with the ornithine-urea cycle (OUC). As shown in figure 1.8, it comprises mitochondrial carbamoyl-phosphate synthase (ur_1) which catalyses the reaction between CO_2 and NH_4^+ to form the substrate for the urea cycle, carbamoyl-phosphate. This, along with ornithine, is converted to citrulline by the reaction catalysed by ornithine carbamoyltransferase (ur_2). In subsequent reactions, citrulline and aspartate are converted to arginine and fumarate, utilising energy. In the last step of the OUC cycle, arginine is cleaved by arginase (ur_5) producing urea and ornithine to complete the cycle. This cycle has been known to be present in vertebrates [64] in order to detoxify ammonia by synthesising urea but it is absent in green algae and plants. Diatoms can utilise both NH_4^+ and urea as a nitrogen source for growth and as there is no need to excrete urea as a waste product, the significance of the OUC in diatoms is not understood; however a role in reallocation of intracellular carbon and nitrogen into amino acids and polyamines has been suggested [65].

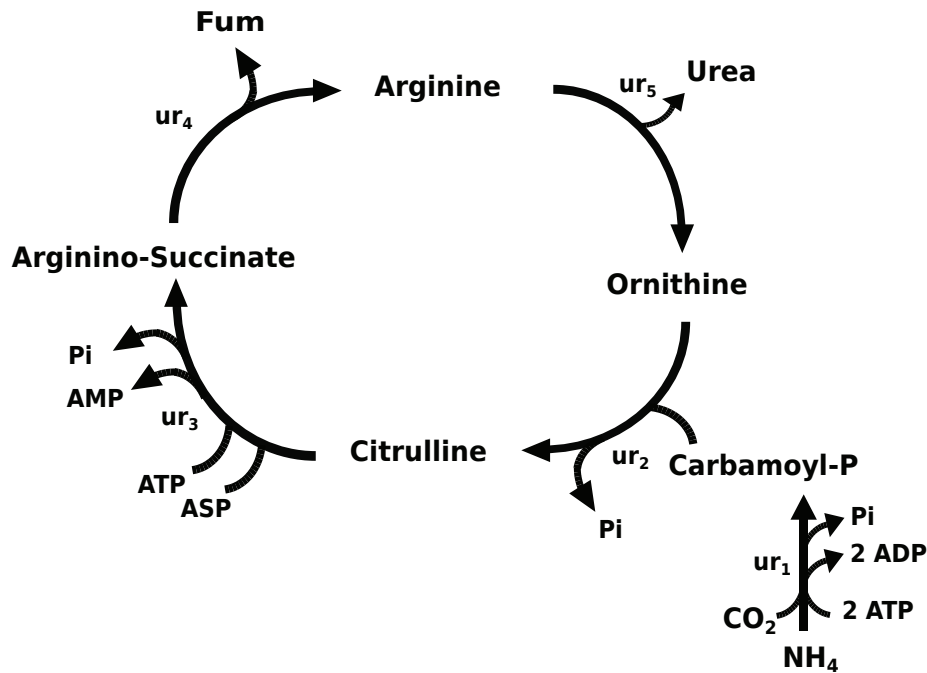


Figure 1.8: Ornithine urea cycle that converts excess ammonia into urea. The cycle consists of five reactions catalysed by enzymes carbamoyl-phosphate synthetase (ur_1), ornithine carbamoyltransferase (ur_2), argininosuccinate synthetase (ur_3), argininosuccinate lyase (ur_4), arginase (ur_5). The first two steps of the cycle occur in the mitochondrion and the rest in the cytosol.

Storage Products

The most abundant carbohydrate storage form in diatoms is chrysolaminaran [66–69]. It is a linear polymer of β -(1-3) and β -(1-6) linked glucose units in a ratio of 11:1 and is localised to vacuole (contrary to starch which has plastidial localisation). In addition, genome of diatoms also encodes for trehalose, mannose and mannitol synthesis pathways. The trehalose pathway in brown algae is inherited from red algae while mannitol pathways through lateral gene transfer from Actinobacteria. The genome of *P. tricornerutum* possesses a complete set of enzymes for synthesis of trehalose and mannose. Mannitol dehydrogenase, that catalyses the conversion of the mannose into mannitol, in the genome of *P. tricornerutum* is homologous with the mannose dehydrogenase enzyme from the brown algae *Ectocarpus siliculosus* [70].

Similar to other higher plants and algae, *P. tricornutum* encodes genes for both acetyl-CoA synthetase and Pyrdh complex. The former is targeted to chloroplast while the latter is targeted to mitochondrion [46]. It also encodes an acetyl-CoA transporter [71] however localisation is not clear (though TargetP localization server predicts a chloroplast transit peptide [72]). Lipids are generally stored in the form of TAGs, as lipid droplets (shown in figure 5.3), which may account for 80% of the total lipid content in the cell [73, 74].

Growth on Organic Carbon Sources

Microalgae can be cultivated under different cultivation modes such as photoautotrophy, heterotrophy, mixotrophy and photoheterotrophy. Under photoautotrophic (also referred as autotrophic) culture conditions, microalgae rely on solar energy and fix inorganic carbon to generate organic matters. Under heterotrophic cultivation, microalgae utilise organic compounds as carbon and energy source. However, some species of microalgae have been reported to assimilate organic sources only in the presence of light what is referred to as photoheterotrophy. Though light is utilised in the latter mode, inorganic carbon is not fixed. In the contrary, mixotrophic cultivation uses light energy to fix inorganic carbon and simultaneously utilise organic compounds as source of energy and carbon [75–78].

Although diatoms are phototrophic organisms, they can simultaneously assimilate organic carbon compounds such as acetate, glucose, glycerol, glycine, under mixotrophy. This has been reported to increase the growth rate, biomass concentration and lipid production in diatoms [79]. Among all organic carbon sources, glycerol has been reported to be absorbed more easily by the cells and significantly increase the lipid production, and has been widely studied as potential source of organic carbon for diatom culture [79–82].

The possibility of purely heterotrophic growth in diatoms is a subject of some controversy. Some species of *Navicula* has been reported to grow on glucose in the absence of light, however glycerol and fructose could support growth only in the presence of light i.e photoheterotrophy [83]. Alternatively,

Zheng et al. [53] reported that *P. tricornutum* could use glucose as a primary source of carbon only in presence of light but cannot sustain growth on glucose in the dark. Huang et al. [54] were able to sustain growth in heterotrophic condition however the growth rate was very low, showing that light was essential for normal growth.

1.4.4 *Phaeodactylum tricornutum*

P. tricornutum is a pennate diatom, belonging to the class Bacillariophyceae and the Phaeodactulaceae family, and is the only species in genus Phaeodactylum. *P. tricornutum* usually can exist in three forms, as shown in figure 1.9: fusiform, a normal cell type for pennate diatom with two arms, oval cells which have no arm and triradiate, which closely resemble fusiform except that they have three arms instead of two. These cells are weakly silicified, or can even completely lack silicon deposition, which made many researchers, in the early stage of classification, to believe that they were not diatoms [84]. In addition, the oval cell possesses only one valve and these unique features set the genus *Phaeodactylum*, which was described as new genus of unicellular algae by Bohlin in 1897, aside from other genera of diatoms [84].

T. pseudonana was the first diatom to have its genome sequenced, and has a genome of less than 34 Mb [62]. This was followed by the genome sequence of *P. tricornutum*, which has a genome size of less than 30 Mb [63]. Comparison of the expressed sequence tags (EST) of *P. tricornutum* [85] to those of a green algae, *C. reinhardtii*, a red algae, *Cyanidioschyzon merolae*, and a centric diatom, *T. pseudonana* [86] suggests that proteins are more highly conserved between *P. tricornutum* and the red algae than between *P. tricornutum* and green algae, which is in accordance with the evolutionary history that suggests diatoms originated as a result of secondary endosymbiosis between red algae and a heterotroph. Though pennate diatoms diverged from the centric diatoms only 90 millions years ago, the genome of *T. pseudonana* differs from that of *P. tricornutum* to the same extend as those of mammals and fish, which diverged approximately 550 millions years ago

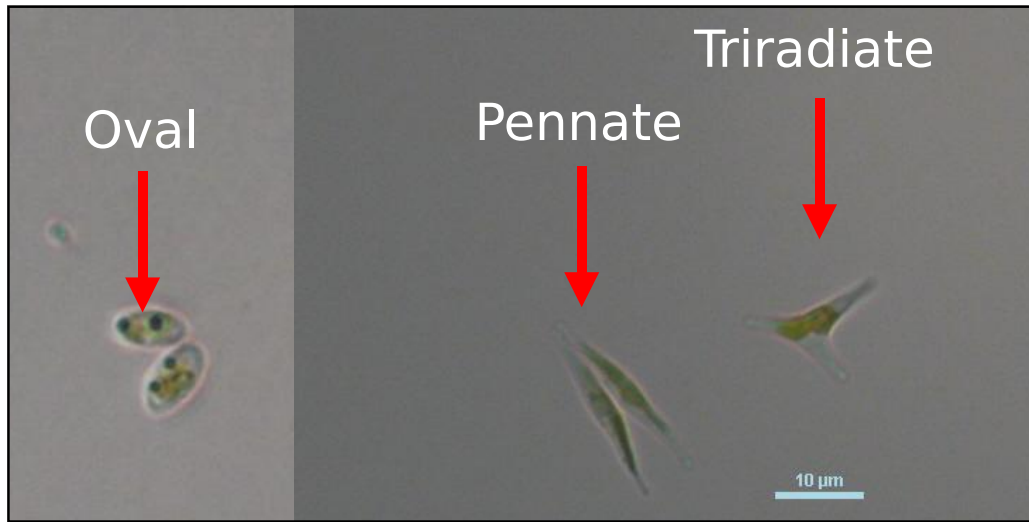


Figure 1.9: Different morphotypes of *P. tricornutum*. Image obtained using Nikon ECLIPSE 50i epifluorescence microscope connected to Nikon DS-Ri1 camera. Red arrows indicate the oval, pennate and triradiate forms.

[63]. These differences are mainly associated with genes involved in central carbon metabolism and the carbon-concentrating mechanism. The plastidial FBPase, cytosolic ATP-citrate lyases and enzymes associated with the phosphoketolase pathway have been reported to be present in *P. tricornutum* but not in *T. pseudonana*. The ATP-citrate lyase catalyses the reaction for the production of acetyl-CoA and OAA from citrate and CoA and its presence was also speculated as the reason for the higher capability of *P. tricornutum* to store lipid as compared to *T. pseudonana*, though the study from Yu et al. [87] reported similar percentage TAG accumulation, 14% of total dry weight, in both species under nitrate starved conditions. Regardless, *P. tricornutum* remains the preferred diatom for study as the biogenesis of the silicified frustule is facultative, which makes genetic material extraction and manipulation easier, bringing great advantages to the experimentalists. Moreover, it mainly stores lipid in the form of TAGs and long-chain polyunsaturated fatty acids (PUFAs), mostly in the form of eicosapentaenoic acid (EPA), making it potentially useful for the production and extraction of valuable lipids [88].

1.5 Conclusion

Algal-derived biofuels are promising candidates for the development of sustainable energy. However, exploitation of algae for biofuel production to sustain the economy not only requires improved technology for algal culture and downstream processing, but also focused research to understand the fundamental biological questions. Interestingly, diatoms are promising algal group for biofuels production. Among all diatoms, *P. tricornutum* is regarded as a model diatoms due to the availability of genome sequence, small genome size, its capability to store high percentage of lipid and grow in absence of silica which eases the experimental manipulation. Nevertheless, they have different biochemistry as compared to other algae and higher plants. Thus, there is a prior need to understand the metabolism of diatoms to enable physiological and genetic manipulation in order to improve the cost-effective production of biofuel.

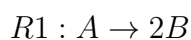
Metabolic models and the application of mathematical techniques over metabolic network, to certain extends, facilitates the analysis of the behaviour of a system *in silico* and to understand cellular phenotypes under specific physiological, environmental conditions and/or genetic manipulations. They allow in-depth insight into the complex networks and have been a promising way to explore and understand the metabolic capability of organisms. This work describes the construction and investigation of metabolic models of *P. tricornutum* to understand its metabolism, which can be further used to derive hypotheses and design experiments to enhance the performance of metabolism and the productivity of lipid synthesis.

Chapter 2

Mathematical Modelling of Metabolism

2.1 Introduction

Mathematical models are an abstract representation of a real-world system using mathematical concepts and language. They use a set of variables and equations to represent the properties of individual components and to describe the relationships between them [89]. They make it possible to perform “what-if” analysis, such as a change in experimental conditions, without the need to perform complicated and expensive experiments [90] and can be used to analyse the functional behaviour of a system. The fundamental basis for modelling metabolism is that the rate of change of concentration of a metabolite is equal to the difference in the sum of production and consumption. For instance: if reaction R1 converts one mole of metabolite A to two moles of metabolite B and reaction R2 converts one mole of metabolite B to one mole of metabolite C, the system can be represented as



Metabolic models therefore consist of a set of metabolic reactions and

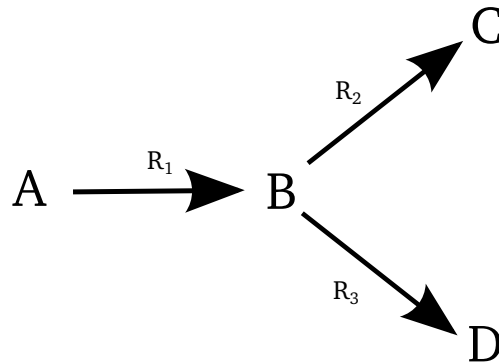
their associated metabolites. The stoichiometric coefficient of the substrates and products with respect to their reactions forms the basis of all metabolic models. These networks can be represented in a mathematical form, where variables are the metabolites and reactions and their kinetics establish the relationship between them [89]. Metabolites in the network are categorised as internal or external. Internal metabolites are produced and consumed as part of intracellular reactions in the system. External metabolites are those that are likely to be in constant exchange across the system boundary such as CO_2 , O_2 , to mimic the media composition. Metabolic models can be broadly classified as *structural* or *kinetic*. Structural models are defined purely in terms of the stoichiometry and thermodynamics, without any kinetic information. Kinetic models, on the other hand, additionally take into account the kinetic information but are generally restricted to smaller networks as the estimate of the kinetic parameters, such as substrate concentration, rate constant, maximum velocity of all enzyme catalysed reactions, is not trivial [89].

2.2 Structural Metabolic Modelling

Many important principles can be described by considering a simple hypothetical network such as that shown in figure 2.1. The rates of change of metabolite concentration can be expressed as a set of ordinary differential equations (ODEs):

$$\begin{aligned}
 dA/dt &= -v_{R1} \\
 dB/dt &= v_{R1} - v_{R2} - v_{R3} \\
 dC/dt &= v_{R2} \\
 dD/dt &= v_{R3}
 \end{aligned}
 \tag{eq.2.2.1}$$

This system of linear equations can be represented in matrix notation as solving large number of individual ODEs can be complicated. Moreover, for large networks, there exist an infinity possible solutions which can be represented as a vector space. Therefore, the right hand side of eq.2.2.1 can be represented in the form of a *stoichiometry matrix*, conventionally



$$\mathbf{N} = \begin{array}{ccc|c}
 & \mathbf{R}_1 & \mathbf{R}_2 & \mathbf{R}_3 & \\
 \begin{array}{c} \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \\ \mathbf{D} \end{array} & \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} &
 \end{array}$$

Figure 2.1: A simple hypothetical network and its associated stoichiometry matrix (\mathbf{N}). Column and row in the matrix represent reactions and metabolites respectively. Each element in the matrix represents the stoichiometric coefficient of the metabolite with respect to the associated reaction. Consumption of a metabolite by a reaction is denoted by a negative coefficient while production is denoted by a positive coefficient.

denoted by \mathbf{N} , of dimension $m \times n$ where rows m represent metabolites and column n represent reaction. Each element, n_{ij} , in the matrix represents the stoichiometric coefficient of the metabolite i with respect to reaction j . The production and consumption of metabolites by a reaction are indicated by positive and negative elements in \mathbf{N} respectively [89]. A stoichiometry matrix in which only the internal metabolites are represented is called an *internal* stoichiometry matrix, conversely a stoichiometry matrix in which the external metabolites are also considered is called an *external* stoichiometry matrix. Most analysis is based on the internal stoichiometry matrix.

The set of ODEs in eq.2.2.1 can be represented using matrix notation as:

$$\begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{dA}/\mathbf{dt} \\ \mathbf{dB}/\mathbf{dt} \\ \mathbf{dC}/\mathbf{dt} \\ \mathbf{dD}/\mathbf{dt} \end{bmatrix}$$

or

$$\mathbf{N} \cdot \mathbf{v} = \mathbf{dS}/\mathbf{dt} \quad (\text{eq.2.2.2})$$

where \mathbf{S} is the vector of the internal metabolite concentrations, \mathbf{N} is the stoichiometric matrix and \mathbf{v} is the vector of reaction rates.

2.2.1 The Steady State Assumption

In a biological system there is continuous exchange of matter between the system and the environment and most metabolism takes place close to the *steady state*, which implies that the rates of formation of internal metabolites is equal to the rates of utilisation. Therefore, at steady state, concentration of metabolites remain constant over time and eq.2.2.2 can be represented as:

$$\mathbf{N} \cdot \mathbf{v} = \mathbf{dS}/\mathbf{dt} = \mathbf{0} \quad (\text{eq.2.2.3})$$

eq.2.2.3 restricts the space of possible flux distribution however, (infinitely) many values of \mathbf{v} could still satisfy the equation and flux vector \mathbf{v} is under-determined. Most structural analysis techniques are based on the consideration of potential solutions (points in flux space) to this equation satisfying some additional criteria and/or to identify properties of all possible solutions.

2.2.2 Null-Space Analysis

Null-space of the stoichiometry matrix, \mathbf{N} , is the space of possible solutions that satisfies the steady state condition. Though null-space contains infinite number of solutions, it can be defined in terms of linear combinations of

columns of a *kernel* matrix, conventionally denoted \mathbf{K} and defined as:

$$\mathbf{N} \cdot \mathbf{K} = \mathbf{0}$$

In this matrix, row vector contains information on the reaction that must apply for all possible solutions and the number of columns represent the dimension of the space that contains all possible steady state solutions. The kernel matrix for the network represented in figure 2.1 is:

$$\mathbf{K} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \\ 1 & -1 \end{bmatrix} \begin{matrix} R_1 \\ R_2 \\ R_3 \end{matrix}$$

Here, the two linearly independent column vectors implies that all the possible solution at steady state lies in the two dimensional space as shown in figure 2.2. However, it should be noted that the column vectors of the kernel are not unique.

Application to Metabolic Networks

Since each row of the kernel matrix represents reaction, the number of invariant properties of reactions in a metabolic network can be identified directly from its kernel matrix. For example, the zero rows in the kernel matrix always correspond to reactions that cannot carry flux in steady state. These reactions are known as dead reactions and represent dead ends in metabolic network. The row vectors that are scalar multiple of each other correspond to reactions which carry flux in fixed proportion at steady state. These set of reactions are termed as *enzyme subsets* [91], and the corresponding enzymes are under co-ordinated control. For example, if we include additional reactions to figure 2.1 to that in figure 2.3 then \mathbf{K} in figure 2.3 is the kernel matrix for the expanded hypothetical network. Here, A_{tx} and R_1 share the row vector $[-1 \ -1]$, R_2 and C_{tx} share the row vector $[-1 \ 0]$, R_3 and D_{tx} share the row vector $[0 \ -1]$. Therefore, these pair of reactions represent enzyme subsets while R_4 has row vector $[0 \ 0]$ and represents a dead reaction.

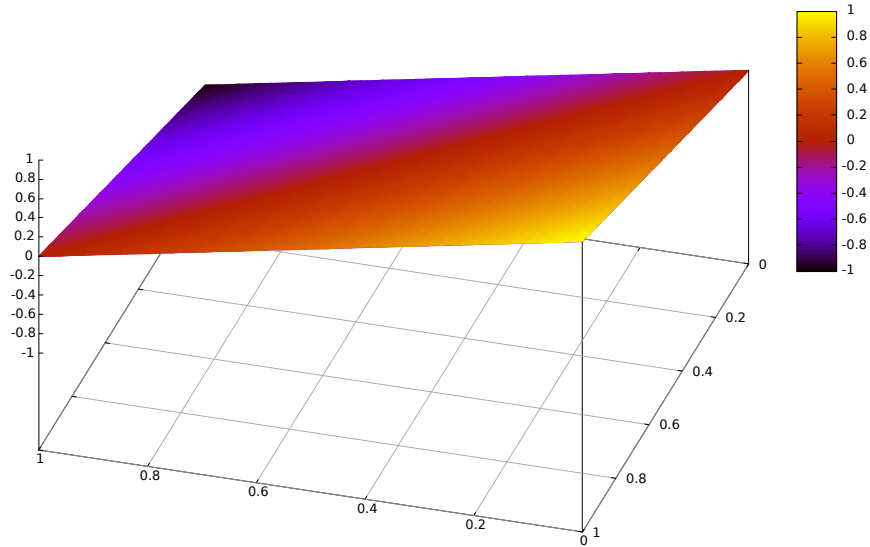
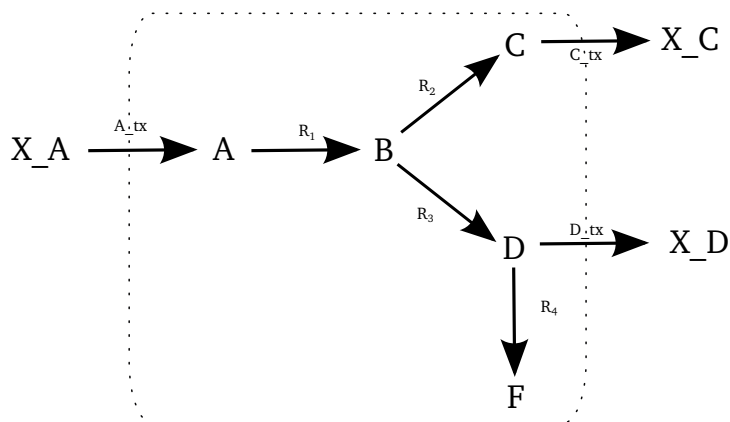


Figure 2.2: The two dimensional solution space for the three dimensional flux space illustrated in figure 2.1

Null-space analysis does not account for the reversibility. The kernel matrix \mathbf{K} would still be the same even if reaction R_1 in figure 2.3 was in opposite direction and row vectors would suggest that reactions, A_{tx} and R_1 , are in same subset. However, irreversible reactions in a subset defined in opposing direction can not carry flux at steady state and are referred as *inconsistent* enzyme subsets. This is usually due to missing reactions or incorrect directionality of the reactions which leads to discrepancy in model and should be fixed during the process of model curation, as will be discussed in section 2.3.1.

In the context of metabolic modelling, the term null-space generally refers to *right* null-space. Apart from that, *left* null-space for a matrix can be calculated which is the same as the right null-space of the transpose of \mathbf{N} :

$$\mathbf{K} \cdot \mathbf{N} = \mathbf{0}$$



$$\mathbf{K} = \begin{bmatrix} -1 & -1 \\ -1 & -1 \\ -1 & 0 \\ -1 & 0 \\ 0 & -1 \\ 0 & -1 \\ 0 & 0 \end{bmatrix} \begin{matrix} A_{tx} \\ R_1 \\ R_2 \\ C_{tx} \\ R_3 \\ D_{tx} \\ R_4 \end{matrix}$$

Figure 2.3: Represents a sample hypothetical network and its kernel matrix. Metabolites with prefix ‘x_’ represents external metabolites. These metabolites are transported in/out of the system using transport reactions denoted by ‘_tx’. The rows in the kernel matrix represents the reactions in the network while the column represents the dimension of the solution space.

While the *right* null-space contains the steady-state flux solutions, the *left* null-space reflects the linear dependencies between metabolites and thus, the moiety conservation relations which is true at all state and not just at the steady state. A typical example of moiety conservation in biological network is NAD^+ and NADH . Most of the reactions involved with NAD^+ and NADH do not consume them as a substrate. They are used as cofactors where one form is converted to the other form and the sum of their concentrations always remain constant. In figure 2.4, B and D are conserved moieties and their sum would be constant at all time. This property can be derived by the analysis of left null-space where conserved moieties have the same coefficient, as shown in figure 2.4.

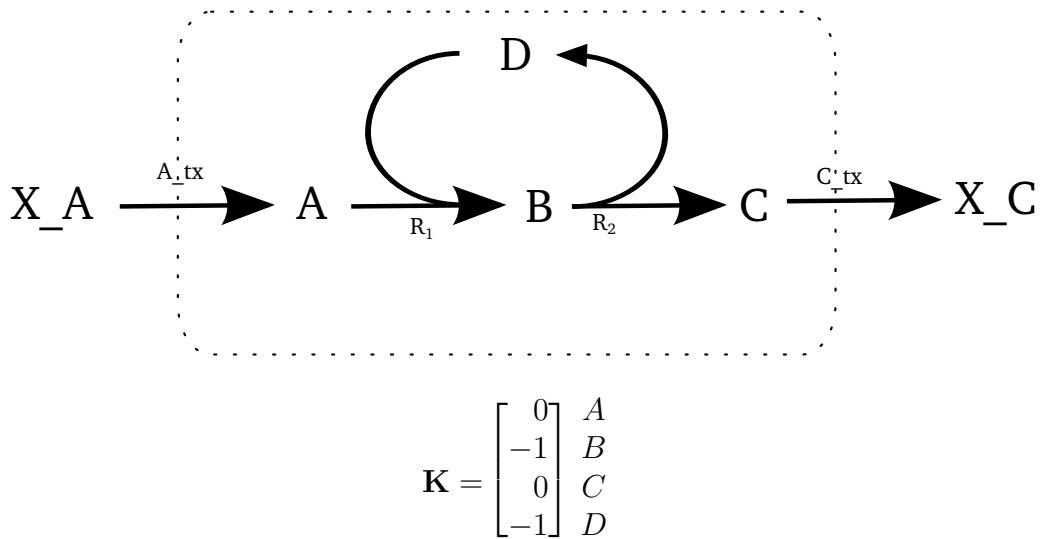
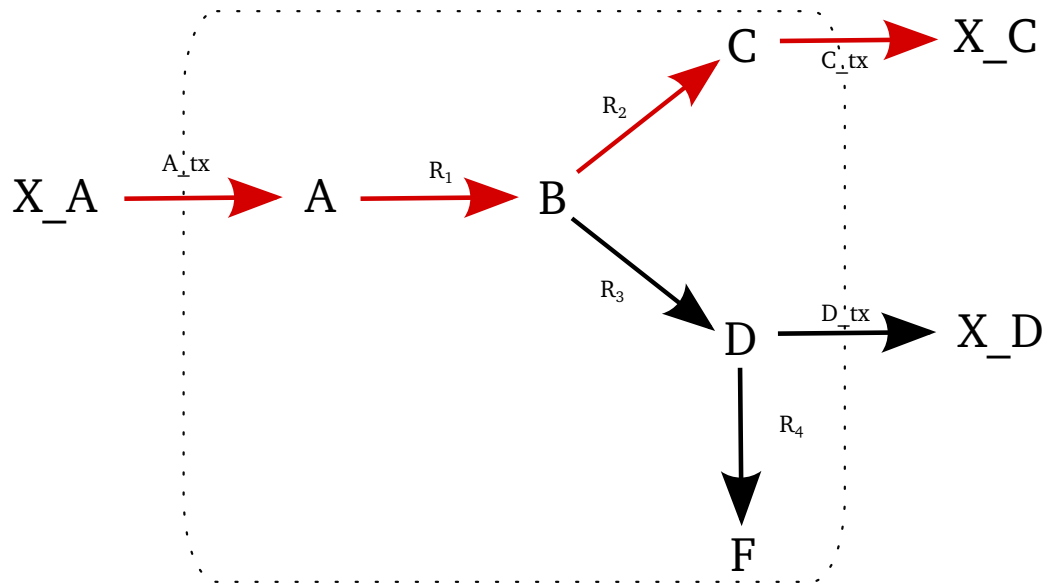


Figure 2.4: A simple hypothetical network representing the conservation relationship. Metabolites with prefix ‘x_’ represents external metabolites. These metabolites are transported in/out of the system using transport reactions denoted by ‘_tx’. The same coefficient for metabolite B and D in the left null space (\mathbf{K}) shows the conservation relationship between these metabolites.

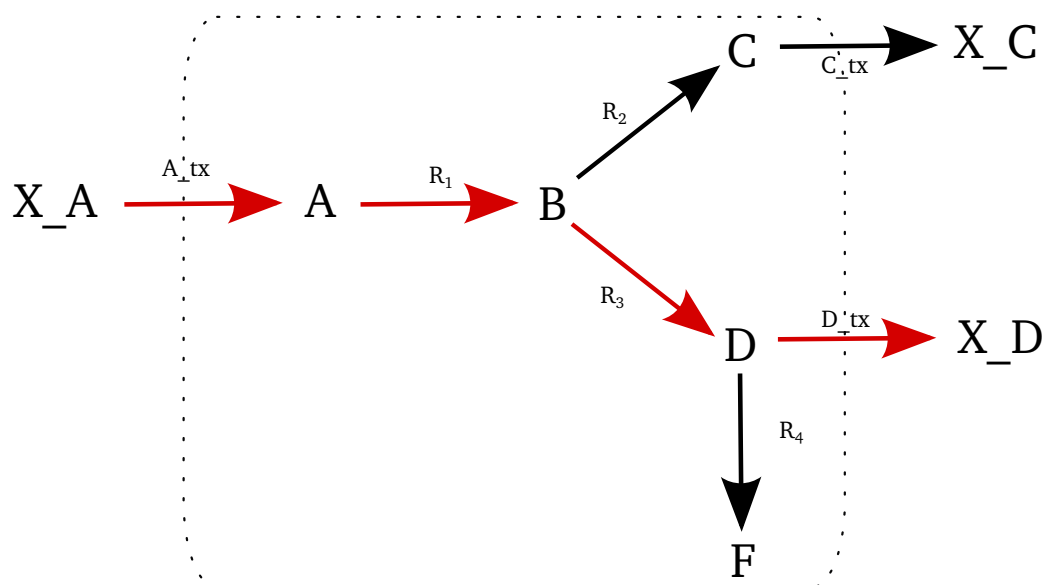
2.2.3 Elementary Modes Analysis

Although null-space analysis provides powerful methods to analyse the network properties, it does not take into account the reversibility of the reactions and hence not all columns of \mathbf{K} , which represent a pathway in the network, are biologically relevant. Moreover, the column vectors of the kernel are not unique. The *elementary modes* concept [92], overcomes these limitations.

Elementary modes are non-decomposable steady-state pathways in metabolic networks. They are all solutions in the flux space obtained by solving eq.2.2.3, i.e steady state constraint, along with the thermodynamic constraint and non-decomposability constraint. A thermodynamic constraint ensures that the irreversible reactions can carry flux only in given direction. The non-decomposability constraint ensures that each elementary mode is an independent minimal pathway in the network and removal of any reaction in an elementary mode will make it infeasible.



(a)



(b)

Figure 2.5: Elementary modes for a simple hypothetical network represented in figure 2.3. Reactions involved in elementary mode are represented with red arrow. There exist one elementary mode each to convert external metabolite x_A to metabolite x_C and x_D as represented in figure a and b respectively.

Figure 2.5 represents the elementary modes for the hypothetical network represented in figure 2.3 where first elementary mode converts one unit of x_A to one unit of x_C and second elementary mode converts one unit of x_A to one unit of x_D .

Elementary modes analysis is a relevant way of connecting the inputs to the outputs in a metabolic network. It makes possible to examine the network flexibility and indicates the importance of reactions in the network. It examines the property of metabolic network and has been successfully used for the purpose of metabolic engineering [93]. For example, Liao et al. [94] used elementary modes analysis to identify the most efficient 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) producing routes in *Escherichia coli* and optimized it by over-expressing enzymes of key reaction to achieve a high molar yield of DAHP. Trinh et al. [95] identified a set of elementary modes involved in joint production of biomass and ethanol and removing the desired gene, i.e. decoupling the biomass formation from ethanol production, enabled to design minimal metabolic functionality of an *E. coli* mutant for efficient production of ethanol. Van Dien and Lidstrom [96] and Poolman et al. [97] used elementary mode analysis to study the metabolic capabilities of *Methylobacterium exorequens* AM1 and physiology of the photosynthate metabolism of the chloroplast stroma respectively.

2.2.4 Flux Balance Analysis

Another approach to overcome the limitation of null-space and the problem of combinatorial explosion in elementary mode analysis is Flux Balance Analysis (FBA). It is a linear programming (LP) based approach for assigning fluxes to reactions, in a metabolic network, at steady state. It determines a solution which either minimises or maximises one or more fluxes in the network subject to a set of constraints, based on the law of mass conservation [98, 99]. Constraints, such as irreversibility of reactions, maximum biomass production, upper limits to the reaction rates, can be set to further constrain the feasible solution space. Specific feasible solutions can then be found that satisfies an optimality criteria. Commonly used objectives in this

study are maximisation of the growth rate for a fixed rate of carbon uptake [99] or minimisation of total flux in the network, as a proxy for economy of investment in enzymic machinery [100, 101]. The steady state assumption along with additional constraints, as defined in eq.2.2.4, is the basis for FBA.

$$\begin{array}{ll} \text{minimise|maximise} & : \mathbf{v}_{targs} \\ \text{subject to} & \left\{ \begin{array}{l} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v \leq \text{constraints} \end{array} \right. \end{array} \quad (\text{eq.2.2.4})$$

where:

- \mathbf{v} is the vector of reaction fluxes.
- $\mathbf{N}\mathbf{v} = \mathbf{0}$ defines steady-state.
- $v \leq \text{constraints}$ defines additional constraints such as thermodynamics constraints, demand for biomass production, limits on reaction rates.

LP approach was first applied on metabolic models by Fell and Small [98] and Watson [102]. Later, it was developed by Varma and Palsson [103, 104] with name FBA. FBA solution gives single flux vector unlike EM and Null-space analysis which gives all the properties of the steady state. Moreover, the LP solutions are not always unique and other possible solutions can exist. Nevertheless, FBA has been widely used to study the capability of metabolic network at different complexity [101, 105–110] and have resulted to improved fermentation conditions, media compositions, gene knockout targets (reviewed in [111, 112]). It also facilitates integration of experimental data such as experimentally measured biomass composition, maximum flux in a reaction.

For further detail on the aspects of FBA, null space analysis and elementary mode analysis readers are referred to Singh et al. [113], attached to the appendix.

2.3 Construction of Genome Scale Metabolic Models

The size of metabolic models can vary from a small set of reactions to large models, extended at the level of genome scale also referred as *Genome Scale Metabolic Models* (GSMs). GSMs describes the metabolic interactions in a given organism based on reaction networks predicted from enzymes encoded by the genome.

Construction of GSMs relies on the availability of annotated genomes associating genes with enzymes. Beside that, assigning correct stoichiometry and directionality to the reactions catalysed by these enzymes, and assuring the consistent naming of metabolites and reactions, are the prerequisites for metabolic model construction. While metabolic models of small networks can be constructed manually by listing a set of desired reactions along with their properties; the construction of large models becomes tedious, time consuming and, more importantly, error prone. These problems can be minimised, and to certain extent model construction can be automated by use of metabolic databases such as KEGG, BioCyc, which store information on genes encoding metabolic enzymes, stoichiometry of the enzyme catalysed reaction and the reversibility of reactions.

KEGG

KEGG (Kyoto Encyclopedia of Genes and Genomes) [114, 115] is a metabolic database that links the genomic information to metabolic function. It consist of three databases: GENES database stores the genomic information for all the completely sequenced genomes and some partial genomes with annotation of gene functions, PATHWAY database stores the functional information and represents the cellular processes such as metabolism, membrane transport, signal transduction and cell cycle etc, LIGAND database stores the information about chemical compounds, enzyme molecules and enzymatic reactions.

BioCyc

BioCyc [116, 117] is set of Pathway/Genome Databases (PGDBs) specific to organisms for which the genome has been sequenced and describes the link between genomic data and functional annotations of the genome including metabolic network. PGDBs are constructed using software Pathway Tools also known as PathoLogic [118] which can then be curated manually. The BioCyc databases are divided into three tiers, based on their quality: Tier 1 databases are curated based on literature study, for multiple years and are of highest quality, Tier 2 databases are computationally generated by the PathoLogic program and have undergone less than a year of curation and Tier 3 databases contain computationally predicted metabolic pathways with no manual curation. In addition to reactions predicted by the genome annotation, PGDBs also contains ‘gap filling reactions’, added by Pathway Tools software. These reactions are based on the assumption that if some, but not all, the genes encoding for enzymes in a pathway are annotated then it is highly likely that the missing gene is the result of genome misannotation. While it is possible to misannotate the genes which can lead to gaps in the network, some of the gap filling reactions are also likely to be false positive, specially when the concept of pathway itself is questionable.

MetaCyc [119, 120] is one of the manually curated PGDB of experimentally elucidated metabolic pathways from all domains of life. It contains 2453 metabolic pathways, involved in both primary and secondary metabolism, as well as associated metabolites, reactions, enzymes, and genes, from 2788 different organisms. DiatomCyc [47] is a PGDB that was constructed based on genome sequence of *P. tricornutum*. It contains 286 metabolic routes, 1719 reactions and 1613 assigned enzymes. It has been curated manually to fill the gaps such as EPA synthesis, photorespiration pathway.

Owing to the availability of organism specific databases in BioCyc and comparatively consistent metabolite naming convention as compared to KEGG [121], MetaCyc and DiatomCyc, along with previously published model of *P. tricornutum* by Hunt et al. [122], will be used for the model construction and curation in this thesis.

2.3.1 Model Construction and Curation

The first step of model construction is to extract the set of reactions encoded by the genome. The source for this complete set of reactions can be an organism specific database or reactions list obtained from organism specific annotated genome or metabolic reconstruction.

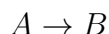
Models built directly from databases acquire any inconsistencies and inaccuracies in the database. They usually contain stoichiometrically unbalanced reactions and/or incorrect empirical formula of the reactants that leads to violation of mass conservation. The missing reactions and inconsistent metabolite identifiers generate discontinuities in the metabolic network while the absence of reversible/irreversible criteria leads to inconsistent enzyme subsets. Hence, the construction of such models often require several iterations of curation mentioned below.

In order to ensure mass conservation, the atomic balance of individual reactions are determined and all the reactions in the model are balanced with respect to carbon (C), nitrogen (N), phosphorus (P), and sulphur (S), as they constitute the major atomic composition of biomass. In our study, oxygen (O) and hydrogen (H) were not considered for atomic balance as water is treated as an external metabolite. Metabolites with unknown atomic composition and reactions involved with such metabolites are replaced, where possible, by balanced reactions with known atomic composition of metabolites, otherwise they are removed. For instance, reactions involved with non-metabolic species (e.g. Charged-MET-tRNAs) are removed from the model while reactions involved with generic compounds (e.g. Alcohols) are replaced with specific metabolites and stoichiometric coefficient balanced on both sides of the reaction. Another cause of mass inconsistencies is the presence of reactions involved with polymers defined in terms of disproportionate number of monomeric units or in terms of metabolite with unknown empirical formulas. For example, MetaCyc reports an amylase reaction where one unit of starch is degraded to one unit each of maltose and glucose:



This reaction would be atomically balanced if starch contains 18C atoms

however, such polymers are usually represented with an -R group in the empirical formula where number of C in -R group is not specified. An algorithm to detect material inconsistencies solely based on the reactions stoichiometry, described by Gevorgyan et al. [123], can be applied even if the atomic composition is not known. For instance, the stoichiometry of reaction



reflects that A and B has same empirical formula. However, if there exists a reaction with stoichiometry



in the network then the two reactions can only be consistent if C had mass of zero. Such metabolites are detected based on left null space on the stoichiometry matrix.

Metabolites with inconsistent identifiers are traced and renamed to avoid the artificial disconnected network. For instance, MetaCyc uses identifier CPD-448 and GLC-1-P for the metabolite α -glucose-1-phosphate and β -glucose-1-phosphate respectively. From the metabolic modelling perspective, the α form of the structure is not differentiated from the β form and to avoid the discontinuity in the network both metabolites are renamed to be regarded as same metabolite. Similarly, reactions that can be involved with either NADH or NADPH form of reductant uses the metabolite identifier as NADH-P-OR-NOP. In a metabolic model, this is treated as a different metabolite and such reactions are corrected for the metabolite identifiers and represented by two separate reactions where one involves NADH while other involves NADPH.

The possible cause of biomass infeasibility are the inconsistent enzyme subset, which usually are results of discontinuity in the network and/or inappropriate reversibility of the reactions, and were resolved by adding appropriate reactions to the network or changing the reversibility, as described in section 3.2.3.

2.4 Metabolic Modelling Tools

The size and complexity of biological networks make not only the construction but also the analysis of network, time consuming and error prone. Hence, to increase productivity, ensure reproducibility and avoid human generated error, computer programs have been widely used for model construction and analysis. These programs are used to convert the file specific properties of the model, such as stoichiometry and kinetic information, to a mathematical object so that the system properties can be analysed by performing mathematical operations.

Several programs have been developed for simulation of models such as BIOSSIM [124], METASIM [125], SCAMP [126], FACSIMILE [127], Copasi [128, 129], METATOOL [91], CellNetAnalyzer [130], COBRA Toolbox [131], ScrumPy [132]. METATOOL is one of the earlier software package developed for structural modelling. It implemented algorithms for elementary mode analysis and enzyme subsets using script based user interface. The COBRA Toolbox in MATLAB is another modelling package widely used for FBA analysis on GSMs.

The ScrumPy package is primarily written in Python and supports both structural and kinetic modelling. It has implemented algorithms for elementary mode analysis, enzyme subsets, conserved moieties, RCC, FBA and has been widely used for the construction and analysis of large and small models [101, 108, 133–135]. All the modelling work in this thesis has been undertaken with ScrumPy. It is released under the Gnu Public Licence, and available for download from <http://mudshark.brookes.ac.uk/ScrumPy>

2.4.1 The Python Programming Language

Python is a high level, dynamic programming language, widely used for variety of applications. It is a free, open source, platform independent programming language that has readable syntax and console based interactive development environment. The intrinsic Python distribution supports data structures such as string, float, lists, tuples, dictionaries which are flexible

and easy to use. It also supports standard libraries such as SciPy, NumPy that contain statistical, numerical and scientific tools for scientific and engineering applications. In addition, it can easily be integrated with other programming languages such as C, C++ using variety of integration mechanisms and can invoke C and C++ libraries. Python also supports the concept of software reuse mechanism such as object-oriented programming (OOP), where data and functions can be encapsulated in objects which in turn is the instance of a class. Such paradigm provides encapsulation of data and methods, avoiding the outer interference and misuse, allows object to inherit properties from different classes and process differently depending on their data type or class and promotes modular and structured management of the information. For more detail on Python, reader are referred to any addition of Lutz and Ascher [136], Lutz [137].

2.4.2 ScrumPy: Metabolic Modelling with Python

ScrumPy provides a high-level modelling interface that utilises and extends the low-level features of Python and standard libraries such as SciPy, NumPy, making it non-obligatory to work with low-level algorithms and advanced programming techniques. In addition, PyoCyc module in ScrumPy facilitates the extraction and mapping of information such as genes, proteins, enzymes, reactions, metabolites, from the local copy of PGDB flatfiles.

ScrumPy models are defined using simple, plain-text (ASCII) files, usually with file extension ‘.spy’. It contain four distinct types of entries:

- Directives - optional commands that specify how a model is to be processed
- Comments - to aid human readability
- Reactions - mandatory content that describe an individual reaction in terms of name, reaction stoichiometry and rate law
- Initialisation - describes parameters and initial metabolite concentrations

An example of ScrumPy model is as follows:

```
Structural()  
#A simple toy model  
R_1:  
    A <> B  
    ~
```

where, the directive `Structural()` specifies that it is a structural model and kinetic information is not to be processed which improves the computation time. Any line in model beginning with `#` represents comment and is not processed. Reaction, `R_1`, represents a reversible reaction. The tilde (`~`) represents ScrumPy's default kinetic law mass action with equilibrium and rate constants of one. In a structural model, the tilde serves as an end of reaction marker. Additional directives are present in ScrumPy to facilitate the model analysis such as `Include()` facilitates to load model spread in multiple files, `ElType()` directive specifies the data type (boolean, float, integer) to be used in the generation of the stoichiometry matrices, `External()` directive is used to declare external metabolites.

A metabolic model, in ScrumPy, can be constructed by extracting the desired reactions using `PyoCyc` modules and its attributes. This property also allows to semi-automate the process of model construction. For instance, provided a list of enzyme commission (EC) numbers the reactions can be extracted and stored in ScrumPy format as shown in figure 2.6.

First, the database is loaded as data object, by typing into the command line interface:

```
>>> db = PyoCyc.Organism('DiatomCyc')
```

where the location of the PGDB flatfile is passed as an argument. The list of (EC) number is looped to extract the reaction from the loaded database object. Reaction has an attribute called `'AsScrumPy'` which returns ScrumPy format of reaction. This can be written in text file and saved as `'spy'` file


```

*ScrumPy - scrapy devel branch*
File Edit Debug Windows Help
ScrumPy - Metabolic Modelling in Python

Doc is out of date in places at time of writing,
(but a jolly good read for all that)

Python 2.7 version.

Version : 1.0-BETA
Last update : 30/9/2016
Revision : 1083
scipy dev branch
Python 2.7.6 (default, Oct 26 2016, 20:30:19)
[GCC 4.8.4] on linux2
Type "copyright", "credits" or "license" for more information.
IDLE 0.8 -- press F1 for help
>>> from Bioinf import PyoCyc
>>> db = PyoCyc.Organism('DiatomCyc')
1 Reaction
1 Enzrxn
1 Compound
1 Protein
1 Gene
1 Pathway
1 Types
>>> ecs = ['2.7.2.1','5.1.3.1','4.1.2.9','2.7.2.1']
>>> for e in ecs[0:1]:
        for r in db[e]:
            print r.AsScrumPy()

"ACETATEKIN-RXN":
  "ATP" + "ACET"<-"PROTON" + "ADP" + "ACETYL-P"
  ~
  #RIGHT-TO-LEFT

"RXN-7958":
  "ATP" + "PROPIONATE"<>"PROTON" + "ADP" + "PROPIONYL-P"
  ~

>>> spy=open("PK.spy","w")
>>> spy.write('# Phosphoketolase pathway\n')
>>> for e in ec:
    str = "# " + e + "\n"
    for r in db[e]:
        spy.write(str)
        spy.write(r.AsScrumPy())
    spy.close()
Ln: 57 Col: 4

```

Figure 2.6: PyoCyc Module and the process of metabolic model construction in ScrumPy.

extension. In a similar fashion, a raw GSM can be generated by extracting all the reactions present in the organism specific PGDB file using the PyoCyc module.

Models, once constructed, are loaded as data object and is initialised by typing into the command line interface:

```
>>> model = ScrumPy.Model('Toy\_Model.spy')
```

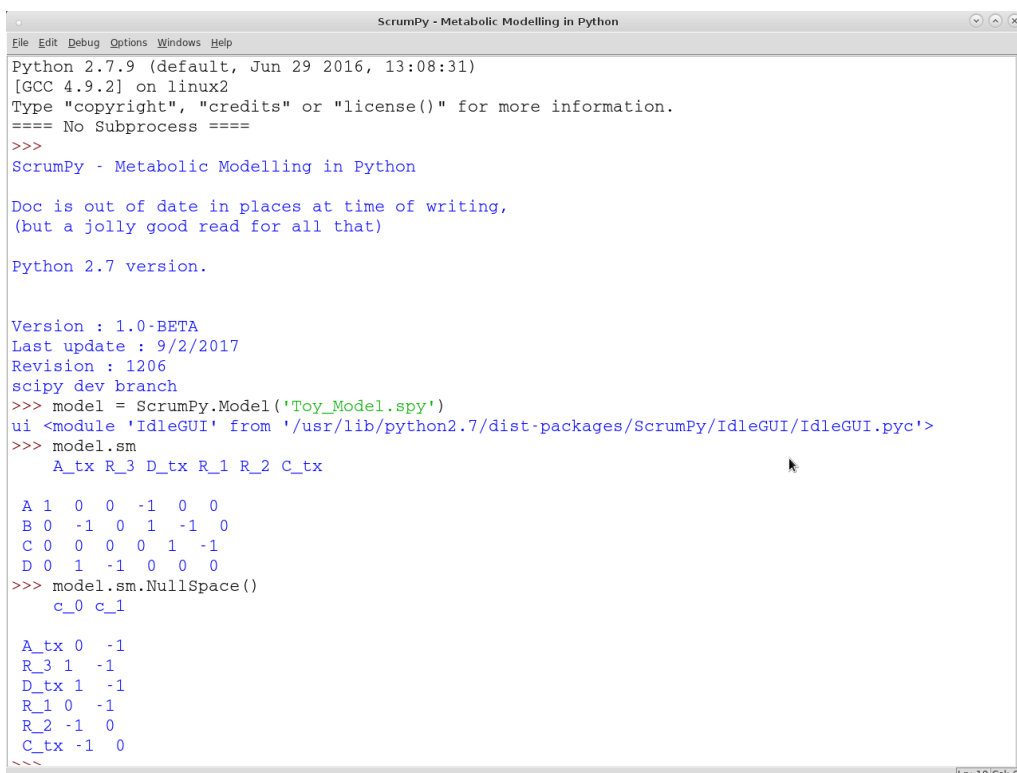
Most of the structural (as well as kinetic) properties such as the stoichiometric matrix, kernel matrix, elementary modes, enzyme subsets are available as model attributes and can be obtained by typing:

```
>>> model.sm  
>>> model.sm.NullSpace()  
>>> model.ElModes()  
>>> model.EnzSubsets()
```

respectively, as also demonstrated in figure 2.7. Most of the common model analysis tasks are available as attributes of the model and can be explored by typing:

```
>>> dir(model)
```

Apart from in-built attributes, more specific and complex analysis of the model can also be performed by importing python modules written to perform specific task. The ScrumPy attributes and additional modules used for the model construction and analysis, in this thesis, can be found in appendix A. The process of model construction and analysis of GSMs of *P. tricornutum*, using methodology covered in this chapter, will be discussed in next chapter.



```
ScrumPy - Metabolic Modelling in Python
File Edit Debug Options Windows Help
Python 2.7.9 (default, Jun 29 2016, 13:08:31)
[GCC 4.9.2] on linux2
Type "copyright", "credits" or "license()" for more information.
==== No Subprocess ====
>>>
ScrumPy - Metabolic Modelling in Python

Doc is out of date in places at time of writing,
(But a jolly good read for all that)

Python 2.7 version.

Version : 1.0-BETA
Last update : 9/2/2017
Revision : 1206
scipy dev branch
>>> model = ScrumPy.Model('Toy_Model.spy')
ui <module 'IdleGUI' from '/usr/lib/python2.7/dist-packages/ScrumPy/IdleGUI/IdleGUI.pyc'>
>>> model.sm
  A_tx R_3 D_tx R_1 R_2 C_tx

A 1  0  0  -1  0  0
B 0 -1  0  1  -1  0
C 0  0  0  0  1  -1
D 0  1 -1  0  0  0
>>> model.sm.NullSpace()
  c_0 c_1

A_tx 0 -1
R_3 1 -1
D_tx 1 -1
R_1 0 -1
R_2 -1  0
C_tx -1  0
>>>
```

Figure 2.7: The ScrumPy command line interface.

Chapter 3

Construction of *P. tricornutum* genome scale models

3.1 Introduction

This chapter describes the construction and curation of genome scale metabolic models of *P. tricornutum* and their analysis. Two GSMs of *P. tricornutum* have been constructed in this work: i. the first model was developed from a previously published model by Hunt et al. [122] ii. the second model was constructed *de novo* from DiatomCyc, the PGDB of *P. tricornutum*. Models are analysed for their properties using the structural modelling techniques described in chapter 2. The general model properties for both the models will be discussed and compared with other published models of *P. tricornutum* [109, 110, 122]. The model developed from Hunt et al. [122] will further be used to understand the difference in the metabolic responses to phototrophic and mixotrophic conditions. The parameters used for the model analysis such as absorbed light intensity and cell maintenance cost were estimated based on the experimental values and are discussed in this chapter.

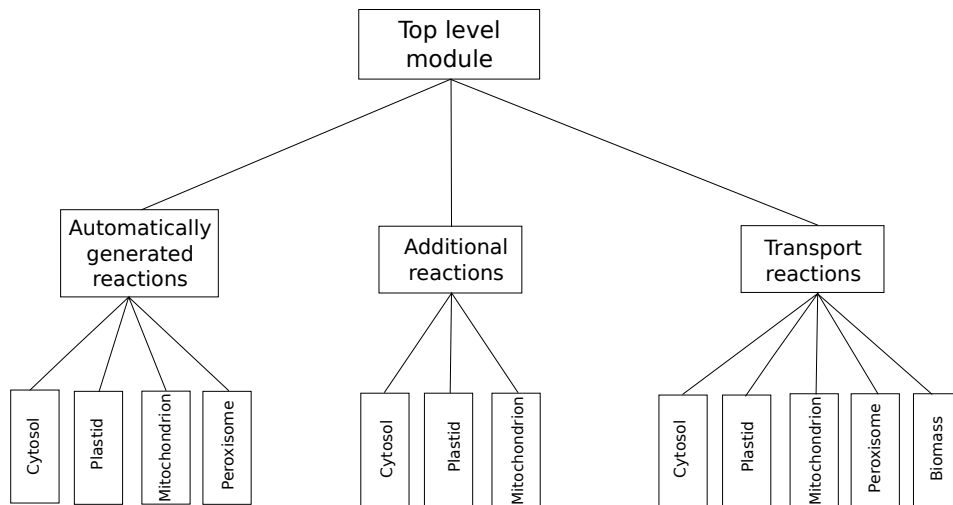


Figure 3.1: Schematic diagram representing the modular structure of the GSMs of *P. tricornutum*. The top level module describes the model type along with various other features of the model and imports the modules spread over multiple files which consist of compartment-specific reactions extracted either directly from the database or added manually, and compartment-specific transport reactions.

3.2 Construction and Analysis of GSMs

3.2.1 Model Structure

The GSM of *P. tricornutum* was structured in a modular fashion. Modules were divided based on compartment-specific reactions, their source and type, as shown in figure 3.1:

- A top level module:
 - It includes the ScrumPy directives, as described in section 2.4.2: `Structural()` to specify that it is a structural model, `ElType()` to specify the data type, `External()` to declare external metabolites which included water and proton and `Include()` to load and import other modules spread in multiple files.

- Automatically generated reactions:
 These modules consist of reactions that are directly extracted from the metabolic reconstruction or biochemical databases. Reactions for specific compartments are placed in separate modules and the suffixes `_Cyto`, `_Plas`, `_Mito` and `_Pero` are added to the reactions present in cytosol, plastid, mitochondrion and peroxisome respectively.
- Transport reactions:
 Transport reactions are differentiated from other reactions by using suffix `'_tx'`. Inter-compartmental transport reactions transfer internal metabolites between the cytosol and the specific compartments. Cytosolic transporters exchange the metabolites between cytosol and the environment. Explicit transporters are defined to account for individual biomass components, as described in Poolman et al. [108] and are assigned the suffix `'_bm.tx'`. Environmental and biomass components are assigned an external counterpart and distinguished from internal metabolites by using prefix `'x_'`. All transport reactions are defined with external species on the left side such that positive flux represents transport of metabolite into the system and negative flux represents transport of metabolites out of the system.
- Additional reactions:
 These modules consist of reactions that are either missing in the metabolic reconstruction or database and are added to the model manually. For instance, if a model is built from an organism-specific database and reactions are added from MetaCyc, then these reactions are included in additional reactions' modules. These modules also include lumped generic reactions, as described in section 3.2.3, that represent the biosynthesis of polymers, e.g fatty acid biosynthesis and carbohydrate biosynthesis, lumped light reactions and electron transport chain. Lumped reactions are represented with small letters to differentiate them from reactions obtained from databases.

3.2.2 Software and Tools

The model construction, curation and its analysis was done using the ScrumPy software package [132] (section 2.4.2). Reactions for the models were extracted, using PyoCyc module in the ScrumPy, from organism specific database, DiatomCyc, and/or MetaCyc as described in 2.4.2. LP based model analysis (as described in section 2.2.4) was done using the Gnu Linear Programming Kit (GLPK) integrated in ScrumPy, where the objective function was to minimise the absolute sum of total flux. The allowed solution was constrained by fixing the minimum flux to the biomass transport reactions, specifying the biomass precursor export. The LP was defined as stated in eq.3.2.1.

$$\begin{aligned} \text{minimise} & \quad : \sum_{i=1}^m |\mathbf{v}| \cdot w_i \\ \text{subject to} & \quad \left\{ \begin{array}{l} \mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\ v_{i..j} = t_{i..j} \\ v_{\text{ATPase}} = \text{ATPase} \end{array} \right. \end{aligned} \quad (\text{eq.3.2.1})$$

where \mathbf{v} is the vector of all reaction fluxes, \mathbf{N} is the stoichiometry matrix, with n rows (metabolites) and m columns (reactions). Each reaction is associated with a weight factor, w_i , which signifies the cost of reaction and was equal to 1 unless otherwise stated. The LP is constrained to obey the steady state assumption defined by $\mathbf{N}\mathbf{v} = \mathbf{0}$. $v_{i..j} = t_{i..j}$ defines fluxes in the biomass transporters. v_{ATPase} defines the energy demand for cell maintenance and polymerisation of biomass compounds. For this, a generic ATPase reaction was added to the model.

3.2.3 Model Construction and Curation

In this thesis, two GSMs of *P. tricornutum* have been constructed as follows:

Model I: GSM developed from a published model (Hunt et al.)

The set of compartmentalised reactions in the model previously described by Hunt et al. [122] was extracted from MetaCyc, using the PyoCyc module in

ScrumPy as described in section 2.4.2, and stored in modular fashion as described in section 3.2.1 to obtain a preliminary model of *P. tricornutum*. The initial model consisted of 318 reactions and was compartmentalised into Cytosol, Chloroplast, Mitochondrion and Peroxisome. The ‘lumped’ reactions of Hunt et al. [122] were expanded into individual reactions. The single lumped biomass equation was replaced with transporters for individual biomass components, to produce 48 biomass components which included all amino acids, nucleotides, lipid, carbohydrate and vitamins. Reactions that were present in Hunt et al. [122] but that could not be extracted automatically from PyoCyc, as they were represented by KEGG identifiers, were added to the additional modules of their respective compartments. Missing reactions in the network were identified based on literature study as well as analysing the model for its ability to account for the production of biomass, and were added to the additional reactions modules of the appropriate compartments.

Kroth et al. [46] identified genes encoding enzymes involved in photorespiration with the exception that glycerate kinase was not found in *P. tricornutum*. This reaction was not reported in the Hunt et al. [122] model though Fabris et al. [47] had predicted the enzyme to be present in *P. tricornutum*. Hence, this reaction was added to the mitochondrion compartment of the model. Similarly, reactions involved with the phosphoketolase pathway Fabris et al. [47] and synthesis of metabolites determined in metabolomics data (such as trehalose and mannitol) [138], along with a number of other reactions (listed in appendix A) were added to the model.

Model II: GSM developed from DiatomCyc

The DiatomCyc flatfile consisting of *P. tricornutum*-specific genes, proteins, enzymes, reactions and compounds information was obtained from Michele Fabris, author of DiatomCyc [47]. All the reactions present in DiatomCyc were then extracted, using the PyoCyc module in ScrumPy as described in section 2.4.2, and stored in modular fashion as described in section 3.2.1. Transporters and reaction compartmentalisation information were obtained from Hunt et al. [122]. Reactions with no compartmentalisation information

were assumed to be present in the cytosol. Reactions not present in DiatomCyc but present in Hunt et al. [122] were added to the additional reactions module. Individual biomass transporters to produce 48 biomass components were added to the model which included all amino acids, nucleotides, lipid, carbohydrate and vitamins.

Model Curation

In addition to the curation process described in section 2.3.1, the reactions in the constructed models were specifically curated, to meet physical as well as biological expectations, as follows:

Polymer Synthesis

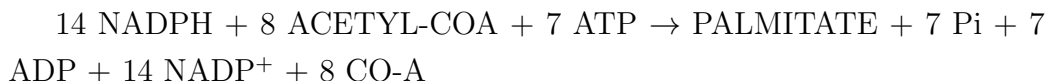
Polymer synthesis reactions involved with metabolites of unknown empirical formula, were replaced by considering polymers to be made of numbers of unit monomers such that the stoichiometric coefficient were balanced on both sides of the reaction. For instance, DiatomCyc reports a carbohydrate synthesis reaction with stoichiometry:



where 1-3-beta-D-Glucans has an empirical formula of $\text{C}_{15}\text{H}_{24}\text{O}_{19}\text{R}_2$ with an unknown R group. Such reactions are replaced by considering polymer to be made of one unit of monomer. For example, Chrysolaminaran, a beta-D-glucan polymer used as storage material in *P. tricornutum*, was considered to be made up of 1 unit of glucose:



Similarly, reactions involved with fatty acid synthesis were deleted from the model and replaced by a generic reaction that converts acetyl-CoA, NADPH and ATP into palmitate as described in Poolman et al. [101, 108]:



The synthesis of other biomass component involving lipid, e.g. TAG synthesis, were defined in terms of palmitate as:



Electron Transport Chain

Since protons and O_2 are declared external in the model, it was not appropriate to link these generic species to the ADP phosphorylation through mitochondrial ETC (shown in figure 3.2) and photophosphorylation (shown in figure 1.4). So, the mitochondrial ETC was represented by set of proton translocation reactions such that the matrix proton species was converted to an inter-membrane/cytosolic protons species using redox energy. These inter-membrane protons were used to drive the ATP synthesis reaction (4 inter-membrane protons were used to phosphorylate 1 ADP to ATP). The P/O ratios of oxidative phosphorylation (the ATP produced per oxygen atom) were 2.5 and 1.5 for NADH and succinate respectively which are consistent with well known literature values [139] for eukaryotes. While the light dependent ETC was represented as two lumped reactions [101], cyclic and non-cyclic photophosphorylation, as follows:

Light Cyclic Rxn:



Light NonCyclic Rxn:



Energy and Redox Conservation

Energy conservation was evaluated by constraining all the energy sources (light energy and organic carbon sources) to zero, setting arbitrary positive flux to the generic ATPase reaction and solving the LP, as described in eq.3.2.1. Existence of a feasible solution reflected the violation of energy conservation. In such a case, the reversibility and directionality of reactions in the solution was checked against thermodynamic data. A similar procedure was implemented to ensure redox conservation.

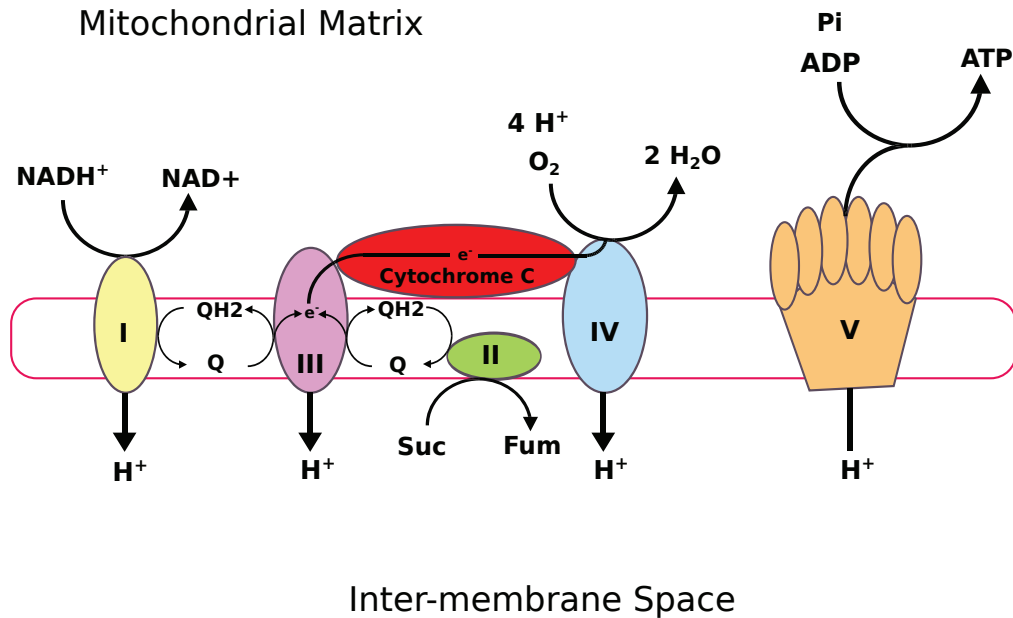


Figure 3.2: Mitochondrial electron transport chain. Electron carrier complexes, I-IV, present in the inner mitochondrial membrane, facilitate the transfer of electron from lower to higher redox potentials. This flow of electrons causes protons to be pumped from the matrix side to the inter-membrane space generating a proton-motive force that drives ATP synthesis through the ATP-synthesizing complex (complex V), embedded in the mitochondrial membrane. QH₂: ubiquinol, Q: ubiquinone, Fum: Fumarate, Suc: Succinate, Pi: Phosphate, I: Complex I, II: Complex II, III: Complex III, IV: Complex IV and V: Complex V

Biomass Feasibility

The LP as described in eq.3.2.1 was used to check for the feasibility of biosynthesis of each biomass component in turn. No feasible solution for a given biomass component reflected either: i. discontinuity in the network, and in such cases missing reactions for the particular biomass component synthesis were added to the model, or ii. inappropriate reversibility of the reactions involved with biomass synthesis. In the latter case, all the reactions in the model were made reversible, LP was resolved and all the reactions carrying negative flux that were defined as irreversible in the previous model were

redefined as reversible or set irreversible in the opposite direction. Energy and redox conservation properties were rechecked after alterations.

3.2.4 Estimation of Experimental Parameters

Light Absorbed per amount of Biomass

The light supply rate per amount of biomass was estimated based on eq.3.2.2 described by Kliphuis et al. [140].

$$r_{Ex} = \frac{I_{abs} \times A}{C_b \times V_c} \quad (\text{eq.3.2.2})$$

where, r_{Ex} is light supply rate per amount of biomass (mol Photon gDW⁻¹ time⁻¹), I_{abs} is the absorbed photon flux density (mol Photon m⁻² time⁻¹), A is the illuminated area (m²), V_c is the culture volume (L) and C_b is the biomass concentration (gDW L⁻¹). Note that the absorbed photon flux density is not the same as the incident photon flux density as portion of light can be reflected and/or transmitted during the experimental process. In this study, absorbed photon flux density for given incident photon flux, was estimated based on the experimental data provided by Accliphot partner CNRS, France (refer to appendix B.1 for the details).

Energy Cost of Cell Maintenance

The energy cost for the cell maintenance was assumed to be distributed between a non-growth specific maintenance cost, such as DNA repair, cell wall maintenance, and pH control and a growth specific maintenance cost, required for transport and polymerisation as eq.3.2.3:

$$\gamma_{ATP} = \gamma_{xATP} * \mu + m_{ATP} \quad (\text{eq.3.2.3})$$

where, γ_{ATP} is the total maintenance cost (mmol ATP gDW⁻¹ hr⁻¹), γ_{xATP} is the growth associated cellular maintenance cost (mmol ATP gDW⁻¹), m_{ATP} is the non-growth associated cellular maintenance cost (mmol ATP gDW⁻¹ hr⁻¹) and μ is the growth rate hr⁻¹.

Geider et al. [141] reported that the maintenance photon flux, i.e photon flux just enough to balance the maintenance cost and not support the growth of the culture, was in the range of 0.5-0.75 $\mu\text{mol Photon m}^{-2} \text{ s}^{-1}$ for *P. tricornutum*. For this range of photon flux density, the light supply rate per amount of biomass was estimated using eq.3.2.2 (approximately 1.85 mmol Photon $\text{gDW}^{-1} \text{ hr}^{-1}$). ATP generated by light reactions at this photon flux, in the absence of growth, was regarded as the non-growth specific maintenance ATP requirement, approximately 0.7 mmol ATP $\text{gDW}^{-1} \text{ hr}^{-1}$.

The growth associated cellular maintenance cost was assumed to be the same as estimated for *C. reinhardtii* [106], i.e 29.89 mmol ATP gDW^{-1} . With a growth rate of $\mu = 0.01 \text{ hr}^{-1}$ (experimental data provided in appendix, B.2), the total maintenance cost was calculated, using eq.3.2.3 to be approximately 1.0 mmol ATP $\text{gDW}^{-1} \text{ hr}^{-1}$.

3.2.5 Model Analysis

The stoichiometry matrix of the model was analysed for general properties such as the number of reactions, transporters and metabolites. PyoCyc module was used to establish the gene associations of reactions. Null-space analysis was used to get an estimate of the number of reactions in the network that were not able to carry flux at steady state. The FBA technique was used to assign flux ($\text{mmol gDW}^{-1} \text{ hr}^{-1}$) to reactions in phototrophic and mixotrophic conditions. For this aspect of the study, an LP formulation as described in eq.3.2.1 was used where flux in biomass transporters were set to the experimentally observed proportion (appendix C.1). The cell maintenance cost was considered to be 1.0 mmol ATP $\text{gDW}^{-1} \text{ hr}^{-1}$ as estimated in section 3.2.4. In phototrophic conditions, light was the only source of energy, and CO_2 and HCO_3^- as the source of inorganic carbon. Mixotrophic conditions had the same constraints as phototrophic except that, in addition, it was allowed to consume glycerol as a source of organic carbon.

3.3 Results

3.3.1 General Model Properties

The final model properties of models are summarised in table 3.1. Approximately, 17% of reactions in model I and 61% of reactions in model II were not capable of carrying flux at steady state i.e were identified as dead reactions. LP analysis showed that both models were able to utilise media components (NO_3^- , NH_4^+ , SO_4 , O_2 , Pi and inorganic and/or organic carbon) as input material to produce biomass in experimentally-observed proportions.

The numbers of reactions including transporters in the LP solution, in phototrophic conditions, for Model I and Model II were 326 and 344 respectively. The minimum photon flux for which a solution to eq.3.2.1 could be found, for both the models, was approximately 4 mmol Photon $\text{gDW}^{-1} \text{hr}^{-1}$. For the corresponding photon flux, the quantum demand was approximately 12 photons per carbon fixed, the photosynthetic quotient was 1.10 O_2 released per carbon fixed and quantum efficiency i.e O_2 released per photon was 0.08.

The number of reactions including transporters in the LP solution, in mixotrophic conditions, for Model I and Model II were 336 and 337 respectively. Both the models utilised approximately 0.1 mmol $\text{gDW}^{-1} \text{hr}^{-1}$ of glycerol in the presence of light. Flux in the photon transporter, i.e. the light requirement for growth, decreases by 50% as compared to phototrophic conditions.

3.3.2 Metabolic Responses

Predicted metabolic responses to phototrophic and mixotrophic conditions were compared based on Model I. The results show different sets of reactions (mainly associated with central carbon metabolism) active in these conditions.

Table 3.1: Summary of the general model properties of the two models constructed in this work. Note that the number of reactions associated with genes is an underestimate as manually added lumped reactions have not been counted and, since metabolites can occur in different compartments the number of metabolites listed here totals less than the sum of metabolites present in the model. *Reactions capable of carrying flux at steady state.

Model Properties	Model I	Model II
Total Reactions	449	1100
Transporters	142	142
Metabolites	355	1028
Reactions with gene association	421	1035
Dead Reactions	84	534
Live Reactions*	365	566

Phototrophic Metabolism

In phototrophic conditions, as shown in figure 3.3, light energy is utilised to fix inorganic carbon through the Calvin cycle. O_2 produced by the light reaction is released out of the system. The RuBP carboxylase reaction fixes CO_2 to meet the carbon demand for the biomass production. Reactions involved with the Calvin cycle are active both in plastid and cytosol however, the plastidial reactions (r1-r14) along with the light reactions carry higher flux ($> 0.1 \text{ mmol gDW}^{-1} \text{ hr}^{-1}$) and TP can be produced solely in chloroplast. (Note that this includes the plastidial TAL (r10) while SBPase (r28) and SBPaldol (r29) is absent in chloroplast). Though the cytosolic SBPase (r28) and SBPaldol (r29) are active, the flux in these reactions are 100-fold less than plastidial reactions associated with the Calvin cycle.

The TP, E4P, R5P, X5P and G6P produced from the Calvin cycle are associated with various pathways contributing towards biomass production. For instance, E4P is utilised for PRPP production leading to nucleotide and amino acid synthesis, R5P is utilised for shikimate production and X5P is used for acetyl-CoA production through the phosphoketolase pathway. The six carbon compounds (G6P and F6P) are utilised for carbohydrate synthesis. PGA can be converted to PEP/Pyruvate, through the lower glycolytic reactions (r15-r17, r42). Pyruvate enters mitochondrion to contribute towards the

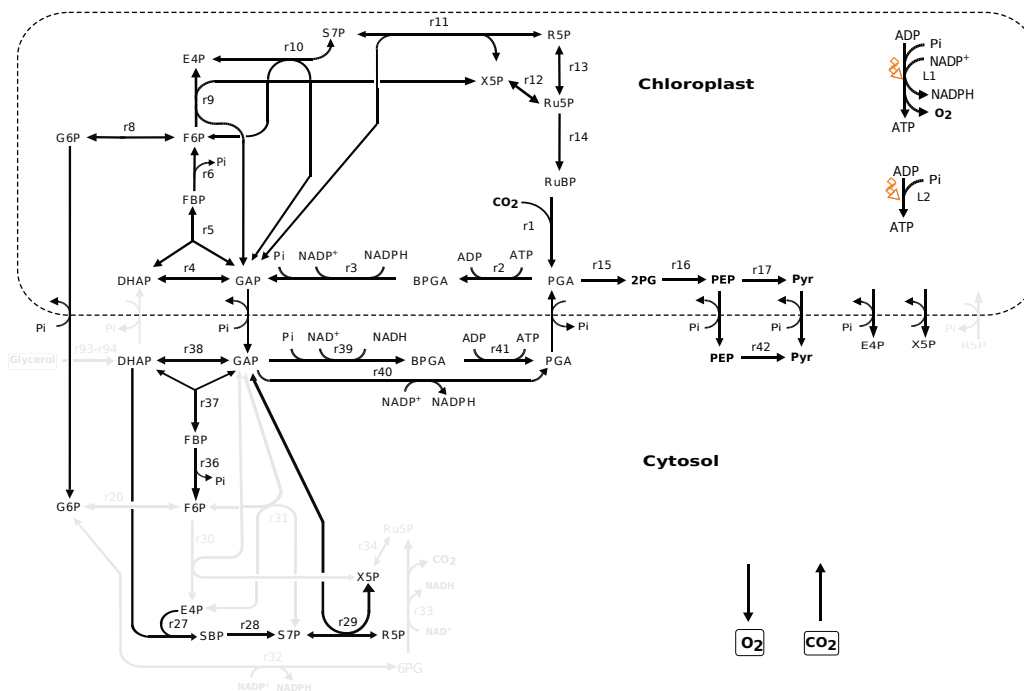


Figure 3.3: Reactions involved with the central carbon metabolism and predicted to carry flux in phototrophic condition are shown by dark lines. Greyed out reactions do not carry flux under this condition. As indicated by unidirectional arrows, CO₂ is imported into the system while O₂ is released out of the system. CO₂ is assimilated by Rubisco (r1) which, along with all other plastidial reactions involved with the Calvin cycle (r2-r14), carry higher flux as compared to other active reactions. The TP, E4P, R5P, X5P and G6P produced from the Calvin cycle is utilised for the production of biomass components (though not shown in figure).

functioning of the TCA cycle as discussed under ‘Mitochondrial Metabolism’.

As there is no direct NAD(P)H transporter from the plastid to the cytosol, reducing power is transferred via the GAP-PGA shuttle. The GAP-PGA shuttle involves reduction of PGA to GAP in the chloroplast through PGK and GAPdh, transport of GAP across the chloroplast membrane, and oxidation of GAP to PGA in the cytosol to yield NADH, ATP, and PGA. The latter metabolite is transported to the chloroplast to complete the cycle and resulting in net transfers of light-generated reducing equivalents. Apart from this shuttle, reductant is transferred across compartments through the GLT/GLN-Ornithine shuttle discussed under ‘Mitochondrial Metabolism’.

Mixotrophic Metabolism

In mixotrophic conditions, as shown in figure 3.4, though the LP formulation permits both uptake of both glycerol as well as inorganic carbon, the LP solution prefers glycerol as the sole carbon source, in the presence of light reactions. O_2 produced by the light reaction is not sufficient to oxidise glycerol and hence O_2 is imported into the system. There is no export of CO_2 as one would expect from catabolic metabolism, rather CO_2 produced is refixed by carboxylase reactions (PEPC and pyruvate carboxylase). Most of the plastidial reactions involved with the Calvin cycle are inactive. SBPase (r28) and SBPaldol (r29) are active and carry 10 fold higher flux as compared to phototrophic condition.

Glycerol is converted to glycerol-3P and then to DHAP by glycerol kinase (r93) and glycerol-3P dehydrogenase (r94) respectively. DHAP enters the gluconeogenic and/or glycolytic routes. As the glycolytic route in cytosol is not complete, cytosolic TPI (r38) converts DHAP to GAP, which then enters the plastid. Plastidial GAPdh (r3) and PGK (r2) carries flux in the direction of PGA production, opposite to that in phototrophic conditions, which is then utilised by the lower half of glycolysis (r15-r17, r42) to produce PEP/pyruvate. DHAP and GAP are also utilised for G6P production (gluconeogenesis) via cytosolic FBPaldol (r37), FBPase (r36) and G6Pisomer (r20). G6P is utilised for carbohydrate production and also by another glycolytic route, the ED pathway which is connected to the lower half of glycolysis in mitochondrion as discussed below.

Mitochondrial Metabolism

The mitochondrial responses, as shown in figure 3.5, for both phototrophic and mixotrophic conditions remain similar except that reactions involved with the ED pathway, lower glycolysis and serine synthesis is active in mixotrophic conditions but not in phototrophic conditions. In mixotrophic conditions, 6PG, produced by G6Pdh (r32 in figure 3.4) in the cytosol, is transported to the mitochondrion where it is metabolised by phosphogluconate dehydratase (r47) and KDPGaldo (r48) to pyruvate and GAP. The latter metabo-

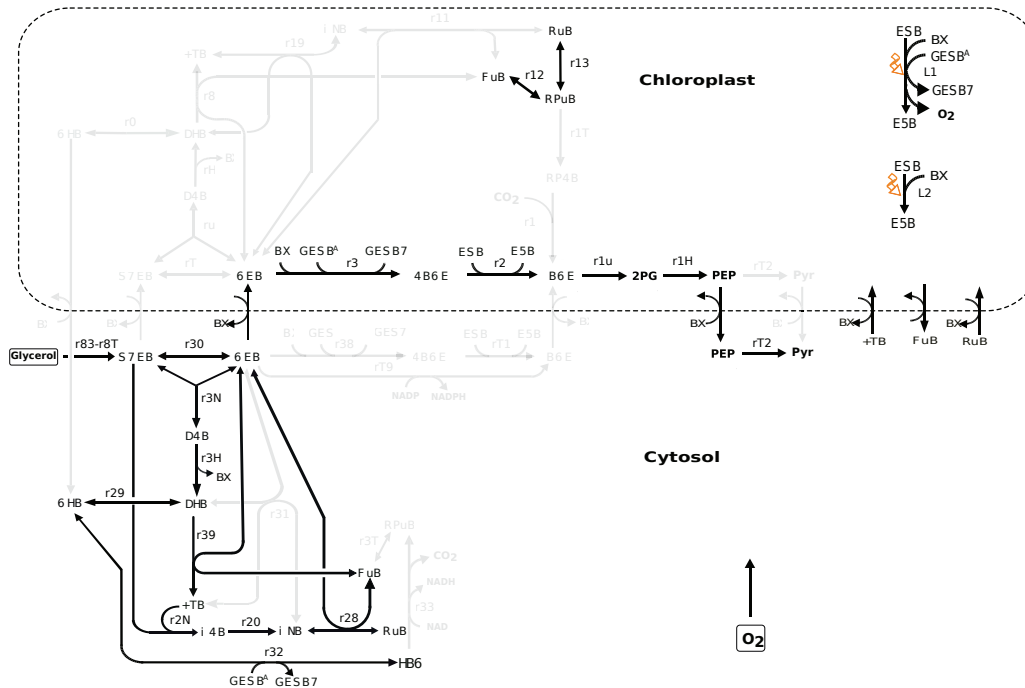


Figure 3.4: Reactions involved with the central carbon metabolism and predicted to carry flux in mixotrophic condition are shown by dark lines. Greyed out reactions do not carry flux under this condition. As indicated by unidirectional arrows, O_2 is imported into the system however, there is no exchange of CO_2 between the system and the environment. The only source of carbon utilised for biomass production is glycerol which is converted to DHAP in two consecutive reactions catalysed by glycerol kinase (r93) and glycerol-3P dehydrogenase (r94) respectively. DHAP enters the glycolytic/gluconeogenic routes. Most of the reactions involved with the Calvin cycle in the plastid are inactive. Plastidial GLTdh (r3) and PGK (r2) carry flux in the direction of PGA formation (glycolysis) contrary to phototrophic condition where these reactions carried flux in the direction of GAP formation (Calvin cycle). The TP, E4P, R5P, X5P and G6P produced from the Calvin cycle is utilised for the production of biomass components (though not shown in figure).

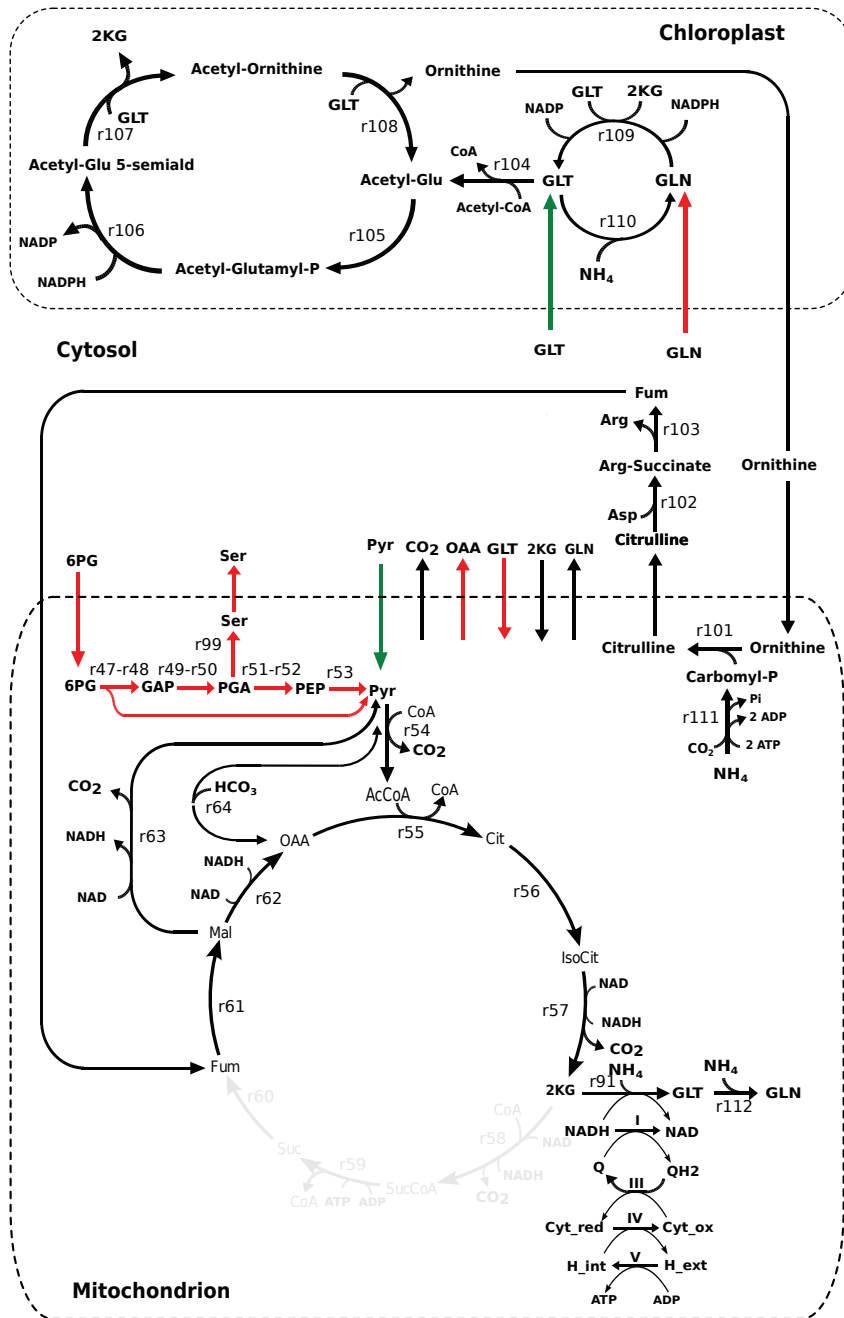


Figure 3.5: Predicted mitochondrial response in phototrophic and mixotrophic conditions. Red, green and black colours represent reactions that are active in mixotrophic, phototrophic and both conditions respectively. Note that the purpose of this figure is to show the mitochondrial response and reductant transfer from chloroplast to mitochondrion hence not all the reactions in cytosol and chloroplast involved with GLT, GLN and 2KG have been shown.

lite is oxidised to PGA, which is utilised for either serine production via a non-photorespiratory biosynthetic route (r99), or pyruvate, which enters the TCA cycle or is carboxylated to OAA.

The mitochondrial ETC (I-V) is active in both conditions however the flux is 10-fold higher in mixotrophic conditions as compared to phototrophic conditions. In addition to the GAP-PGA shuttle, reductant is transferred from the chloroplast to the cytosol via GLT/GLN-Ornithine shuttle and from the cytosol to the mitochondrion through a citrulline/fumarate shuttle, as shown in figure 3.5. Ornithine is produced in the plastid by addition of an acetyl group to GLT along with a series of reactions involving oxidation of NADPH and is transported to the cytosol where it can be used for proline synthesis (not shown in the figure) and further transported to mitochondrion. In mitochondrion, it is utilised for the synthesis of citrulline, which is an intermediate for OUC. The OUC utilises citrulline and aspartate, produced from the transamination of OAA, to produce arginine and fumarate. The latter metabolite enters the TCA cycle to produce 2KG and reductants for the ETC. In mixotrophic condition, OAA produced in mitochondrion, from the TCA cycle and carboxylation of Pyr, acts as the precursor for aspartate synthesis however, in phototrophic condition OAA is produced from carboxylation of PEP in cytosol. Reactions of the TCA cycle between 2KG and fumarate are inactive and 2KG is converted to GLT and GLN, via GLTdh and GLN synthase respectively, and transported to plastid to complete the cycle.

3.4 Discussion

3.4.1 Model Construction and Curation

The model built from *P. tricornutum* PGDB, Model II, required removal of 819 reactions and 172 metabolites from the final model. This was mainly due to undefined metabolites and unbalanced reactions involved with such metabolites. While many of these metabolites were linked to non-metabolic process such as Charged-LEU-tRNAs, Sulfur-Carrier-Proteins-ThiI, and could

be removed without affecting the metabolic networks, other metabolites representing generic forms such as Nucleoside-Triphosphates, Alcohols, Triacylglycerols had to be replaced with defined instances of the metabolite. On the other hand, Hunt et al. [122] model consisted only of metabolic reactions and hence the model developed from Hunt et al. [122] bypassed the need to remove reactions and metabolites linked to non-metabolic process. However, since the reactions for both the models were extracted from MetaCyc/DiatomCyc, they inherited the inaccuracies and inconsistencies of the database such as inconsistent metabolite names, incorrect reversibility and directionality of reactions. Curation of models were achievable through manual intervention as described in sections 2.3.1 and 3.2.3.

Polymer synthesis reactions, in both the models, were replaced by lumped reactions. For instance, chrysolaminaran, was considered to be made up of one unit of glucose and reactions for chrysolaminaran synthesis were replaced by lumped reaction, as shown in 3.2.3. Similarly, palmitic acid synthesis reactions were represented by single lumped reaction that utilised 8 units of acetyl-CoA. This reduces the size of the model but it can be argued that if the aim of the model analysis is to understand the routes for precursor synthesis, then this can still be fulfilled. On the other hand, if the purpose of the model analysis would have been to specifically look at the reactions involved with fatty acid synthesis of different chain lengths, as in Levering et al. [110], then the lumped reaction would have been replaced by individual reactions.

3.4.2 Comparison of GSMs

Table 3.2 summarises the size of GSMs of *P. tricornutum* constructed in this work along with other GSMs published till date. The GSM developed by Hunt et al. [122] was reconstructed using extensive knowledge of the finished genome sequence and literature-based manual analyses and it is small as compared to other published GSMs of *P. tricornutum*. As a result, the model developed from Hunt et al. [122], Model I, is also comparatively smaller than that developed from DiatomCyc, Model II. The size of latter

Table 3.2: Comparison of GSMs of *P. tricornutum*: Properties of GSM have been compared between 3 published models. All the models includes Cytosol, Mitochondrion, Peroxisome and Chloroplast. ^aChloroplast is further subdivided into stroma and thylakoid.

Model Properties	Hunt et al. [122]	Kim et al. [109]	Levering et al. [110]	Model I (this study)	Model II (this study)
Reactions	318	849	4456	591	1242
Metabolites	335	587	2172	355	1028
Genes	680	607	1027	421	1035
Software used	CellNet Analyzer	MOST	COBRA Toolbox	ScrumPy	ScrumPy
Compartments	4	4 ^a	4 ^a	4	4

model is comparable to GSMs of other organisms constructed using the same methodology, e.g. *Salmonella typhimurium* [135] had 1162 reactions based on 4.8 Mb genome, *A. thaliana* [101] has 1,406 reactions based on 157 Mb genome [142], rice *Oryza sativa* [108] has 1736 reactions based on 430 Mb genome [143]. This is also, to certain extent, comparable to the GSM of *P. tricornutum* published by Kim et al. [109] with 849 reactions.

On the other hand, the GSM of *P. tricornutum* published by Levering et al. [110] has an extraordinarily large number of reactions as compared to other published GSMs. This model is based on *P. tricornutum* genome annotation Phatr2 which was reconciled with the reference network obtained from the network of *C. reinhardtii* [144] and two genome-scale models for *Synechocystis* sp. PCC6803 [145, 146]. Contrary to our models, which include lumped reaction for lipid synthesis with an emphasis on the precursor production, this model incorporates metabolites and reactions to account for a range of fatty acid and lipid production with more than 60% of the reactions involved with lipid metabolism. The core model with reduced lipid metabolism, however, has 1029 reactions and is in same range as other GSMs. (Note that models in this work were constructed and most of the preliminary analysis was complete by early 2014. Hence, Hunt et al. [122], which was already made available by the authors, was taken as the reference for models construction while the other two models, Kim et al. [109] and Levering et al. [110], were published quite recently).

However, it is important to note that the capability of a metabolic network cannot be determined merely by the number of reactions because, despite of the presence of a reaction, it might not be able to carry flux at steady state. For instance, though the number of reactions in Model II is greater than in Model I, nearly half of the reactions in Model II (refer table 3.1), are dead and cannot carry flux at steady state. Moreover, if we consider the number of reactions utilised for the production of biomass components calculated with the same objective function, i.e minimisation of total flux, then irrespective of model size and even the organism, it appears to be in the same range for *P. tricornutum*, *A. thaliana* [101], *O. sativa* [108] and *S. typhimurium* [135].

3.4.3 General Model Properties

Model analysis predicts quantum demand, photosynthetic quotient and quantum efficiency for *P. tricornutum* to be 12 photon/C fixed, 1.10 O₂ released/C fixed and 0.08 O₂ released/photon respectively when NH₄⁺ is used as nitrogen source. These values increase to 14 photon/C fixed, 1.4 O₂ released/C fixed and 0.10 O₂ released/photon when NO₃⁻ is used as nitrogen source. Nevertheless, these values are close to the experimental observation where quantum demand, photosynthetic quotient and quantum efficiency for *P. tricornutum* have been reported to be in the range of 8.8-14.0 photon/C fixed [147], 1.2-1.6 O₂ released/C fixed [148, 149] and 0.12 [141] respectively.

In mixotrophic conditions, the model consumed ≈ 0.1 mmol gDW⁻¹ hr⁻¹ of glycerol. In the experimental study, significant effect of glycerol on growth was observed after 120 hours of culture [138] and the total glycerol consumed during this period was estimated to be approximately 0.1 mmol gDW⁻¹ hr⁻¹ (refer to appendix B.3 for experimental detail), consistent to model finding. Though the structural analysis results are merely used for qualitative analysis, here the quantitative similarity between the model and the experimental observation can be regarded as an assurance for the biological significance of the model results.

3.4.4 Estimation of Maintenance Cost

Our estimation of the non-growth specific maintenance cost based on maintenance photon flux is approximately $0.7 \text{ mmol ATP gDW}^{-1} \text{ hr}^{-1}$. Non-growth associated maintenance costs have been reported to range from 0.36 to 7.6 $\text{mmol gDW}^{-1} \text{ hr}^{-1}$ for *Lactobacillus plantarum* and *E. coli*, respectively. For *C. reinhardtii* this was assumed to be $1.5 \text{ mmol gDW}^{-1} \text{ hr}^{-1}$ [106]. Hence, our estimation seems to be within the experimentally observed range, and close to the lower end.

As the growth-associated maintenance cost has not been studied independently and is not well known in *P. tricornutum*, we assumed it to be the same as *C. reinhardtii* ($29.89 \text{ mmol ATP gDW}^{-1}$) which was estimated by fitting a metabolic model to the experimentally determined biomass yield by changing the ATP requirement [106].

In our study, total cell maintenance cost was estimated to be approximately $1.0 \text{ mmol ATP gDW}^{-1} \text{ hr}^{-1}$ which does not come as a surprise knowing that cell maintenance cost for *P. tricornutum* can be very low [141]. This is also in the range used by Levering et al. [110] for their model analysis. On the contrary, though Kim et al. [109] assumed the growth and non-growth associated maintenance cost to be same as *C. reinhardtii* for their model analysis, they estimated the total maintenance cost in the range of 30-38 $\text{mmol gDW}^{-1} \text{ hr}^{-1}$. This overestimation appears to be the result of either not taking into account the growth rate for the growth-associated maintenance cost or simply assuming the growth rate of *P. tricornutum* to be 1.0 hr^{-1} , which is not realistic given that a typical growth rate for *P. tricornutum* in phototrophic condition ranges from 0.01 to 0.1 hr^{-1} [150–152]. Though, the overestimation of maintenance cost might not change the predicted flux pattern and relative flux distribution, it is likely to have a quantitative impact on the absolute reaction fluxes.

3.4.5 Minimum Photon Flux Requirement

Model analysis predicts that the minimum photon flux required for biomass production in experimentally observed proportion (as in appendix C.1) to

be 4 mmol Photon gDW⁻¹ hr⁻¹. Geider et al. [141] reported growth of *P. tricornutum* at a photon flux intensity as low as 1.3 $\mu\text{mol Photon m}^{-2} \text{s}^{-1}$, which is only marginally higher than that of maintenance photon flux. This value, when used to get an estimate of the minimum photon flux required for growth using eq.3.2.2, was approximately 4.3 mmol Photon gDW⁻¹ hr⁻¹, in the same range as the minimum photon flux required for biomass production by the models. Though this value would depend on various factors such as biomass composition, concentration and growth rate, that the experimental estimation and model analysis result in the same range of minimum flux requirement is quite promising and can be regarded again as an assurance for the biological relevance of the model.

3.4.6 Metabolic Responses

Phototrophic Metabolism

The model analysis suggest that, under light conditions TAL and FBPase along with other Calvin cycle enzymes present in plastids are active and capable of assimilating CO₂ and producing the TP, even in the absence of SBPase. Based on the structural modelling approach, it has already been reported that the operation of the Calvin cycle requires at least two active reactions out of SBPase, FBPase and TAL [97].

In higher plants, Calvin cycle enzymes along with OPPP are present in the plastid and are under redox regulation. G6Pdh, 6PGdh, TAL are upregulated in the dark while GAPdh, FBPase, SBPase are upregulated in light. In diatoms, SBPase is present in the cytosol and is not under redox control, while FBPase and TAL are present in both plastid and cytosol and only the plastidial FBPase is under redox control. The possible reason that these reactions are not redox regulated is because the OPPP in diatoms is in the cytosol, which might prevent futile cycling with the Calvin cycle even in the absence of redox regulation.

In addition, the model result suggests that since SBPase is localised in cytosol, TAL not being under redox regulation allows the functioning of the Calvin cycle solely based on reactions present in the plastid. Though the

significance of different localisation and regulation of enzymes involved in the Calvin Cycle in *P. tricornutum* has not been studied experimentally, apart from the observation that the abundance of TAL increases in under nitrogen replete phototrophic conditions Yang et al. [153]. Further experimental study to establish the significance of TAL in diatom metabolism is something worth exploring.

Mixotrophic Metabolism

Model analysis in mixotrophic conditions suggests that glycerol enters metabolism via DHAP, which can enter glycolysis and/or contribute towards gluconeogenesis. Further, it suggests that both glycolytic routes, EMP and ED pathway, are active and pyruvate produced through these routes can enter TCA cycle or be carboxylated to OAA, by pyruvate carboxylase.

ED pathway is known to produce less energy per molecule of glucose as compared to the EMP pathway however, the gene encoding the key enzyme of this pathway, KDPGaldo, has been reported to be highly upregulated at the beginning of the light period suggesting their role in the early light period, when pyruvate production through the EMP pathway is low [71].

Contrary to our observation, Kim et al. [109] reported no flux ED pathway. Further, they suggest that in the absence of a TP transporter in mitochondrion, both the ED pathway and the lower half of glycolysis would be dead. However, our model analysis suggests that the ED pathway along with the lower half of the glycolysis are active in mixotrophic as well as low light conditions (see section 4.3.1) even in the absence of any mitochondrial TP transporter. The only transporter to this route in our model is for 6PG, consistent to Hunt et al. [122] model, that connects the first two reactions of the ED pathway, which is also shared by the OPPP, to the ED pathway in mitochondrion. The end product of the ED pathway enters the lower half of the glycolysis, followed to the TCA cycle and the network is complete even in the absence of TP transporters. Our result is also consistent with the isotopomer analysis [53] study which showed flux through the ED pathway under mixotrophic conditions.

Mitochondrial Metabolism

In higher plants, the well known mechanism of reductant transfer across compartments includes the MAL/OAA shuttle and/or the TP/PGA shuttle [154]. The MAL/OAA shuttle is carried out by MALdh, an enzyme that reversibly catalyses the reduction of $\text{NAD}^+/\text{NADP}^+$ to NADH/NADPH by the oxidation of malate to OAA. In the cytosol, NADH is oxidised by OAA to NAD^+ and malate. The latter can enter the plastid and/or mitochondrion in exchange for OAA that is produced by oxidation of MAL by NADPH dependent MALdh in the plastid and NADH dependent MALdh in mitochondrion [155]. In *P. tricornutum*, MALdh has been reported to be present only in mitochondrion which limits the possibility of an operational MAL/OAA shuttle, leaving behind an unanswered question: ‘how is reductant transferred across compartments?’.

Our FBA analysis suggests that the reductant is transferred out of the chloroplast through TP/PGA which is quite well reported both experimentally [156] as well as from the model analysis [108]. Kim et al. [109] also reports an active TP/PGA shuttle from their *P. tricornutum* model analysis however only in mixotrophic but not in phototrophic conditions.

In addition to TP/PGA shuttle, our model analysis suggest for reductant transfer through a network that interconnects the ornithine biosynthesis pathway in the plastid to the OUC and TCA cycles. This route involves the exchange of ornithine with GLT/GLN, in the plastid, which is consistent with the route suggested by Levering et al. [110]. However, contrary to Levering et al. [110], whose model analysis suggested that ornithine transported to mitochondrion is converted to GLT and then to GLN, by 1-pyrroline-5-carboxylate dehydrogenase and GS respectively, without any correlation to OUC, our analysis shows that ornithine metabolism has greater complexity and is interconnected to OUC and TCA cycle. Ornithine produced is utilised by the OUC cycle whose intermediate, fumarate, is fed to the TCA cycle to produce 2KG. Reactions between 2KG and fumarate do not carry flux which is consistent with model analysis of plants [101] and *P. tricornutum* [109], and labeling analysis in plants [157] under light and dark conditions.

These observations highlight the importance of the OUC in diatoms and is in agreement with the proposed connection between the urea cycle, glutamine synthetase/glutamate synthase (GS/GOGAT) and TCA cycle by Prihoda et al. [158].

3.5 Conclusion

In this study, we constructed GSMs of *P. tricornutum* based on a previously published model and PGDBs and, scrupulously curated reaction stoichiometries to ensure conservation of mass and energy. The models developed, in our work, is not only able to produce biomass in experimentally observed proportion but also to reflect the experimentally observed behavior in both phototrophic and mixotrophic conditions. Without use of any kinetic constraints, the minimum photon used for the biomass production under phototrophic condition and glycerol uptake under mixotrophic condition is close to the experimental ranges and thus supports, at least to a certain extent, the semi-quantitative significance of the model analysis. Moreover, the study emphasises the previously reported observation that, regardless of network size, the number of reactions utilised for biomass production is comparable in different organisms. Model analysis suggests that though *P. tricornutum* has different regulation and localisation of the Calvin cycle enzymes, inorganic carbon can be fixed solely in the chloroplast.

In addition, it sheds light on the possible novel route for reductant transfer across compartments through GLT/GLN-ornithine shuttle that is interconnected to TCA cycle and OUC in mitochondrion. It also suggests possible conditions such as mixotrophy, light limitations (see section 4.3.1) when the ED pathway and lower half of glycolysis in mitochondrion would be active. Thus, the results not only highlights the difference in the metabolic responses between phototrophic and mixotrophic conditions but also the significance of presence of the ED pathway, lower half of glycolysis and OUC in *P. tricornutum* which are unusual for algae.

However, a FBA solution provides just one optimal solution, whereas there might exist multiple optimal and sub-optimal solutions. Pertaining

to this, in the next chapter, the model will be analysed subject to varying constraints in order to examine the metabolic re-adjustment with varying environmental/physiological conditions and to better understand the possible alternative metabolic responses.

Chapter 4

Reaction Response to Variation Analysis

4.1 Introduction

Prominent marine microalgae, such as diatoms, can cope with highly variable light conditions. They display physiological, biochemical, and behavioural responses to optimise their photosynthesis and growth under changing light conditions. Further, they show different physiological changes to phototrophic and mixotrophic conditions. One of the industrial partner of AccliPhot, Fermentalg, studied *P. tricornutum* grown in phototrophic and mixotrophic conditions on glycerol and reported higher growth rate and lipid accumulation under mixotrophic conditions [138]. However, the metabolic responses mediating the adjustment towards variation in light conditions, glycerol metabolism and lipid production are still largely unknown. Hence, to understand the metabolic changes during such conditions and further investigate the potential routes likely to be involved with these processes, the GSM of *P. tricornutum* was analysed using an FBA technique where the solution space is scanned repeatedly over an increasing constraints, as suggested by Poolman et al. [101]. The method determines the reactions that respond in a coordinated fashion to the changing constraint and can therefore be used to identify smaller sub-networks in a GSM.

4.2 Method

The LP formulation, shown in equation eq.4.2.1, was solved repeatedly with incrementing values for a specific constraints: i. light intensity, ii. lipid demand, and iii. glycerol uptake. As compared to equation eq.3.2.1, there are two additional constraints: cyclic photophosphorylation could not exceed noncyclic, and an arbitrary limit was set on the sum of the rubisco reactions, as described in Poolman et al. [108]. The latter constraint imposes an upper limit on the rubisco reaction and can be regarded as a proxy for saturation of the Calvin cycle. Note that the limit set on rubisco, in this study, is arbitrary and is not claimed to be quantitatively close to its *in vivo* value. Rather, it is used to understand how metabolism would re-adjust given that the enzymes in the Calvin cycle saturate at some point. In addition, the biomass composition, in our study, is allowed to vary, contrary to previous studies where fluxes in biomass transporters are fixed [101, 108]. This novel approach of relaxing the upper bound on the biomass constraint allows the model to change the biomass composition based on the conditions and adds to the metabolic flexibility. It also brings the model closer to a realistic scenario where organisms tend to vary their biomass composition based on nutrient availability and environmental factors.

$$\begin{array}{ll}
 \text{minimise} & : |\mathbf{v}| \\
 \text{subject to} & \left\{ \begin{array}{l}
 \mathbf{N}\mathbf{v} = \mathbf{0} \\
 \mathbf{v}_\nu = \nu \\
 \mathbf{v}_{i..j} \geq t_{i..j} \\
 \mathbf{v}_{ATPase} = ATPase \\
 \mathbf{v}_{LightNonCyc} \geq \mathbf{v}_{LightCyc} \\
 \mathbf{v}_{Carboxylase} + \mathbf{v}_{Oxygenase} \leq \mathbf{C}
 \end{array} \right. \quad (\text{eq.4.2.1})
 \end{array}$$

where the notations and constraints are the same as in equation eq.3.2.1, with the addition of $\mathbf{v}_\nu = \nu$ which defines the photon flux into the system and $\mathbf{v}_{i..j} \geq t_{i..j}$ which defines fluxes in the biomass transporters with no upper constraint. $\mathbf{v}_{LightNonCyc} \geq \mathbf{v}_{LightCyc}$ ensures that flux in cyclic

photophosphorylation cannot exceed flux in noncyclic photophosphorylation. $v_{\text{Carboxylase}} + v_{\text{Oxygenase}} \leq C$ sets a limit on the sum of the carboxylase and oxygenase reaction.

4.2.1 Photon Flux Variation Analysis

This analysis was performed to understand the potential metabolic adjustment to cope with increasing light intensity. For this aspect of the study, the lower bounds of the biomass transporters in eq.4.2.1 were constrained with the same experimentally observed biomass proportions as used in section 3.2.5, but with no limit on the upper bound. An arbitrarily value of 4.0 mmol gDW⁻¹ hr⁻¹ was used to set an upper limit to the sum of the rubisco carboxylase and oxygenase reactions and the photon flux was incremented gradually until the change in carbon flux through the system saturated and the response of the reactions remained unchanged.

4.2.2 Lipid Demand Variation Analysis

For this aspect of the study, eq.4.2.1 was solved repeatedly with incremented demand for lipid production, v_{Lipid} , instead of increasing photon flux into the system, until no feasible solution was possible.

The biomass composition (appendix C.2) for cultures grown in phototrophic conditions at 60 $\mu\text{mol Photon m}^{-2} \text{s}^{-1}$ was obtained from Fermentalg and hence, to understand the potential metabolic response to changing lipid demand under this experimental condition, the lower bounds of biomass transporters were constrained according to the experimental observations and the light intensity used in the experiment was used to constrain the upper bound of the photon flux (i.e 12 mmol Photon gDW⁻¹ hr⁻¹, refer to appendix B.3 for calculation). There was no limit on the upper bounds for the biomass transporters. The upper limit for the rubisco reaction, for this analysis, was 0.8 mmol gDW⁻¹ hr⁻¹. This was estimated by solving equation eq.4.2.1 with a fixed photon flux of 12 mmol Photon gDW⁻¹ hr⁻¹ and the flux obtained in rubisco was regarded as the maximum limit for this analysis.

4.2.3 Glycerol Uptake Variation Analysis

This analysis was performed in order to understand glycerol metabolism and the effects of increasing glycerol uptake on biomass composition under experimental conditions, where the axenic cultures of *P. tricornutum* grown in mixotrophic condition on glycerol was exposed to a light intensity of $60 \mu\text{mol Photon m}^{-2} \text{s}^{-1}$. The flux in the glycerol transporter was varied from 0.01 to 1.0 $\text{mmol gDW}^{-1} \text{hr}^{-1}$ and the allowed photon flux, v_ν , into the system was constrained to be less than or equal to the light intensity used in the experiment ($10 \text{ mmol Photon gDW}^{-1} \text{hr}^{-1}$, appendix B.3).

The experimentally observed biomass composition in mixotrophic condition (appendix C.2) under a similar light intensity was used to constrain the lower bounds of the biomass transporters, while the upper bounds were set arbitrarily 5 times higher. Thus, allowing the biomass components to increase up to 5 fold. In addition, the weight of reactions involved with metabolites that were components of biomass but were below the threshold level of determination (appendix C.2), were increased by 100 fold, on the assumption that reactions involved with such metabolites will carry minimal flux and since the objective function is being minimised, the increased weight imposes a severe penalty on including the reactions in the solution, whilst allowing them to appear in solution if essential to achieve a steady state.

The mixotrophic analysis in section 3.3.2 revealed that flux in rubisco reduces to zero however, some experimental studies suggest simultaneous assimilation of organic and inorganic carbon [138, 159]. Hence, to take this into consideration an arbitrary lower bound of $0.2 \text{ mmol gDW}^{-1} \text{hr}^{-1}$ was set on the rubisco reaction. This constraint was imposed to understand the combined effects of photosynthesis and glycerol assimilation.

4.3 Results

4.3.1 Reaction Responses to Photon Flux Variation

The number of reactions in the solutions to this condition varied between 324 and 338. Of these, 118 reactions showed more than 1% variation over the range of photon flux and these fall, approximately, into three recognisable areas of metabolism: nitrogen metabolism, energy dissipation mechanisms (reactions involved with energy consumption with no net carbon fixation) and carbon metabolism.

Nitrogen Assimilation

The model was allowed to take up both NH_4^+ and NO_3^- as nitrogen sources. Figure 4.1 shows that as the photon flux increases there is a transition between NH_4^+ and NO_3^- . At lower light intensities, NH_4^+ is used as the N source for biomass formation, but with increasing light intensity, NO_3^- is preferred over NH_4^+ . NO_3^- is reduced to nitrite by nitrate reductase which is further reduced to NH_4^+ by nitrite reductase in the chloroplast. NH_4^+ then enters the biomass components through GLTdh and/or GLN synthase. Though, the biomass composition is allowed to vary, the flux in total nitrogen uptake is constant with no net increase in nitrogenous biomass components. Rather, the increase in photon flux leads to the shift in the preference of nitrogen source.

Energy Dissipation Mechanisms

An excess of energy going into the system can be compensated for energy dissipation mechanisms, which involve sets of reactions that consume ATP and/or NADPH/NADH, to avoid over-reduction of reductant pools. Figure 4.2 shows flux in reactions involved with energy dissipation mechanisms with increasing photon flux, and their network diagrams.

In the flux range labelled D and beyond in the figure 4.2, adenylate kinase (AK) along with pyruvate-orthophosphate dikinase (PPDK) and pyruvate ki-

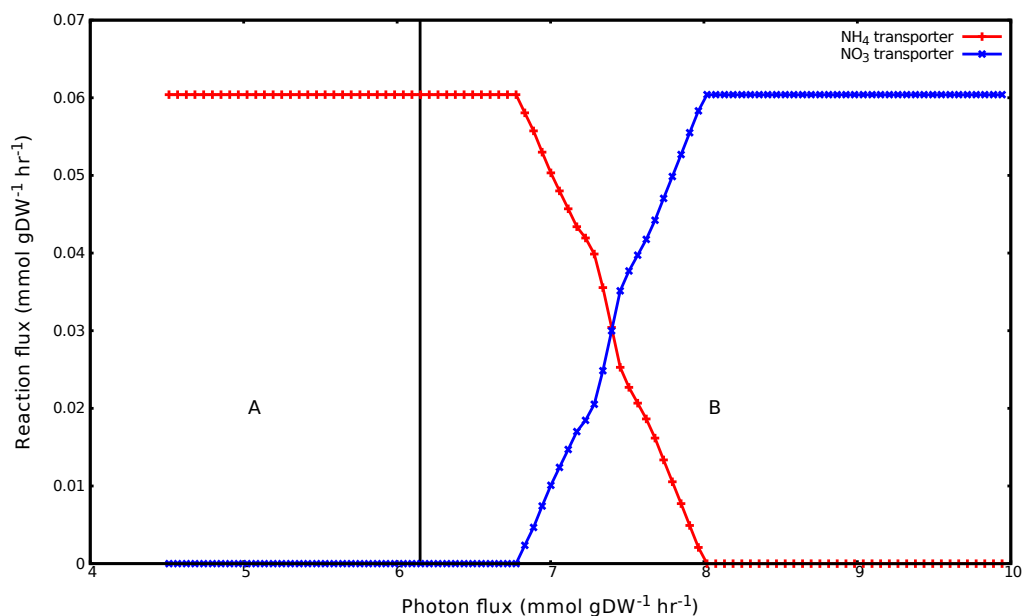


Figure 4.1: Variation in nitrogen source utilisation with varying photon flux. The X-axis represents the photon flux into the system and the Y-axis represents the flux carried by the reactions. At low light intensity (region A) NH_4^+ is utilised as nitrogen source and at increased light intensity (region B and beyond, refer figure 4.3.1 and figure 4.3.1 for light intensities in these regions) NO_3^- is preferred as nitrogen source.

nase (PK) are involved in energy dissipation, utilising three ATP. In region E, flux in the GOGAT cycle increases, which interconverts GLT and GLN at the expense of 1 molecule of each ATP and NADPH. The second GLT molecule produced by the GS-GOGAT cycle is reduced to 2KG and NH_4^+ , by GLTdh, and is again fed to GS-GOGAT to complete the cycle. In region F, the xanthophyll cycle turns on. It comprises interconversion between antheraxanthin and zeaxanthin/violaxanthin in conjunction with the ascorbate-glutathione cycle, oxidising two NADPH molecules. Though AK and PPDK reactions are active in region C, PK is inactive and the cycle is not complete to act as energy dissipation mechanism; rather, these reactions are associated with lipid production, discussed in the next section.

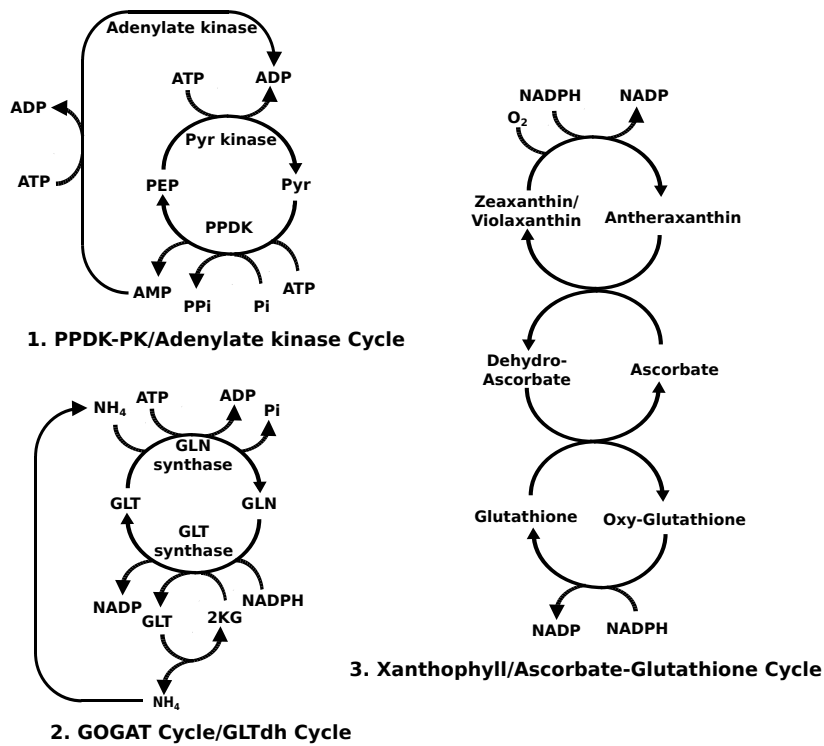
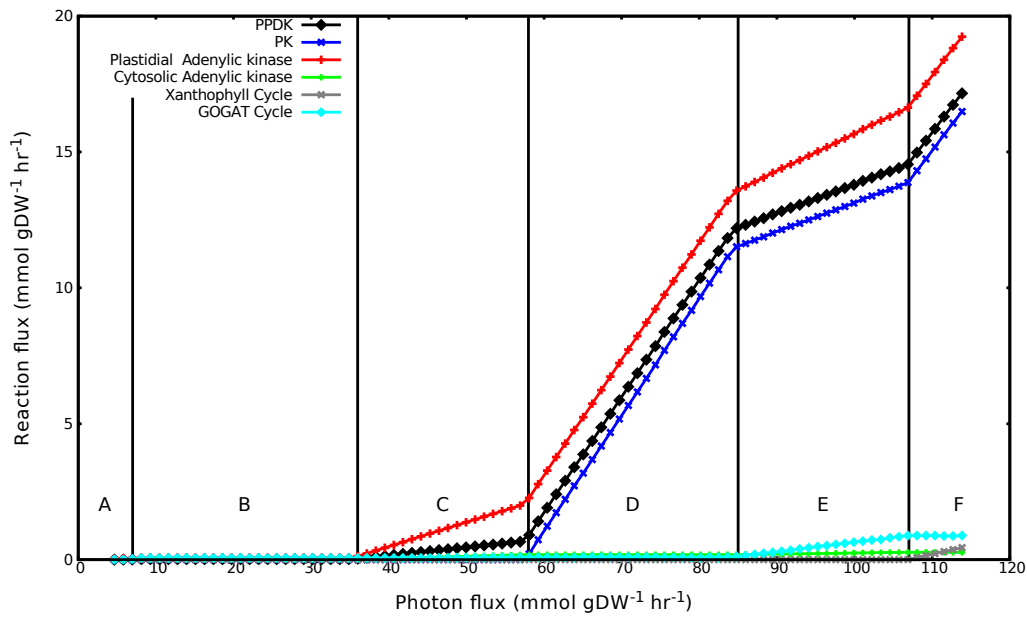


Figure 4.2: Top: Flux in reactions involved with energy dissipation mechanisms with increasing photon flux. The plot has been divided into six major regions labeled A to E based on the flux patterns. Bottom: Network diagram of the energy dissipation mechanisms shown in the top figure. PPK: pyruvate-orthophosphate dikinase, PK: pyruvate kinase, GOGAT: glutamine oxoglutarate aminotransferase.

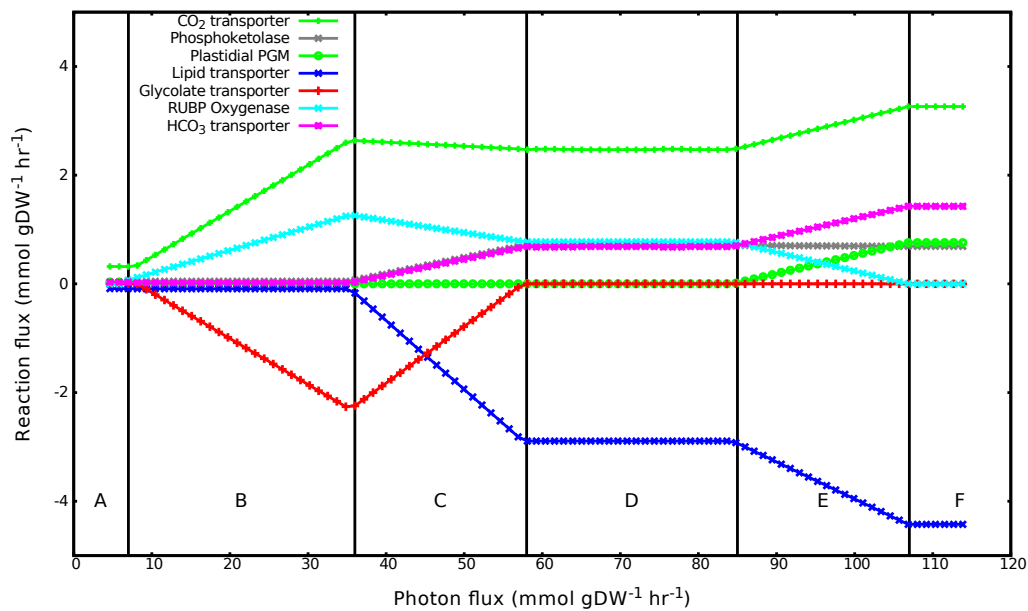


Figure 4.3: Reaction responses to changing photon flux. Positive values in transport reactions represent import to the compartment and negative values represent export. Flux in transport reactions is represented in terms of numbers of C atoms being imported or exported out of the system. The plot has been divided into six major regions labeled A to E based on the flux patterns.

Carbon Metabolism

Figure 4.3 shows the change in some of the reaction fluxes associated with carbon metabolism, in response to changes in photon flux. Other reactions, involved with carbon metabolism, that varied over the range of photon flux followed one or other of the flux patterns shown in figure 4.3.

A: Represents the minimum photon flux region. There is no increase in biomass components (i.e biomass transporters carry base-line flux) and/or excretion of by-products. Figure 4.4 is a magnified view of region A and the network diagram comprised of reactions with changing flux in this region. As shown in figure 4.4, most flux is accounted for the transfer of reductant across the compartments and generation of ATP via the mitochondrial ETC. GAP and PGA are exchanged between chloroplast and cytosol, resulting in net transfer of reductant and ATP

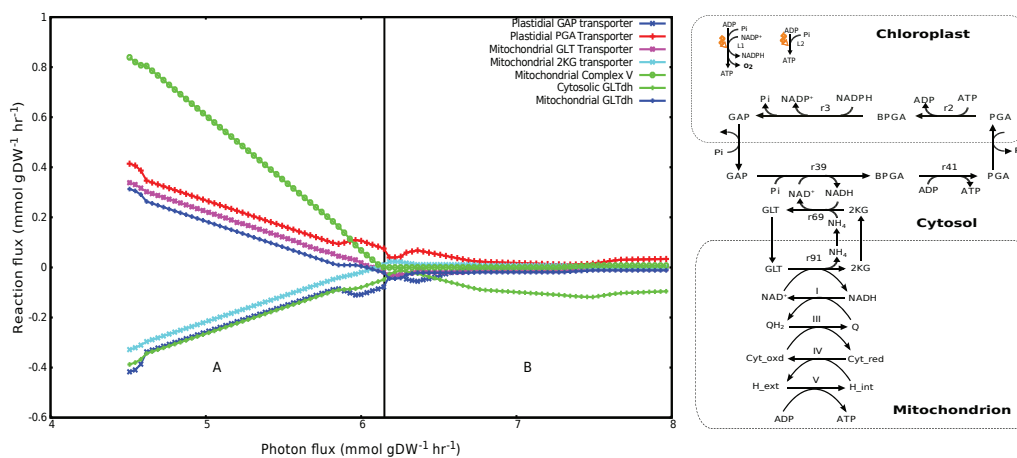


Figure 4.4: Chloroplast and mitochondrial reaction responses at low photon flux. Right: Reactions exhibiting change in flux at low photon flux. Left: The network composed of reactions shown in right panel. Reductant generated in chloroplast through light reactions is transferred across chloroplast via GAP-PGA shuttle which is further transferred to mitochondrion via GLT-2KG shuttle where it is used by the mitochondrial ETC to generate ATP. GAP: glyceraldehyde 3-phosphate, PGA: 3-phospho-D-glycerate, GLT: glutamate, 2KG: 2-ketoglutarate, BPGA: 1,3-diphosphateglycerate, GLTdh: GLT dehydrogenase

from chloroplast to the cytosol, as also seen in section 3.3.2. Cytosolic GLTdh oxidises NADH to convert 2KG to GLT which is transported to the mitochondrion. Mitochondrial GLTdh catalyses the conversion of GLT to 2KG, reducing NAD^+ to NADH, which is utilised by the mitochondrial ETC in order to generate ATP and 2KG is transported to the cytosol to complete the shuttle.

Apart from reductant transfer, reactions involved with the ED pathway, and serine biosynthesis through a non-photorespiratory route are active at low light, as also seen in section 3.3.2. Beyond this region, flux in these reactions turns off and serine is produced through the photorespiratory route.

B: Carbon is imported in the form of CO_2 ; however, there is no increase in any biomass components. Extra influxed carbon is exported out of the system as glycolate, which is mirrored by a corresponding increase

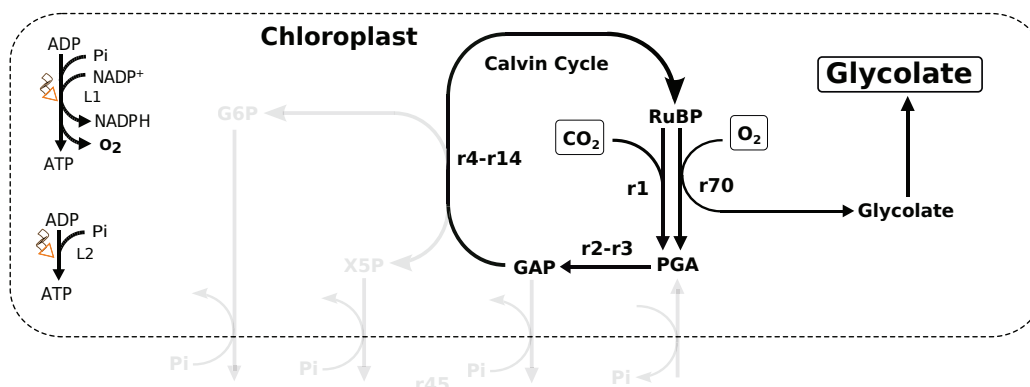


Figure 4.5: The network diagram for reactions carrying varying flux in region B of figure 4.3. Photorespiration turns on and glycolate produced through photorespiration is excreted out of the system. Simultaneously, CO_2 is assimilated through the Calvin cycle and PGA/GAP produced during the process is utilised for biomass production (eliminated from figure for clarity).

in the rubisco oxygenase reaction. The network diagram comprised of reactions that carry varying flux in this region is shown in figure 4.5.

C: Here, glycolate export decreases and although the rubisco oxygenase reaction also decreases, this decrease does not mirror that of glycolate export: at the C-D transition, there is no glycolate export although rubisco oxygenase flux is maintained, indicating the glycolate is being recycled. Also in this region, export of lipid increases, as does the import of HCO_3^- and flux in the phosphoketolase pathway. This region represents the transition between region B, where glycolate produced is excreted out of the system, and D, where the lipid export increases as described in the next paragraph.

D: In this region, a constant level of lipid export is accounted for increased flux in the phosphoketolase pathway, recycling of glycolate and the import of HCO_3^- . Though RuBP oxygenase is active, glycolate is not exported out of the system; rather it is completely recycled as shown in figure 4.6. During the process of recycling, glycolate is oxidised by GDH to glyoxylate which is further metabolised to glycine by glycine aminotransferase. Glycine undergoes two fates: glycine decarboxylase

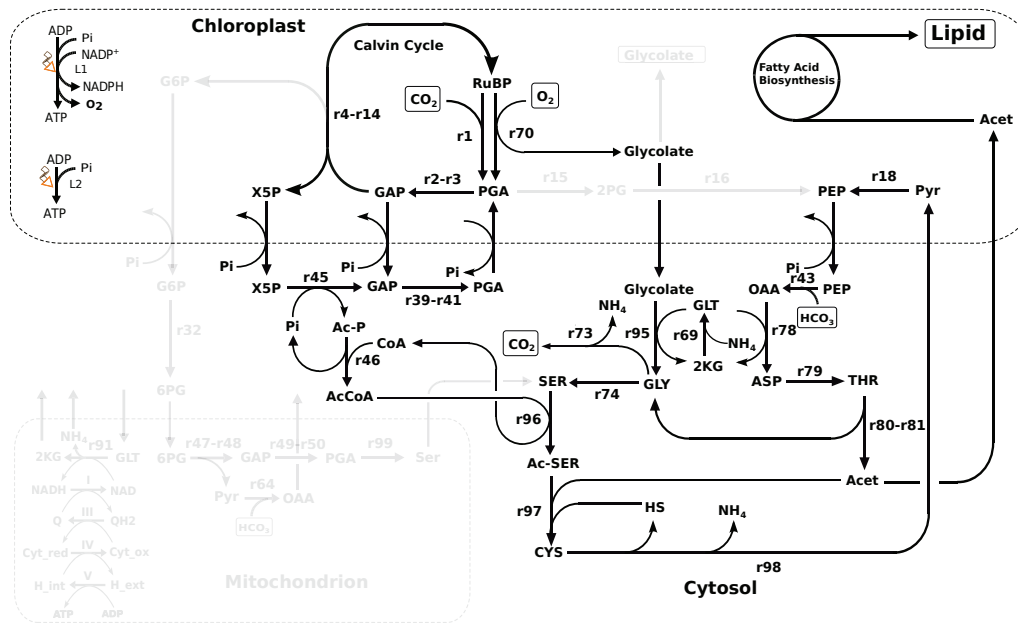


Figure 4.6: The network diagram for reactions contributing towards lipid production in region D of figure 4.3. Greyed out reactions do not carry flux in this region. Glycolate produced, during the process of photorespiration, is recycled to PEP which is carboxylated and utilised through threonine metabolism to produce glycine and acetate. Acetate produced during this process and through the phosphoketolase pathway contributes towards lipid production.

complex (r73 in figure 4.6) catalyses the oxidative cleavage of glycine to CO_2 , NH_4^+ , and a methylene group and/or glycine combines with methylene group to release serine through the action of serine hydroxymethyltransferase (r74). Serine is converted to acetyl-serine, then cysteine with concomitant release of acetate which is used for lipid production. The carbon skeleton of acetate derives from acetyl-CoA, produced through the phosphoketolase pathway (r45-r46). Cysteine undergoes hydrolytic deamination to produce pyruvate which is metabolised by PPK (r18) at the expense of ATP to PEP. HCO_3^- is utilised by PEPC (r43) for the production of OAA, which is in turn utilised for threonine synthesis (r78-r79), followed by its degradation to produce glycine and acetate (r80-r81), which further contributes towards lipid production.

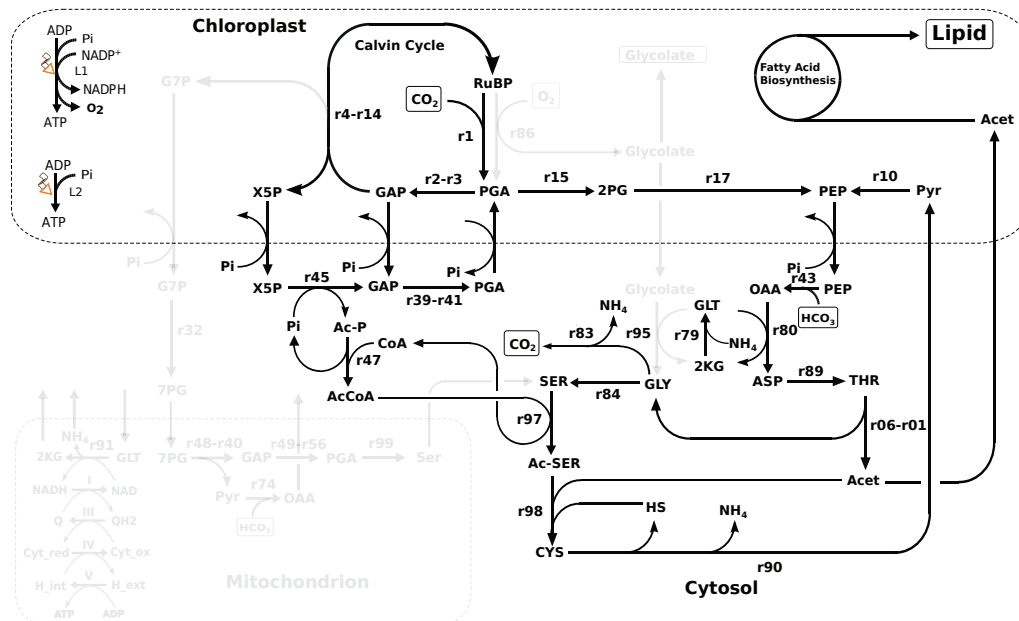


Figure 4.7: The network diagram for reactions contributing towards lipid production in region F of figure 4.3. Greyed out reactions do not carry flux in this region. Precursor for lipid synthesis is produced through the phosphoketolase pathway and threonine metabolism. PEP is solely produced through the glycolytic route contrary to region D (figure 4.6) where it is produced through glycolate recycling.

- E:** Here is an increase in lipid export, but the contribution from glycolate decreases and the required carbon is taken up in the form of HCO_3^- and CO_2 . The increase in lipid production is supported by flux in plastidial PGM. This region represents the transition between region D and F.
- F:** RuBP oxygenase turns off, hence glycolate is not produced at all. Lipid export is fully accounted by the import of HCO_3^- supported by plastidial PGM. As shown in the network diagram, figure 4.7, precursor for lipid production, acetate, is produced through the phosphoketolase pathway and threonine metabolism as seen in regions C, D and E. However, in this region, PEP required for the carboxylation of HCO_3^- is solely produced through the lower half of glycolysis present in the chloroplast, contrary to region C, D and E where it was produced through recycling of the glycolate.

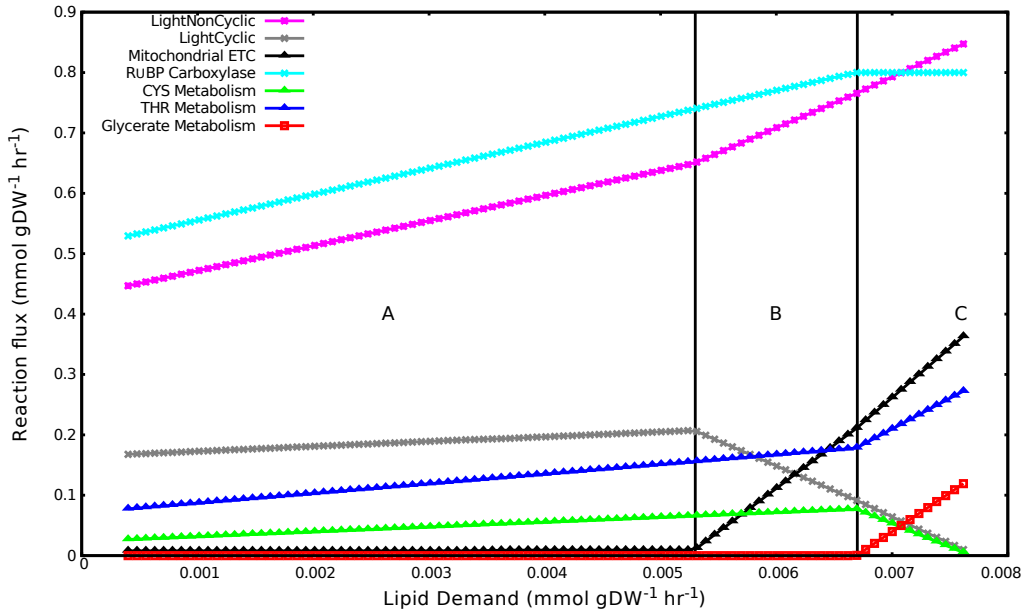


Figure 4.8: Reaction responses to increase in lipid demand. The plot has been divided into three major regions labeled A to C based on the flux patterns. Region A: reaction response when flux in photon transporter and rubisco reaction is not saturated, Region B: reaction response when flux in photon transporter saturates due to upper constraints on it, Region C: reaction response when flux in both photon transporter and rubisco saturates due to upper constraints on them. ETC: electron transport chain, CYS: cysteine, THR: threonine, RuBP: ribulose-1,5-bisphosphate.

4.3.2 Reaction Responses to Lipid Demand Variation

The number of reactions in the solutions for this condition varied between 332 and 352, and of these, 96 reactions showed variation over the range of imposed lipid demand. These reactions could be distinctively categorised to the following areas of metabolism: Calvin cycle, lipid synthesis, threonine metabolism, glycerate metabolism, cysteine metabolism, phosphoketolase pathway and reductant transfer. Figure 4.8 shows the reaction responses to increase in lipid demand. Other reactions varying over the lipid demand followed one or other pattern in figure 4.8.

A: Flux in the light reactions and rubisco carboxylase increase with increasing lipid demand. ATP and reductant required for the increasing lipid

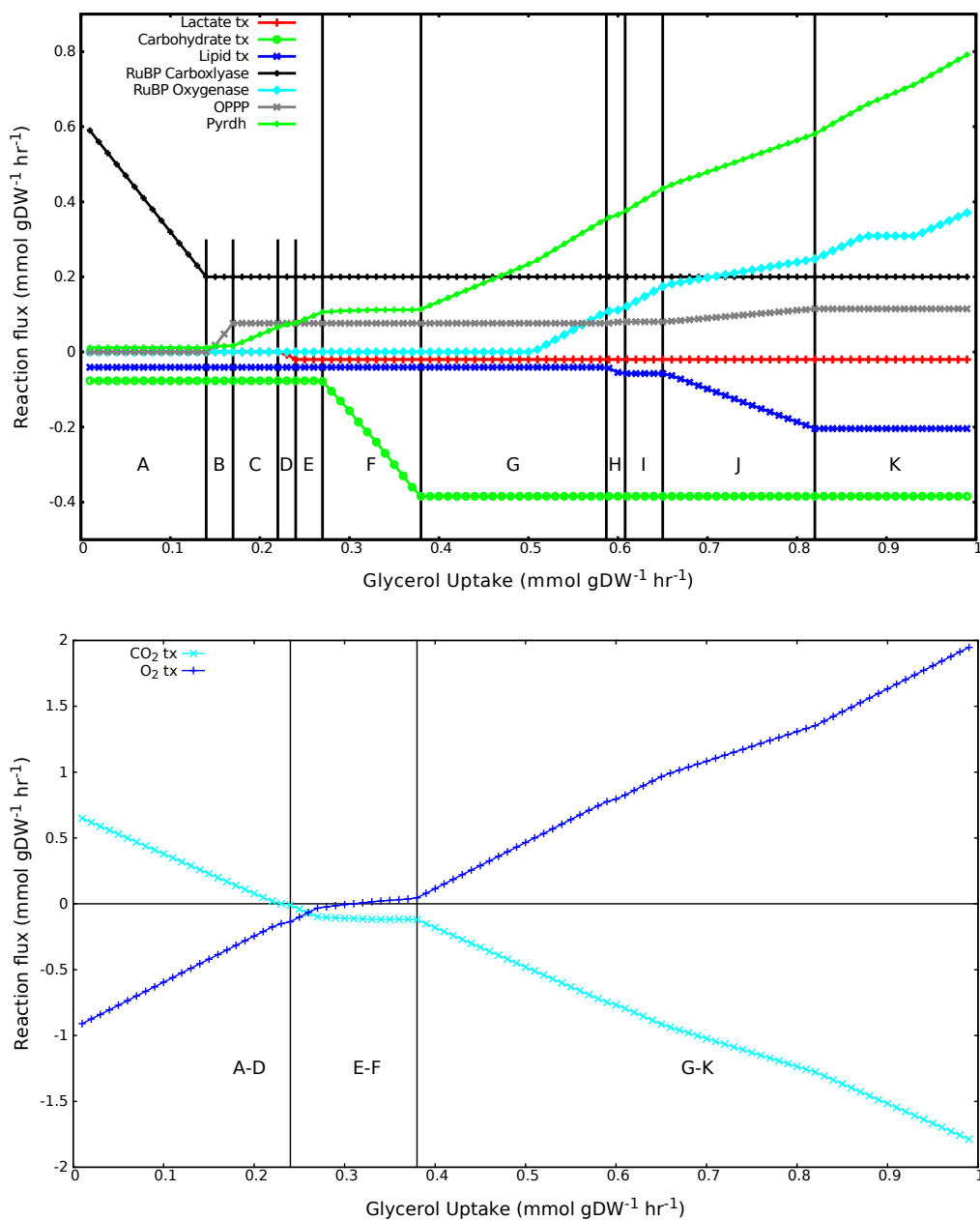


Figure 4.12: Reaction responses to changing glycerol flux. Positive values in transport reactions represent import to the compartment and negative values represent export. Flux in transport reactions (tx) is represented in terms of number of C atoms. The plot has been divided into seven major regions labeled A to K based on the flux patterns. RuBP: ribulose-1,5-bisphosphate, OPPP: oxidative pentose phosphate pathway, Pyr dh: pyruvate dehydrogenase.

A-D: There is no increase in biomass components above the base-line level. CO_2 is imported to the system while O_2 is exported out of the system. At low glycerol uptake, i.e region A, a large portion of C in biomass is contributed by inorganic carbon fixation by RuBP carboxylase; however flux in RuBP carboxylase and plastidial reactions associated with the Calvin cycle (r1-r14) decreases with increase in glycerol uptake.

As shown in figure 4.13, glycerol is converted to glycerol-3P by glycerol kinase (r93) followed by glycerol-3P dehydrogenase (r94) to DHAP. TPI (r38) catalyses the interconversion between DHAP and GAP which can then enter gluconeogenesis (r20, r36, r37) to produce G6P and glycolysis (r2, r3, r15, r16, r42) to produce pyruvate. In region B, flux in the OPPP increases as gluconeogenesis produces G6P, which serves as precursor for the OPPP. In region C, flux in glycolysis increases producing pyruvate that then enters the TCA cycle through PyrDh. In region D, pyruvate is fermented to lactate, by LDH (r92), and is excreted out of the system. CO_2 produced in these regions is refixed within the system.

F: Excess of carbon from glycerol uptake, i.e carbon remaining after the base-line production of biomass and lactate excretion, is diverted towards carbohydrate production. DHAP and GAP enters gluconeogenesis through cytosolic FBPaldol (r37), FBPase (r36) and G6Pisomer (r20), as shown in figure 4.13, and G6P produced through this process is utilised for carbohydrate synthesis.

H and J: Flux in photorespiration and lipid production increases with increasing glycerol uptake. Glycolate produced by RuBP oxygenase is recycled through the glyoxylate cycle, as shown in figure 4.14, which also shares reactions with the TCA cycle. Precursor for lipid synthesis is produced in the process of threonine degradation. Glycine produced during this process is further metabolised to PGA through glycerate metabolism which enters glycolysis to produce pyruvate, as also seen in section 4.3.2, that then enters the TCA cycle. Also the variant of the TCA cycle is active, where instead of 2KG being metabolised to

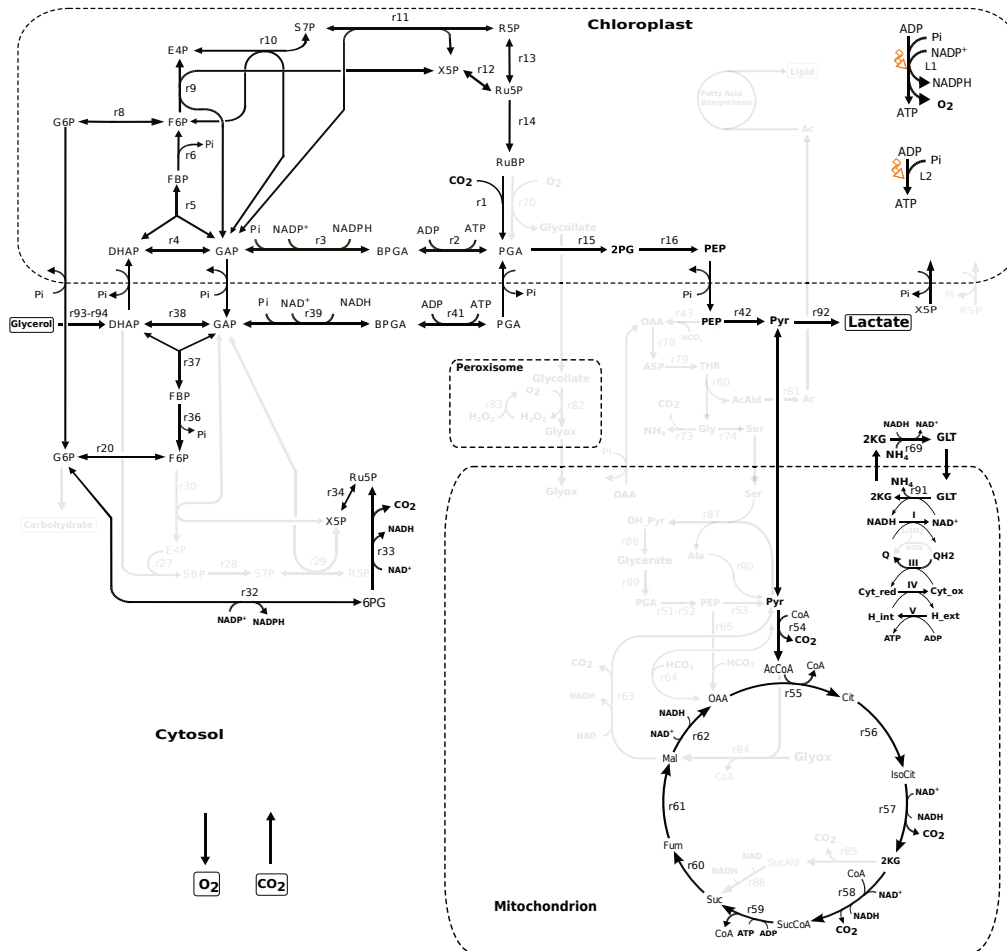


Figure 4.13: The network composed of reactions exhibiting change in flux in response to increasing glycerol uptake in regions A-D of figure 4.12. Glycerol and CO_2 is assimilated to support biomass production. Flux in plastidial Calvin cycle decreases with increase in glycerol uptake. Complete TCA cycle along with mitochondrial ETC is active. In region D and beyond, lactate is produced and excreted out of the system. Greyed out reactions do not carry flux in this region.

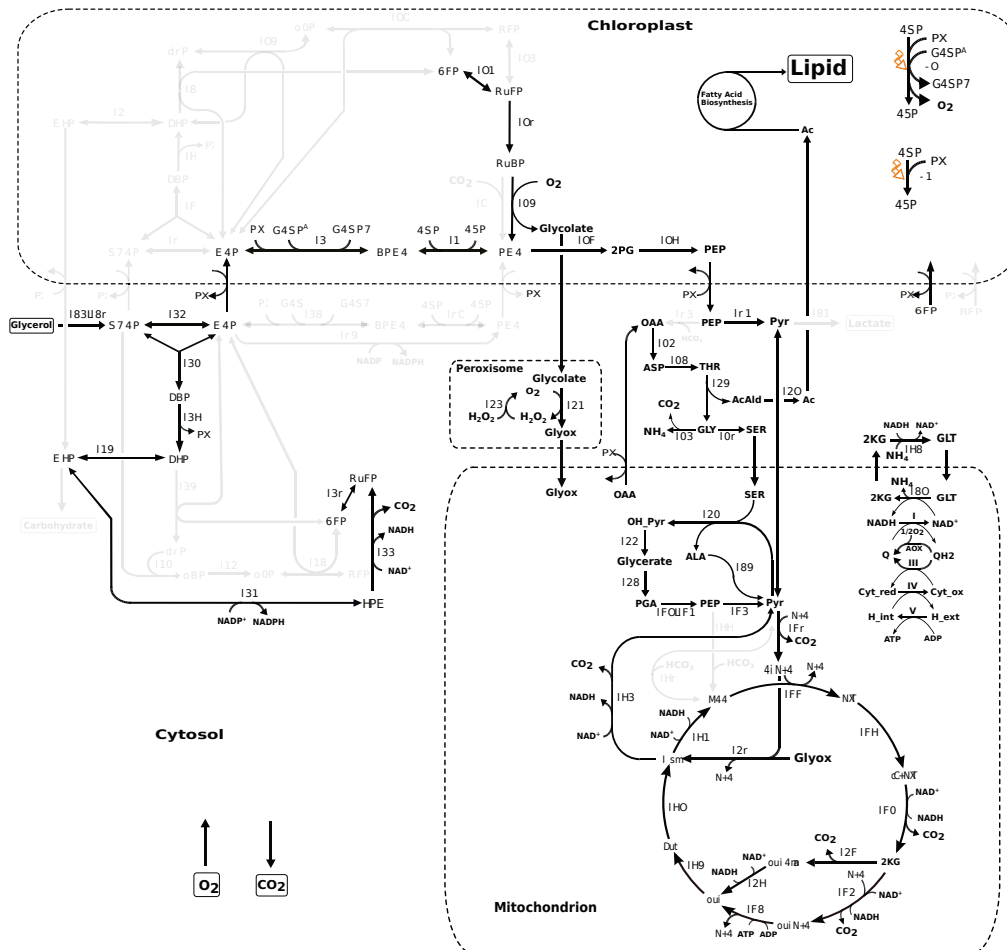


Figure 4.14: The network composed of reactions exhibiting change in flux in response to increasing glycerol uptake in region H and J of figure 4.12. Glycerol is metabolised through glycolytic route to produce pyruvate which then enters TCA cycle. Photorespiration turns on and glyoxylate produced during the process is metabolised through glyoxylate cycle. O₂ is imported into the system and is also produced by light reaction to oxidise glycerol while CO₂ produced by TCA cycle, OPPP and ME is released out of the system. Greyed out reactions do not carry flux in this region.

succinate via SucCoA, it is metabolised through SucAld. The reductant produced during the process is utilised by the mitochondrial ETC to generate ATP required for maintenance and biomass production. Malic enzyme (ME) and alternative oxidase (AOX) are also active at high glycerol uptake. CO₂ produced by TCA cycle, OPPP and ME is released out of the system.

E, G, I and K: In these regions, there is no change in lipid and carbohydrate production. Excess carbon and energy is adjusted by increased flux in ascorbate, tocopherol and chlorophyll production. However, flux changes in these reactions are less than 1% and have not been studied in detail. Apart from that, excess carbon is excreted as CO₂, and flux in photorespiration and other energy dissipation mechanisms, as discussed in section 4.3.1, adjusts the energy.

4.4 Discussion

4.4.1 Reductant Transfer

Model result suggests that reductant is transferred from the chloroplast to the cytosol via the PGA-GAP shuttle, as also seen in section 3.3.2, and reductant is transferred from the cytosol to the mitochondrion via the GLT-2KG shuttle. While the role of the PGA-GAP shuttle that we see in our model analysis is well known both experimentally [156] and through model analysis [108, 109], here we propose a novel route for reductant transfer across mitochondrion through the GLT-2KG shuttle, where GLT and 2KG are exchanged between cytosol and mitochondrion and reversible GLTdh present in both the compartments facilitates the interconversion. In the cytosol, GLTdh drives GLT synthesis while in mitochondrion it catalyses the deamination of GLT.

In vivo, the functionality of this shuttle would be determined by the directionality of GLTdh present in different compartments. Krebs et al. [160] suggested GLTdh to favour direction of GLT breakdown producing

NH_4^+ which can be utilised by mitochondrial carbamoyl-P for urea synthesis in mammalian liver. In higher plants, it is believed to favour the synthesis of GLT. However, it has a low affinity for NH_4^+ , with a K_m value higher than 5 mM, and as such a high concentration of NH_4^+ is unlikely in plants, another route of GLT synthesis via the GS/GOGAT cycle is thought to be dominant [161]. However, this consensus was questioned with the finding that GLTdh activity increases both in the cytosol and mitochondrion with increased NH_4^+ concentration in phloem companion cells [162].

GLTdh in *Gracilaria sordida*, a red alga, has affinity for NH_4^+ with a K_m 3.0-3.6 mM. The deamination process is favoured by NAD^+ dependent GLTdh while the amination mechanism is favoured by NADPH [163]. If this enzyme is present in more than one cellular compartment with different redox potentials or with separate coenzyme dependencies then it might independently regulate the deamination of GLT and amination of 2KG [161, 164]. Under such conditions, the possibility that it is involved in redox transfer cannot be ruled out.

Model analysis suggests that mitochondrial GLTdh is active in the direction of deamination. The fact that the OUC is present in *P. tricornutum* and that deamination of GLT would provide NH_4^+ for carbamoyl-P synthesis in mitochondrion along with the possibility of higher NAD^+ concentration compared to NADH (as the latter is constantly drained into the ETC), makes it likely that such a phenomenon could also be observed in a biological scenario. On the other hand, the higher concentration of NH_4^+ in the cytosol along with a higher concentration of NADH, that is transferred from chloroplast to the cytosol, could drive GLTdh in the direction of GLT synthesis. This could possibly transfer redox across the compartments. However, to make any firm conclusion on the role of GLTdh in redox transfer across compartments, it is essential to establish the redox potential and NH_4^+ concentration in different compartments of *P. tricornutum*.

4.4.2 Nitrogen Assimilation

The model prefers assimilation of NH_4^+ at low light however the preference shifts to NO_3^- with increasing light intensity, which provides a sink for reductant generated by the excess photons. This result is consistent with model observation made by Poolman et al. [101, 108] under similar conditions. In the latter study, the fluxes in biomass transporters were fixed. On the contrary, in our study the fluxes in biomass transporters are allowed to vary and model analysis still suggests only an increase in storage compounds (lipid and carbohydrate) and no increase in nitrogenous compounds, which is consistent with the biological findings where the organism prefers to increase lipid and carbohydrate as compared to proteins.

In plants the interaction between photosynthesis and nitrogen assimilation has been well studied. Nitrogen assimilation shows fluctuations in accordance with the daily light intensity and is stimulated under high light condition [165]. The nitrate reductase has been reported to be present in excess to accept excess electrons under high light conditions [166]. Furthermore, the plastidic enzymes for nitrogen assimilation interact with the thioredoxin system, which regulates the activation of photosynthesis enzymes under light and dark conditions [166]. Hence, the shift in preference of nitrogen source, in the model results, seems to be biologically relevant when both sources of nitrogen are available.

4.4.3 Energy Dissipation Mechanisms

Potential mechanisms of energy dissipation revealed in the model analysis involve the GS-GOGAT cycle, AK along with the PPDK/PK cycle and the xanthophyll cycle.

In our model analysis, nitrogen assimilation is constant irrespective of nitrogen source and light intensity, as discussed in section 4.4.2. Rather, flux in the GOGAT cycle increases and GLT produced by the GS-GOGAT cycle is reduced back to 2KG and NH_4^+ by GLTdh to complete the cycle. This GS-GOGAT/GLTdh cycle oxidises NADPH and utilises ATP without net increase in nitrogen fixation and hence acts as an energy dissipation

mechanism. From experimental studies, GS-GOGAT is thought to be the main route of assimilation of ammonium produced by nitrate reduction [167, 168] and is known to be light-regulated. Further, the gene expression has been reported to increase at higher light intensities [169, 170] suggesting its role in energy dissipation and increased nitrogen assimilation.

Another potential mechanism of energy dissipation, in our model analysis, involves AK coupled to the PPDK/PK cycle. PPDK catalyses the ATP and Pi dependent formation of PEP and PPi from pyruvate. PK catalyses ATP dependent conversion of PEP to pyruvate and thus, PPDK and PK forms a futile cycle dissipating excess energy. This overall cycle consumes only ATP and hence further suggests that there would exist other dissipation routes to accommodate excess reductant. Experimental studies suggest the role of PPDK in energy dissipation in *P. tricornutum* [171] while AK is known to catalyse the nucleotide phosphoryl exchange reaction and plays an important role in maintaining cellular energy through regulation of nucleotide ratios [172]. However, further experimental data, such as protein expression of these enzymes at high light levels, would bring better understanding of the importance of this complete cycle.

In our model analysis, the xanthophyll cycle turns on after the reactions involved with the Calvin cycle and lipid production saturate, i.e when lipid production no longer utilises the increasing level of reductants, and is functional in conjugation with the glutathione-ascorbate cycle. The glutathione-ascorbate cycle, involves the antioxidant metabolites, ascorbate and glutathione, and detoxifies hydrogen peroxide produced as a waste product in metabolism and also acts as a sink for reducing power [173, 174]. Although, in this section of model analysis hydrogen peroxide is not being produced, or at least flux in such reactions are varying by less than 1%, the glutathione-ascorbate cycle is active along with the xanthophyll cycle to re-oxidize the NADPH pool. Cells grown in high light have been reported to double the pool size of XC pigments relative to cells grown in low light [175]. In addition to the xanthophyll cycle, diatoms possess a diadinoxanthin cycle and the levels of Ddx and Dtx were strongly affected by light intensity [32, 36].

4.4.4 Photorespiration

Model analysis suggests that photorespiration turns on in high light conditions and high glycerol uptake. Glycolate produced during the process can be either excreted out of the system or can be recycled within the system. The first step of glycolate recycling, in the model, involves its oxidation to glyoxylate, by glycolate dehydrogenase (GDH) and/or glycolate oxidase (GOX), which is further metabolised to glycine and serine.

In common with higher plants, marine algae are known to enhance photorespiration at high light, high O₂ concentration, low CO₂ concentration and elevated temperature. However, unlike higher plants, glycolate produced during the process can either be excreted out of the system or further metabolised, consistent with our model results. It has been reported that phytoplankton can excrete 5-40% of carbon as photorespiratory products. Leboulanger et al. [48] reported an increase in glycolate concentration in sea water during the day and decrease during the night, speculating that production largely outpaced the consumption of it by the heterotrophs, while at night there was net uptake by heterotrophs.

The specificity factor of Rubisco to discriminate CO₂ from O₂ is higher in diatoms than in cyanobacteria and green algae. Though this results in a lower rate of O₂ fixation in diatoms than green algae [176], excretion of glycolate was observed in *Thalassiosira weissflogii* cultures when shifted from low to high light, but not in continuous high light. There was no excretion of glycolate in cells placed under constant high light despite the high expression of glycine decarboxylase, suggesting the presence of a glycolate recycling mechanism [49]. Paul and Volcani [177] suggested the well-known route of glycolate recycling, present in higher plants and involving transamination to glycine to be present in *Cylindrotheca fusiformis*. Though the isotope labelling results suggest glycolate being recycled to serine through glycine in *P. tricornutum* [53, 54], the gene for glycerate kinase was not identified until recently [46, 55].

P. tricornutum has the metabolic capability to produce serine from both the non-photorespiratory and photorespiratory routes [52]. The model re-

sults suggest serine being produced through the photorespiratory route, i.e recycling of glycolate, in high light, and through the non-photorespiratory route via PGA in mitochondrion under low light, section 4.3.1, which seem to be energetically rational. A consequence of using the non-photorespiratory route is that it utilises the ED pathway and lower glycolysis within mitochondrion to produce PGA, thus NADH produced during the process can be utilised by the mitochondrial ETC for ATP synthesis.

The metabolic route through which serine can enter the Calvin cycle has not yet been experimentally reported in diatoms. As a result of our model analysis we propose two routes for serine degradation to pyruvate: i. through glycerate metabolism ii. through cysteine metabolism, a possible novel metabolic route for serine metabolism. In the former route, serine is metabolised to glycerate which is converted to PGA, at the expense of ATP by glycerate kinase, and enters the lower half of glycolysis present in the mitochondrion to produce PEP/pyruvate. Higher plants convert serine to glycerate via hydroxypyruvate, and then phosphorylate with glycerate kinase in the chloroplast which is also the conventional known route of glycolate recycling to PGA in higher plants [31]. The latter route involves synthesis of cysteine from serine, involving cysteine synthase, which is further degraded to produce pyruvate by cysteine desulfidase.

The overexpression of cysteine synthase in plants has been reported to be associated with alleviated reactive oxygen species damage [178]. In *P. tri-cornutum* an increased level of this enzyme has been reported under nitrogen stress [153]. Cysteine desulfidase is known to be present in bacteria and is involved in utilization of cysteine as a source of sulfur. It also protects the cells from growth inhibition due to high levels of cysteine which interfere with isoleucine, leucine, and valine biosynthesis via inhibition of threonine deaminase [179]. Though the enzyme activity has been reported to be present in *Arabidopsis*, the significance of this enzyme in plants is not well known [180]. Our model results suggest the possible role of these enzymes in serine metabolism; however, the biological significance of this route can be confirmed only with further studies on the kinetics and protein expression level of these enzymes.

With increase in light intensity, it is likely that O_2 concentration in chloroplast increases which in-turn would increase flux in photorespiration. However, in our model analysis flux in rubisco oxygenase decrease to zero at higher light intensity. This observed difference in the behaviour can be the artefact of over-stressing of the model. At high light, use of excess energy for lipid production along with other mechanisms to dissipate energy seem to minimise total flux as compared to photorespiration and as our objective function is minimisation of total flux, the first mechanism is preferred over photorespiration.

4.4.5 Lipid Production

Model analysis suggests that flux in lipid transporter (i.e. lipid production) increases under high light and high glycerol uptake. These results are consistent with experimental observations where algae including diatoms have been reported to increase lipid synthesis under high light [181–184] and availability of glycerol [80]. This increased lipid is generally stored in the form of TAGs as lipid droplet, as shown in figure 5.3. In physiological context, increased flux in lipid transporter in the model analysis can be regarded as lipid droplet formation.

Model analysis suggests that there exist two potential routes for lipid production in *P. tricornutum*: i. the phosphoketolase pathway, ii. threonine degradation. In the former route, phosphoketolase catalyses the conversion of X5P to GAP and acetyl-P. GAP enters the Calvin cycle/OPPP while acetyl-P is converted into acetate which acts as the precursor for lipid production. In the latter route, i.e through threonine degradation, PEP/pyruvate is carboxylated to OAA that enters the threonine synthesis pathway followed by its degradation to produce glycine and acetate. Glycine is further metabolised while acetate is utilised for lipid production. PEP/pyruvate required for this process can be synthesised through glycolysis and/or through recycling of glycolate. Hence, the model results not only shed light on the significance of the presence of the phosphoketolase pathway but also suggest the ultimate contribution of photorespiration towards lipid production.

The overexpression of enzymes involved in the phosphoketolase pathway has resulted in increased lipid production in *S. cerevisiae* [185]. Due to the increase of acetyl-CoA production, this route has been described as a potential alternative carbon route for different industrially relevant metabolites [186–188]. The lipid production route through threonine metabolism in our model analysis resembles the synthetic pathway proposed by Bar-Even [189] for formate assimilation. In the naturally occurring metabolic route, formate is fixed by the enzyme formate-tetrahydrofolate ligase (FTL) and in subsequent steps, the C is transferred to glycine to generate serine. Serine is converted to PEP via glycerate, carboxylated to OAA, and reduced to malate. Malate is then cleaved to glyoxylate which regenerates glycine, and acetyl-CoA, which serves as a biomass precursor. The proposed synthetic route, by Bar-Even [189], deaminates serine to PEP which is further carboxylated to OAA, instead of serine being converted to glycerate. OAA is further metabolised through threonine synthesis and degradation routes, as seen in our model analysis, to produce glycine and acetyl-CoA. Though proposed independently, the route for acetate production from our model analysis is consistent with the synthetic route proposed by Bar-Even [189].

However, it should be noted that these are not the only metabolic routes for lipid production in the metabolic network. Another reaction known to be present in *P. tricornutum* and suggested to contribute to lipid production is ATP-citrate lyase. In our model result, this reaction contributes towards lipid production; however the flux in this reaction changes by less than 1% and hence has not been shown explicitly in model results. In addition, the current version of the model does not include a plastidial Pyrdh reaction and acetyl-CoA transporter, which also contributes towards lipid synthesis as described in section 1.4.2. If these reactions are included in the model, then there exist two additional route, for acetyl-CoA synthesis: i. through Pyrdh complex using pyruvate in the plastid ii. acetyl-CoA produced through the phosphoketolase pathway is transported to the chloroplast without being converted to acetate.

4.4.6 Effect of Glycerol on Metabolism

In our model analysis, glycerol enters central carbon metabolism by its conversion to glycerol phosphate, which provides the glycerol backbone for TAG synthesis, by glycerol kinase, and then by glycerol-3-phosphate dehydrogenase to DHAP which is an intermediate for glycolysis and gluconeogenesis. Pyruvate produced through glycolysis can be fermented to lactate via LDH, and/or enters the TCA cycle, and/or carboxylated to OAA and utilised as a substrate for lipid synthesis, as discussed in 4.4.5, while G6P produced through gluconeogenesis contributes towards carbohydrate (chrysolaminarin) synthesis. Further, model analysis suggests active photorespiration in mixotrophic conditions, and glycolate produced during this process is metabolised through glyoxylate/TCA cycle. In addition, not just the conventional TCA cycle but also the variant of TCA cycle which involves decarboxylation of 2KG to SucAld followed by its oxidation to succinate is active though this variant of TCA cycle generates less ATP as compared to the conventional TCA cycle.

Experiments conducted under similar conditions, by our industrial partner Fermentalg, suggest glycerol-mediated effects on lipid and carbohydrate metabolism [138, 190]. In the same study, the microarray analysis shows up-regulation of genes encoding the enzymes TPI, FBPaldol, pyruvate carboxylase and genes associated with FA and TAG biosynthesis while metabolomic results show higher concentrations of pyruvate, lactate, diacylglycerol, trehalose, mannitol and xylose. Further, the experimental analysis on the fatty acid composition of TAGs in glycerol-treated cells suggests that their accumulation is mainly due to *de novo* synthesis and partially due to the degradation of some pre-existing membrane lipid [138, 190], confirming the contribution of glycerol to lipid as seen in the model. The same experimental study also suggest an increase in carbohydrate although the increase was observed in trehalose and mannitol, and not just chrysolaminarin as predicted by the model analysis. The possible reason for this difference in the observation might be due to increased osmotic pressure upon glycerol addition as the previous studies in both prokaryotes and eukaryotes [191–193], have shown

that both trehalose and mannitol are induced by osmotic stress.

The glycerol metabolism of *P. tricornutum* was previously investigated, using isotope labelling experiments with ^{13}C -glycerol as carbon source. These studies highlighted that, under mixotrophic conditions, *P. tricornutum* cells mostly convert this compound into glycine and serine through the photorespiratory route [53, 54], consistent with our model findings. However, no significant changes in the serine and glycine cellular contents were observed in our metabolomics data, while other amino acids show significant modifications upon glycerol addition [138, 190] and the probable reason can be the tight coupling between production and consumption of these compounds, as seen in our model analysis.

4.5 Conclusion

Overall, the model analysis suggest changes in central-carbon metabolism with increases in photorespiration and lipid production under high light conditions and availability of glycerol under mixotrophic conditions. Physiologically, microalgae [194] and marine diatoms [195, 196] including *P. tricornutum* [197] are known to increase lipid under high light and glycerol availability [138]. A proteomics study of *T. pseudonana* at high light showed upregulation of proteins associated with NPQ, fatty acids synthesis, lipid metabolism, Calvin cycle, photorespiration, light harvesting complexes [196] which seem to be consistent with our model observations. Further, model results also emphasis the role of photorespiration in glycerol metabolism as suggested by previous labelling experiments.

Hence, it is conceivable that the approach of scanning the solution space to varying constraints using FBA has enabled the identification of a number of metabolic modules close to realistic physiological behavior. While many of them are known to be associated with given conditions on the basis of their biochemistry, other responses are novel reflecting the metabolic capability and diversity of the network. Based on the model analysis, described in this chapter, experiments were designed to increase the lipid productivity in *P. tricornutum* and is described in next chapter.

Chapter 5

Experimental Setup and Results

5.1 Introduction

Results from the previous chapter suggest that lipid production in *P. tricornutum* increases both when exposed to high light and with the availability of glycerol in mixotrophic conditions. This has also been reported from experimental studies [80–82, 138, 195–197]. Further, model analysis suggested an important contribution from HCO_3^- towards lipid production both in phototrophic and mixotrophic conditions. Thus, I, with Valeria Villanova and Dr. Adeline Le Monnier, designed and performed experiments to test the effect of high light and glycerol, in the presence of HCO_3^- , on growth and biomass components during my secondment at Fermentalg. Additionally, model analysis suggests one of the potential routes for lipid production would be through threonine metabolism. Hence, the effect of threonine on culture growth and lipid production was also tested in the presence of HCO_3^- . This chapter mainly describes the experimental setup and results obtained for phototrophic and mixotrophic (on glycerol and threonine) conditions during the secondment. Note that the mixotrophic culture in this chapter will refer to growth on glycerol as the source of organic carbon, if not stated otherwise.

5.2 Method

5.2.1 Experimental Setup

Axenic cultures of *P. tricornutum* were grown in 2L photo-bioreactors (Applikon Schiedam, The Netherlands) in enriched artificial seawater (Enriched ESAW) medium as described by Villanova [138]. All cultures were sparged continuously with air at a flow rate of 0.5 L/min. The pH was maintained at 8 by automatic on-demand injection of 0.25 N HNO₃. Temperature was controlled at 20°C by circulating cold water. Cultures were exposed to continuous light, using an external light panel as shown in figure 5.1. As listed in table 5.2, the light intensity was increased gradually from 40 to 300 $\mu\text{mol Photon m}^{-2} \text{ s}^{-1}$ at regular intervals of time. At same time intervals, 15 mM of HCO₃⁻ was added to all three cultures and this was increased to 30 mM at later stages when the biomass concentration was high. In addition, two of the cultures were supplemented with 50 mM glycerol and 10 mM threonine respectively. The illustration of the experimental design is shown in figure 5.2. The effect of increasing light intensity, HCO₃⁻ and organic carbon was measured in terms of growth rate, percentage of sugar and FA (table 5.2). Though the initial plan was to perform at least 3 biological replicates for each conditions, only 3 photo-bioreactors were available in Fermentalg during the period of my secondment and thus, all the results presented here are based on single experiment for each condition.

5.2.2 Nitrogen and Phosphate Concentration

The demand for phosphate and nitrogen was met by supplementation with NaH₂PO₄ and HNO₃ respectively. Their concentrations in the supernatant were determining using a Merck RQflex reflectometer with test strips (Reflectoquant nitrate and phosphate). HNO₃ was also used to regulate the pH of the culture thus, the cultures were never deficient in terms of nitrogen while NaH₂PO₄ was added to the culture as needed.



Figure 5.1: The experimental setup in Fermentalg. Photo-bioreactor was surrounded by the light panel. Samples were removed from photo-bioreactors through the outlet channel connected to the sterilised syringe. Similarly, additional HCO_3^- , glycerol and threonine was injected into the medium through the inlet valve using the sterilised syringe.

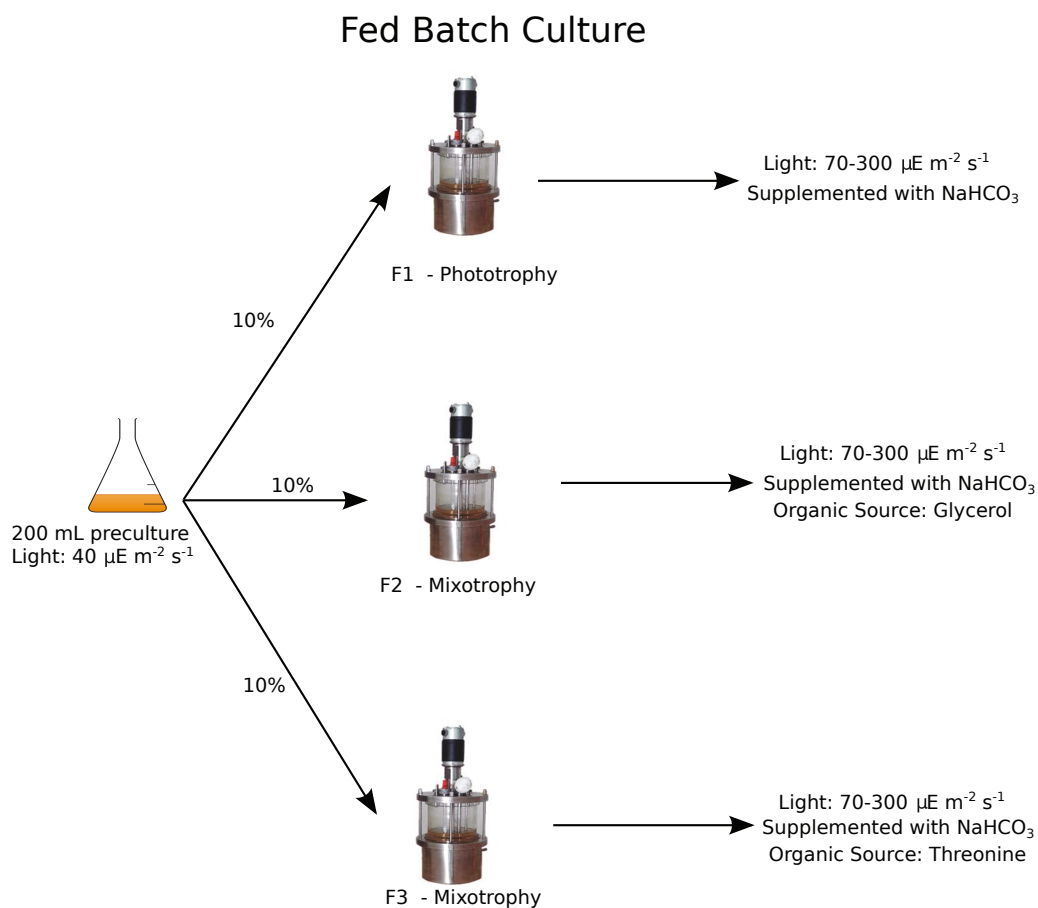


Figure 5.2: Schematic representation of the experimental design. Cultures were grown in three photo-bioreactors (F1, F2 and F3). All three cultures were exposed to increasing light and were supplemented with HCO_3^- , in addition to CO_2 . Two of the photo-bioreactors, F2 and F3, were further supplemented with organic carbons, glycerol and threonine, respectively. The effects of increasing light, HCO_3^- and organic substrates were measured in terms of the growth rate and biomass components.

5.2.3 Measurement of Growth

Growth of the culture was measured at regular intervals of time, as shown in table 5.2, in terms of optical density and dry weight. The optical density was measured at 750 nm using a double beam UV/visible spectrophotometer. Sample with optical density higher than 1.0 were diluted to get the absorbance within this range and then multiplied by the dilution factor. Dry weight was measured using Millipore membrane filter. For this aspect, millipore membrane filter was placed in an aluminium cup and dried for 24h in an oven at 105°C. Later, 2 to 5 ml of culture was filtered through these Millipore membrane filter which was then rinsed with sea water (to remove excess salt), placed in the same aluminium cup and dried for 24h in an oven at 105°C. The difference in weight of filter and aluminium cup before filtering the biomass and after filtering the biomass was regarded as the dry cell weight and expressed as g/L.

5.2.4 Nile Red Staining and Microscopy

TAG accumulation was monitored by Nile Red fluorescent staining. For this, 160 μ L cell suspension (1-5 million cells/mL) was added to 40 μ L of Nile Red dye (2.5 μ g/mL stock concentration, in 100% DMSO) in a 96-well microplate and mixed. The Nile Red fluorescence was analysed qualitatively under a microscope, after 20 minutes of incubation in the dark at room temperature.

5.2.5 Biomass Composition Measurement

The biomass composition of culture in terms of percentage carbohydrates, protein, lipids and fatty acid (FA) was measured by the analysis team in Fermentalg (for details refer to Villanova [138]), hence the protocol/method is not described here. In later sections, total carbohydrate and FA has been used for data analysis. Since, estimation of total lipid and protein required large volume of cell pellets, it was measured for few time points while the former were estimated more frequently (as shown in table 5.2).

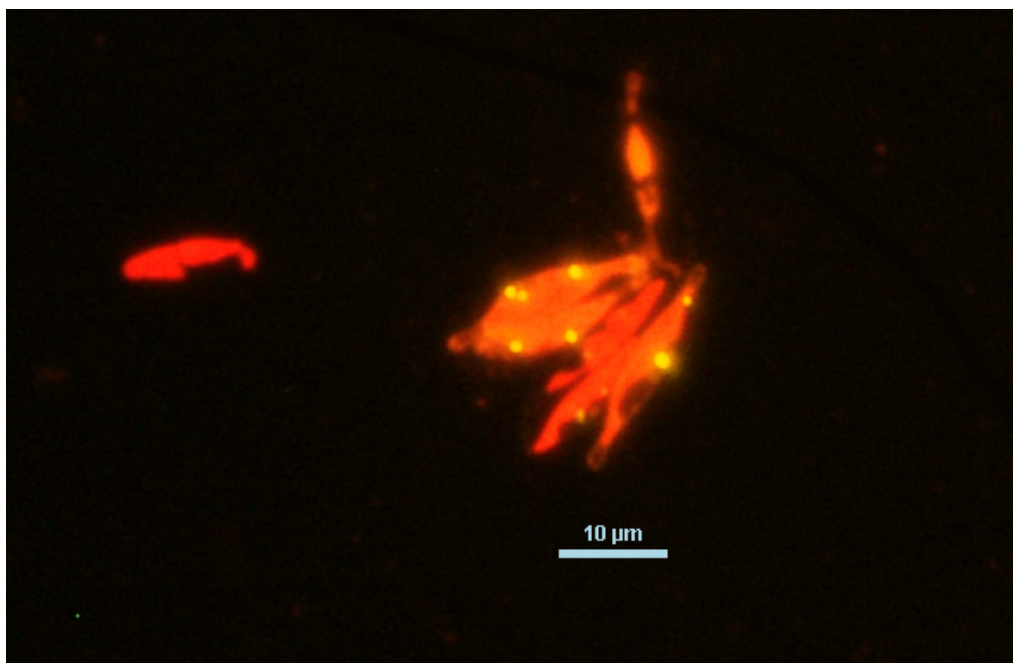


Figure 5.3: Lipid accumulation visualised using Nile Red stain. The yellow inclusions are the lipid bodies.

5.3 Results

Data presented in this section and their interpretation are based on single experiment performed for each condition. Due to lack of biological replicates, no statistical analysis could be performed. In certain scenarios (shown below) effect of varying parameters such as light, HCO_3^- and organic sources can be clearly seen however, in others it is not possible to derive a conclusive decision if results are significantly different or not and interpretations are based on mere visual observation.

5.3.1 Effect of bicarbonate on growth rate

The effect of HCO_3^- was analysed by comparing the growth rates of cultures grown in the presence of HCO_3^- to those grown in the absence of HCO_3^- under phototrophic and mixotrophic conditions, as shown in figure 5.4. Note that here data for cultures in absence of HCO_3^- were taken from previously

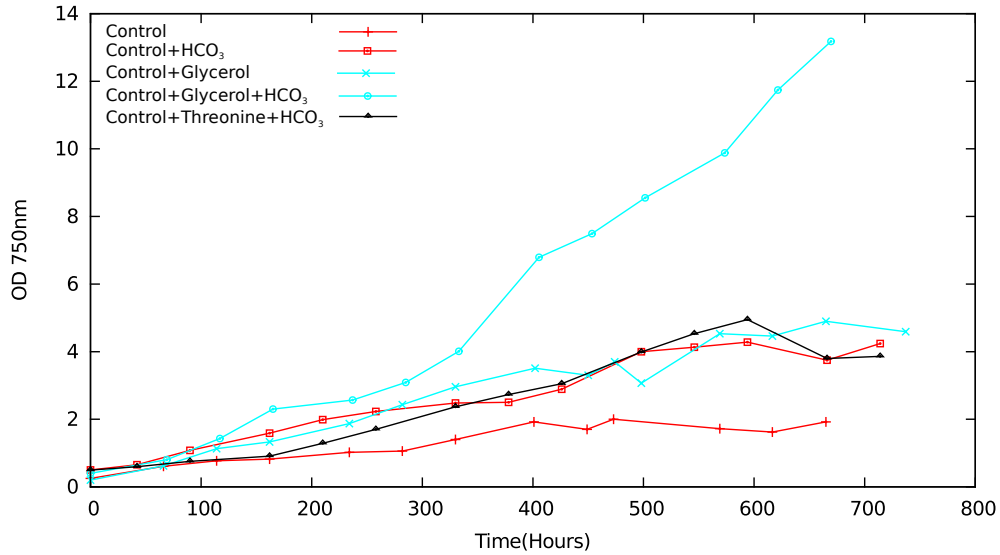


Figure 5.4: Growth rates in different culture conditions. Control cultures were grown under phototrophic conditions without any addition of HCO_3^- . Addition of HCO_3^- increased the growth rates in both phototrophic and mixotrophic conditions.

performed experiments, in Fermentalg, under similar conditions. Control cultures, i.e. cells grown under phototrophic condition without any addition of HCO_3^- , had the lowest growth rates. On addition of HCO_3^- , growth rates increased in phototrophic conditions and this was comparable to the growth rate in mixotrophic conditions in the absence of HCO_3^- . Mixotrophic cultures in the presence of HCO_3^- had the highest growth rate. The cultures grown with additional threonine in the presence of HCO_3^- had similar growth rates to those in phototrophic conditions in the presence of HCO_3^- .

5.3.2 Effect of bicarbonate as a function of absorbed light

Further, the effect of HCO_3^- was analysed as a function of absorbed light per amount of biomass. For this, absorbed light per amount of biomass for a given incident light intensity was calculated according to eq.3.2.2 described

Table 5.1: Growth response in terms of optical density, percentage FA and sugar and estimated light absorbed per amount biomass for cultures (from single experiment) grown with and without supplemental HCO_3^- in phototrophic and mixotrophic conditions, in the same external light intensity of $300 \mu\text{mol Photon m}^{-2} \text{ s}^{-1}$.

	Control	Control + HCO_3^-	Control+Glycerol	Control+Glycerol + HCO_3^-
Optical Density	2.5	4.1	8.2	9.9
Percentage FA	9.8	10.7	9.2	9.7
Percentage Sugar	6.8	4.1	8.0	2.9
Absorbed Light (mmol Photon gDW^{-1})	1320	740	375	170

in chapter 3. Table 5.1 lists the optical density, percentage storage compounds and estimated absorbed light per amount of biomass for the cultures grown with and without supplemented HCO_3^- (note that, as mentioned in earlier section, data for cultures in absence of HCO_3^- were taken from previously performed experiments) at the same external light intensity of $300 \mu\text{mol Photon m}^{-2} \text{ s}^{-1}$. Based on the pattern observed, we suggest that addition of HCO_3^- increases the growth rate and decreases total sugar in both phototrophic and mixotrophic conditions. There seems to be no significant difference in total FA (at least in mixotrophic condition). However, to make any firm claim these experiments need to be repeated with biological replicates. Nevertheless, with increase in growth, the light intensity available for the culture decreases and as shown in table 5.1, the cultures supplemented with HCO_3^- has less light supply rate as compared to its counterparts, i.e not supplemented with HCO_3^- ; still, increase in growth can be observed in the former cultures suggesting its contribution.

5.3.3 Effect of light intensity and bicarbonate availability

The effect of increasing external light intensity in the presence of different molar concentrations of HCO_3^- on growth and storage compounds in phototrophic and mixotrophic conditions is listed in table 5.2. For the given external light intensity, the absorbed light per amount of biomass was es-

Table 5.2: Growth rate (in terms of OD and dry weight), storage compounds (percentage FA and sugar) and estimated light absorbed per amount biomass at different light intensities and HCO_3^- concentrations in phototrophic and mixotrophic conditions. Note that data is based on single biological replicate.

	Time		Incident Light ($\mu\text{mol Photon m}^{-2} \text{ s}^{-1}$)	Volume Added		Measured Values			Estimated Values	
	Hours	($\mu\text{mol Photon m}^{-2} \text{ s}^{-1}$)		Glycerol (mL)	Threonine (mL)	OD	DW (g/L)	FA %	Sugar %	HCO_3^- mM
Control+ HCO_3^-	42	40	-	-	0.65	0.69	7.89	4.8	2.3	47
	90	70	-	-	1.08	0.97	10.07	6.9	17.5	125
	162	100	-	-	1.59	1.03	8.29	7.5	16.6	253
	210	130	-	-	1.99	1.44	9.78	8.4	14.7	291
	258	160	-	-	2.23	1.37	10.4	11.7	15.8	424
	330	210	-	-	2.48	1.55	11.32	9.0	15.2	521
	378	260	-	-	2.5	1.54	19.14	7.8	14.6	615
	426	300	-	-	2.89	2.1	12.71	6.4	30.5	606
	498	300	-	-	4.0	2.21	10.53	7.0	26.7	660
	546	300	-	-	4.13	2.55	10.71	4.1	31.3	738
	594	300	-	-	4.28	2.6	8.0	8.2	24.8	763
	69	40	9.0	-	0.80	0.75	4.22	4.8	1.87	51
	117	70	0.0	-	1.43	1.07	6.46	7.3	15.3	112
	165	100	7.0	-	2.3	1.36	9.9	9.5	14.7	141
Control+Glycerol+ HCO_3^-	237	130	7.0	-	2.57	1.83	10.45	7.5	14.4	166
	285	200	0.0	-	3.09	1.96	19.27	7.7	14.2	224
	333	300	8.0	-	4.01	2.78	12.64	6.1	26.6	215
	405	300	0.0	-	6.79	3.54	10.89	3.3	21.9	198
	453	300	6.5	-	7.49	4.45	11.53	11.4	24.1	194
	501	300	0.0	-	8.55	5.48	8.99	12.3	19.3	167
	573	300	0.0	-	9.88	6.6	9.64	2.9	20.0	167
	621	300	0.0	-	11.74	6.8	9.41	10.9	17.2	173
	669	300	0.0	-	13.18	7.3	12.24	14.9	17.2	173
	42	40	-	18.0	0.59	0.71	5.45	4.6	2.3	45
	90	70	-	-	0.75	0.72	6.85	4.9	2.4	178
	162	100	-	-	0.90	1.29	5.96	6.4	16.6	237
	210	130	-	-	1.29	0.95	5.65	5.8	15.6	498
	258	160	-	-	1.7	0.94	8.79	10.8	15.4	711
330	210	-	0.0	2.37	1.46	9.87	8.2	15.0	646	
378	260	-	-	2.73	1.49	16.69	7.6	26.5	740	
426	300	-	-	3.05	1.96	11.09	5.1	24.6	718	
498	300	-	-	3.99	2.07	9.92	4.7	22.8	738	
546	300	-	-	4.53	2.5	10.41	9.9	20.3	770	
594	300	-	-	4.95	2.83	7.77	9.1	19.9	723	

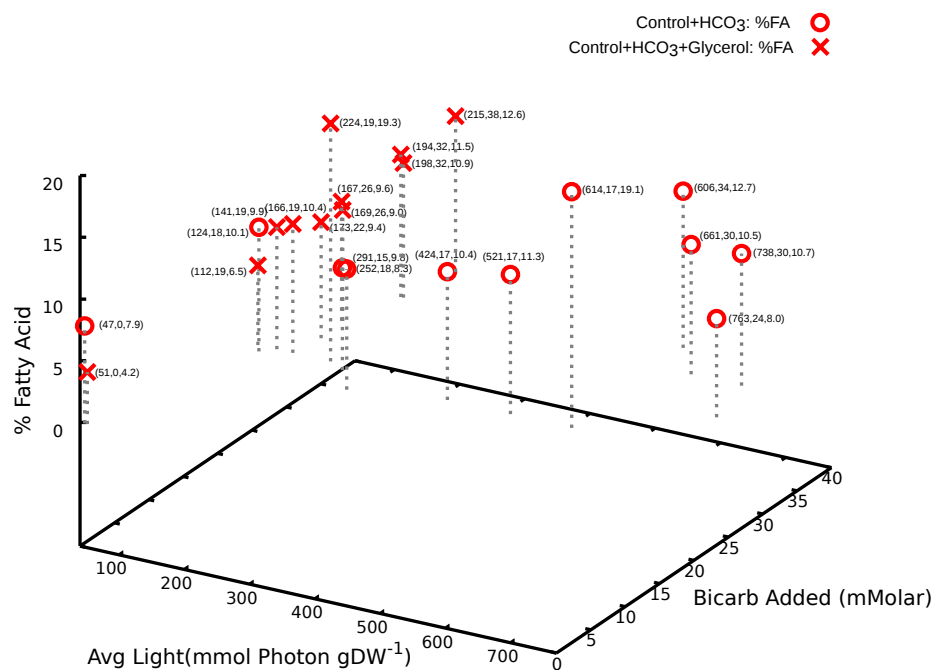
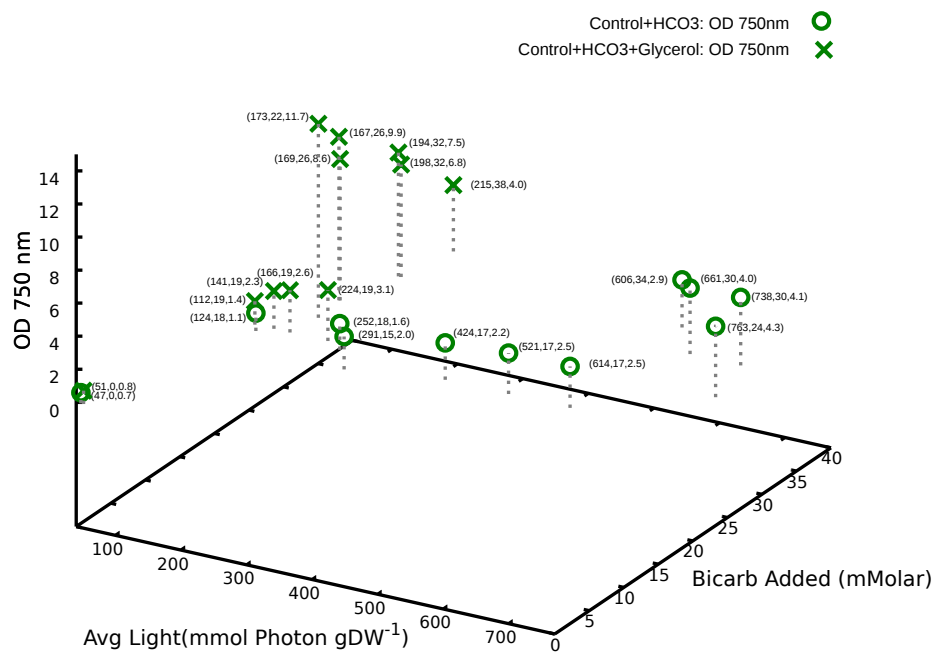


Figure 5.5: Graphical representation of data presented in table 5.2. Growth rate in terms of optical density and total percentage of FA, in phototrophic and mixotrophic (on glycerol) conditions, is plotted as a function of available light and added molar concentration of HCO₃⁻.

estimated according to eq.3.2.2 described in chapter 3. For ease of visualisation and interpretation, the OD and percentage FA for phototrophic and mixotrophic conditions (listed in table 5.2) have been plotted as a function of molar concentration of HCO_3^- and absorbed light per amount of biomass in figure 5.5. Higher light intensities and higher HCO_3^- concentrations shows positive effects on growth as well as lipid production in phototrophic and mixotrophic conditions. At similar HCO_3^- concentration, increase in light intensity increases growth rate as well as percentage FA. The percentage of FA, in both phototrophic and mixotrophic conditions, ranged from 5 to 20% and the highest percentage of FA, 20% of total biomass, in phototrophic conditions as well as in mixotrophic conditions, was observed at a HCO_3^- concentration of ≈ 15 mM.

The cultures grown with threonine had no significant differences in growth rate and total FA as compared to phototrophic conditions (table 5.2). Though, high light and high HCO_3^- increased the growth rate and FA production as compared to low light and low HCO_3^- , it was the same or even lower as compared to the phototrophic conditions.

5.4 Discussion

The most common practice to supplement the algal culture with inorganic carbon is to bubble air into the growth medium. The bio-availability of carbon can be increased by bubbling air with a higher CO_2 concentration. Another method to increase the concentration of DIC is to add HCO_3^- to the medium. Regardless of which method is used to supplement inorganic carbon, DIC can exist in different forms: dissolved CO_2 , carbonic acid (H_2CO_3), bicarbonate (HCO_3^-) and carbonate (CO_3^{2-}), depending on the pH of the solution. The uptake of some forms of DIC could be favoured over others among microalgae species due to the different affinities of the active transport enzymes for the species. For example, *Nannochloris maculata* and *Nannochloris atomus* have a greater affinity for dissolved CO_2 and H_2CO_3 as compared to *Nannochloropsis oculata* which has higher affinity for HCO_3^- . Nevertheless, an overall increase in DIC has been reported to increase the

growth rate and the amount of storage compounds in microalgae [198]. In addition, most microalgae can simultaneously assimilate organic carbon such as glycerol, fructose, glucose, lactose, mannose, acetate and these conditions have been reported to increase lipid and biomass productivity [80, 138].

Microalgal growth and lipid accumulation is also affected by various other factors such as pH, temperature, light and nutrient supply. Many microalgae have the ability to produce substantial amounts of TAG under stress conditions, such as high light, high pH, and limitation of nutrients such as nitrogen, phosphate, sulphate or silica [199–202]. Excess light intensities overly-reduce the NADPH pool and most microalgae are known to avoid this overreduction by increasing lipid production [23]. Nitrogen and phosphate starvation has been a popular method to induce lipid production in many microalgae [200, 203, 204] including *P. tricornutum* [87]. However, the drawback of this approach is that it limits the growth and one of the challenges, in the field of microalgal biofuel, is to increase the lipid production without constraining growth.

Model analysis suggested an increase in lipid production at high light intensity. Further, it also suggested a contribution of HCO_3^- fixation toward lipid production, both under phototrophic and mixotrophic conditions (refer to section 4.3). The experiment designed on the basis of model analysis resulted in an increase in growth rate and lipid production, as shown in figure 5.5, with increases in light intensity and addition of HCO_3^- . Though our experimental results are based on single biological replicate and hence, we cannot make a firm claim, results are consistent with other experimental studies. Earlier studies [205, 206], report that addition of HCO_3^- halts the cell cycle in *Scenedesmus sp. WC-1* and *C. reinhardtii* and addition of HCO_3^- above this minimal concentration (to arrest cellular cycling) enhances TAG accumulation. However in *P. tricornutum*, addition of HCO_3^- induces TAG accumulation without stopping the cell cycle [205], which is also consistent with our result where both increases in growth and lipid production are observed on addition of HCO_3^- . Quantitatively, we observe the highest increase in lipid production at a HCO_3^- concentration of 15 mM. This value is consistent with Guihéneuf and Stengel [207] who reports TAG accumu-

lation on HCO_3^- addition in *P. lutheri*, though under nitrogen starvation. On the other hand, light intensity, in our experiments, was increased from 40 to 300 $\mu\text{mol Photon m}^{-2} \text{s}^{-1}$ (upper limit was restricted because of the availability of light panels in Fermentalg). This irradiance itself is below the extremes that phytoplankton cells can encounter in environment which can reach upto 740 $\mu\text{mol Photon m}^{-2} \text{s}^{-1}$ for shallow mixed coastal waters and even 1200 $\mu\text{mol Photon m}^{-2} \text{s}^{-1}$ on fluctuating light regimes [208]. Moreover, shading of light due to high cell density further reduces the actual light intensity available for cell growth. This can be clearly seen in our results, table 5.1, where cultures with high algal growth has much less available light though exposed to same external light intensity. Therefore, it is reasonable to assume that cultures in our experiments were light limited and it would be possible to further increase lipid and growth by using even higher light intensities or thinner liquid films to avoid shading [82].

One obvious possible reason for the low growth rates and storage products for cultures grown without HCO_3^- , at least in phototrophic condition in our experiments, could be the severe carbon limitation as the only source of carbon in this condition is the sparged air (0.04% CO_2). In mixotrophic conditions, it is less likely that the culture is carbon limited, as it has organic sources of carbon. However, addition of HCO_3^- in these conditions still increases the growth and lipid production, suggesting assimilation of both organic and inorganic carbon by the culture. Addition of glycerol increased the growth rate by 2 fold, however as compared to phototrophic conditions, there was no significant difference in total FA (even total lipid though data not shown here). This might be an underestimation because we are just analysing the total FA, and total lipid was measured for just two time points. Nevertheless, the results do suggest carbon partitioning between growth and storage. On the contrary, addition of threonine had no significant effect on growth nor on lipid production. The speculated possible reason is that threonine might not have been absorbed/transported in the cells. Even biolog study shows no significant effect of threonine on growth [138]. Though much has been studied on the amino acid transport system in *E. coli* [209], not much is known for *P. tricornutum*.

5.5 Conclusion

Experimental designs based on the model analysis, i.e addition of HCO_3^- and increasing light intensity, resulted in increases in growth and lipid production in both phototrophic and mixotrophic conditions. Though this experiment was not aimed to focus on the metabolic routes, we can predict the metabolic routes from the modelling results that would have prompted this metabolic shift at high light intensities and availability of HCO_3^- . Our speculation that threonine addition would increase the lipid production needs to be studied further for the transport mechanism. Further, to be more conclusive, it would be worth to repeat these experiments with biological replicates in order to perform statistical analysis and have a quantitative estimate of glycerol and threonine consumption which was not done owing to various technical reasons. In addition, the results suggest that, despite cultures being exposed to high light intensities, the available light for biomass production is relatively low and overcoming this limitation would possibly further increase lipid and growth.

Chapter 6

General Discussion

6.1 Overview of Results

In this study, genome scale metabolic models of *P. tricornutum* were constructed and used to predict metabolic fluxes in phototrophic and mixotrophic conditions. We have described and used a novel approach to FBA where flux in biomass transporters are allowed to change, with minimisation as objective function, rather than setting fixed flux constraints on biomass transporters. This flexibility allowed the models to change the biomass composition depending upon the energy/carbon source. This analysis was further used to design the experiments, described in chapter 5, in order to optimise lipid production.

In chapter 3, two compartmentalised GSMs of *P. tricornutum* are described: one is an updated version of model published by Hunt et al. [122] while other is the *de novo* model constructed based on PGDB of *P. tricornutum*. Both the models were curated and validated for mass, energy and redox balance. They were analysed for the fundamental properties such as quantum demand, photosynthetic quotient and results were consistent with experimentally estimated values. Though, the Calvin cycle enzymes are distributed in plastid and cytosol in *P. tricornutum*, model was able to produce triose phosphate solely from plastidial Calvin cycle enzymes without the involvement of cytosolic Calvin cycle enzymes. Model analysis suggests that in

the absence of SBPase and SBPaldol in chloroplast, plastidial Calvin cycle is operation with active FBP and TAL reactions. Further analysis suggests the differences in the metabolic behaviour in phototrophic and mixotrophic conditions. The glycolytic routes such as EMP and ED pathway were favoured in mixotrophic condition while the flux through the Calvin cycle was favoured in phototrophic condition. Contrary to higher plants which have MALdh in different compartments and the reductant is transferred through a MAL/OAA shuttle, only mitochondrial MALdh has been reported in *P. tricornutum* and the mechanism of reductant transfer across compartments is not well understood. Model analysis suggests, in addition to PGA-GAP shuttle which is common to higher plants, reductant transfer through GLT/GLN-ornithine and citrulline/fumarate shuttle, as shown in figure 3.5, which might be linked to ornithine urea cycle (OUC) and TCA cycle as described in 3.3.2.

In Chapter 4, a metabolic model developed from Hunt et al. [122] was further analysed by solving FBA at varying *a*) light intensity, *b*) lipid demand and *c*) glycerol uptake. This method was used to identify the reactions that were correlated with the respective varying constraints (i.e light intensity, lipid production and glycerol metabolism) and the results suggest for some novel metabolic routes in addition to the metabolic route well reported in biochemistry books. For instance, *P. tricornutum* has two routes for serine synthesis: one via a photorespiratory route and another via a non-photorespiratory route [52]. Model results suggest that the non-photorespiratory route is active at low light while the photorespiratory route is active at high light which seems to be biologically relevant owing to the availability of energy. Serine produced as a result could be metabolised to re-enter central carbon metabolism through either glycerate metabolism, which is quite well known in plant biochemistry, or cysteine metabolism. The latter route has not been well reported in literature and further analysis of this route is required to determine its operation *in vivo*. In these results, precursors for lipid production are mainly contributed by the phosphoketolase pathway, glycolate recycling, HCO_3^- fixation and/or threonine degradation. This is an interesting finding as it highlights the importance of the phosphoketolase pathway in *P. tricornutum*, and the contribution of photorespiration toward

lipid production. The latter route, through threonine degradation, has also been suggested by Bar-Even [189] as the synthetic route for lipid precursor production from thermodynamic and energetic point of view. The model analysis also suggests that photorespiratory mechanism and novel route for glycolate recycling, in addition to the known conventional route through glycerate kinase and glyoxylate cycle. In this chapter, model analysis suggests an additional metabolic route for reductant transfer involving GLT-2KG shuttle where GLT and 2KG are exchanged between cytosol and mitochondrion and reversible GLTdh present in both the compartments facilitates their interconversion with net transfer of reductant from cytosol to mitochondrion. Overall, this chapter suggests potential metabolic routes that could be involved in biomass synthesis and energy dissipation and how preference over metabolic routes would shift depending upon the changing conditions.

Chapter 5 describes the experimental work (designed based on the model analysis) that was performed by myself, Valeria Villanova and Dr. Adeline Le Monnier during my secondment at Fermentalg. Three photo-bioreactors were used to grow *P. tricornutum* culture. All three were supplemented with HCO_3^- with increasing light intensities. Two of the cultures were further supplemented with glycerol and threonine respectively. The growth rate and biomass composition were analysed to see the effect of HCO_3^- , light and organic carbons on cultures. It should be noted that this experimental setup was not designed to study the metabolic route for lipid production rather the aim was to design the experiment, based on model results, to increase the lipid production. The experimental work resulted in an increase in both biomass productivity as well as lipid production suggesting contribution of light, HCO_3^- and glycerol. However, culture supplemented with threonine had no significant increase in growth as compared to phototrophic condition and thus, it can be anticipated that threonine may not be the best choice of substrate for *P. tricornutum* culture as had been proposed from the model analysis. This might be because of two reasons: either *P. tricornutum* do not have threonine transporters or cannot metabolise threonine in an effective way to support growth. An effective way to test this assumptions would be to measure threonine concentration in supernatant of culture and get an

estimation of threonine uptake by cells however, this was not possible during our experiment due to technical limitations.

6.2 Limitations and Future Works

Though, model results seem to be interesting and convincing, it is not spared from limitations of techniques used in this project and there are additional work that would be interesting and worth pursuing from modelling perspective. Foremost, though two GSMs were developed in this thesis, only Model I, that developed from Hunt et al. [122], was analysed in detail due to the time constrain. It would be interesting to analyse Model II, which has been constructed from PGDB of *P. tricornutum* under similar environmental and physiological conditions as Model I. Since, Model II has expanded metabolic network, it would not come as surprise if additional metabolic routes turns on or even if different metabolic response is observed due to higher degree of interconnection between metabolic routes. Further, lipid synthesis in both models is represented as lumped reaction. This is a useful assumption if the aim of study is to understand the metabolic routes for precursor production (as was in our study) however, in future it would be relevant and worthy to expand lumped reaction for lipid production, as in Levering et al. [110], in order to better understand environmental and physiological effect on fatty acid composition.

6.2.1 Alternative Objective Functions

In our study, the objective function is minimisation of the total flux which can be related to maximisation of efficiency in terms of machinery cost. However, organisms can prefer a trade off between various factors such as maximisation of efficiency and/or growth at different environmental conditions. Therefore, analysing the model with alternative objective functions such as maximisation of growth or even two-step optimisation approach, first with a linear objective function of maximisation of growth followed by a quadratic objective function of minimisation of the total fluxes, as described by Grafahrend-

Belau et al. [210], would be worth the study. Our approach of minimising the total flux, described in section 3.2.2, assumes the cost of each reaction (i.e enzyme machinery costs), w_i , to be equal to 1. However, this assumption is unlikely to be true *in vivo* as enzymes vary largely in terms of their size, number of subunits and catalytic activity. Thus, even enzyme machinery costs varies but such information is not available for all the enzymes in a GSM. Taking this into consideration, Cheung [211] proposed use of random weighting factors with minimisation as objective function to explore the effect of variability in the enzyme machinery cost on the resulting flux solution. In this approach, w_i of each reaction is assigned a random value between 0 and 1 from a uniform distribution and multiple flux solutions are computed by repeatedly solving optimisation problems, each time with a different set of random weighting factors. An averaged flux solution can be then calculated by averaging the multiple flux solutions generated with different sets of w_i .

6.2.2 Integration of Expression Data

Integration of expression data can be used to improve the prediction of metabolic flux distributions and to get the model results closer to the biological scenario [212, 213]. Various methods, such as regulatory FBA (rFBA) [212], probabilistic regulation of metabolism (PROM) [213], E-Flux [214], Gene inactivity moderated by metabolism and expression (GIMME) [215], transcriptional controlled FBA (tFBA) [216], have been proposed for the integration of expression level to metabolic and regulatory networks. rFBA simulates the steady state behaviour of the metabolic regulatory network where gene transcription factor interactions are defined using boolean logic (AND, OR, NOT). First, the presence of all regulated enzymes in the metabolic network is determined for a given time interval and if an enzyme is not present then the flux through that enzyme catalysed reaction is set to zero. FBA formulation is same as in equation eq.2.2.4, in addition to regulatory ON/OFF constraints:

$$V_j^c = 0 \quad \forall_j \in A^c$$

where V_j^c denotes flux through reaction j in condition c and A^c denotes set of reactions for which the corresponding gene are not expressed in condition c . PROM follows similar concept as rFBA but instead of using Boolean logic to turn ON and OFF the gene, it calculates probabilities of reaction to be active based on gene expression. E-Flux and GIMME uses expression data to restrict reaction fluxes. For example, E-Flux restrict the upper bound of reaction flux as shown:

$$V_j^c \leq \beta_j^c$$

where $\beta_j^c = f(\text{expression level of genes associated with reaction } j \text{ in condition } c)$. Alternative approach to incorporate expression data to the FBA, suggested by Hartman [135], involves changing the weight of the reactions as a function of expression level of gene/transcript associated with reaction:

$$wt_j = \frac{1}{2^{FC}}$$

where, wt_j is the weight of reaction j and FC represents fold change in expression level of gene/transcript associated with reaction j :

$$FC = \log_2\left(\frac{signal_{test}}{signal_{reference}}\right)$$

Therefore, if the gene associated with reaction is over-expressed then the cost of reaction is decreased and vice versa. Solving FBA as described in equation eq.3.2.1, with minimisation as the objective function, will tend to have reactions with decreased weight in order to decrease total flux. Hence, the solution set will have reactions associated with over-expressed genes and reactions associated with under-expressed genes would be included only if it was essential to get a steady state solution. To certain extend, we used this approach to incorporate transcriptomics data available for different light conditions [71]. The preliminary results (though not shown in this thesis) suggests increase in not only lipid but also carbohydrate in high light which is consistent to experimental results [138]. However, this study need to be explored in more details to make the best choice of available techniques and incorporate the available high throughput data.

6.2.3 Integration of Metabolomics Data

In addition to expression data, metabolomics data can be used/incorporated in metabolic models to constrain flux predictions and to relate network structural properties to metabolite levels (reviewed in Töpfer et al. [217]). Jerby et al. [218] makes use of metabolites that were detected in a given organ or tissue to reconstruct tissue-specific networks from a generic metabolism model. Another method for context-specific network extraction integrates metabolomics data by adding turnover metabolites and a respective sink reaction to the generic model. Therefore, a flux through a respective metabolic reaction can be obtained, and experimentally detected metabolites can be integrated by enforcing a minimum flux for the turnover of the respective metabolite [219]. Hoppe et al. [220] makes use of metabolite concentration to add thermodynamic constraints on the reversibility of reactions in order to predict more reliable flux distributions than a generic model. However, one should be cautious while incorporating metabolomics data to large scale metabolic models, which generally uses constraints based approach, that do not fully explain changes in the steady state concentration of internal metabolites.

6.3 Conclusions

In conclusion, it was possible using GSMs, constructed in this thesis, to get a better understanding of *P. tricornutum* metabolic network and its metabolic capability. Here, we also introduced new approach of model analysis using FBA where biomass production is allowed to vary in combination with minimisation as objective function. This modelling approach adds flexibility to model response and reveals potential metabolic routes and behaviour under a range of varying environmental and physiological conditions. In particular, this study reveals functioning of the Calvin cycle solely in chloroplast despite different localisation of the Calvin cycle enzymes in *P. tricornutum*. It also suggests novel metabolic routes for glycolate recycling, reductant transfer and contribution of phosphoketolase pathway, photorespiration and threonine

metabolism towards lipid production, and highlights potential candidates for metabolic engineering. Experiments performed, based on the model analysis, increased the biomass productivity and the lipid production to meet the overall goal of the project.

Bibliography

- [1] J. A. Foley. Global Carbon Emissions Projected to Reach Record High in 2013. Technical report, November 2013. URL <http://www.natureworldnews.com/articles/4990/20131119/global-carbon-emissions-projected-reach-record-high-2013.htm>.
- [2] BP Energy Outlook, 2013. BP Energy Outlook 2030. Technical report, January 2013. URL http://www.bp.com/content/dam/bp/pdf/energy-economics/energy-outlook-2015/bp-energy-outlook-booklet_2013.pdf.
- [3] S. N. Naik, V. V. Goud, P. K. Rout, and A. K. Dalai. Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 14(2):578–597, 2010.
- [4] B. K. Highina, I. M. Bugaje, and B. Umar. A review on second generation biofuel: a comparison of its carbon footprints. *European Journal of Engineering and Technology*, 2(2):117–125, 2014.
- [5] Y. Li-Beisson and G. Peltier. Third-generation biofuels: current and future research on microalgal lipid biotechnology. *OCL*, 20(6):1–8, 2013.
- [6] J. Lu, C. Sheahan, and P. Fu. Metabolic engineering of algae for fourth generation biofuels production. *Energy and Environmental Science*, 4: 2451–2466, 2011.
- [7] NREL. Aquatic species program close-out report, 1988. URL <http://www.nrel.gov/docs/legosti/fy98/24190.pdf>.

- [8] M. Hildebrand, A. K. Davis, S. R. Smith, J. C. Traller, and A. Raffaella. The place of diatoms in the biofuels industry. *Biofuels*, 3:221–240, 2012.
- [9] P. G. Falkowski, R. T. Barber, and V. Smetacek. Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science*, 281(5374): 200–206, 1998.
- [10] C. Hoek, D. Mann, and H. M. Jahns. *Algae: An Introduction to Phy-cology*. Cambridge University Press, 1995.
- [11] K. K. Sharma, H. Schuhmann, and P. M. Schenk. High lipid induction in microalgae for biodiesel production. *Energies*, 5(5):1532–1553, 2012.
- [12] B. R. Moser. Biodiesel Production, Properties, and Feedstocks. In *Biofuels: Global Impact on Renewable Energy, Production Agriculture, and Technological Advancements*, pages 285–347. Springer, New York, 2011.
- [13] A. Prabu and R. B. Anand. Production and Application of Biodiesel—A Case Study. *International Journal of Engineering Research and Devel-opment*, 2(28-41):206–216, 2012.
- [14] M. G. Kelly, H. Bennion, E. J. Cox, B. Goldsmith, J. Jamieson, S. Jug-gins, D. G. Mann, and R. J. Telford. Common freshwater diatoms of Britain and Ireland: an interactive key, Environment Agency, Bristol, 2005. URL <http://craticula.ncl.ac.uk/EADiatomKey/html/>.
- [15] R. Brayner, A. A. Coutè, J. Livage, C. Perrette, and C. Sicard. Micro-algal biosensors. *Analytical and Bioanalytical Chemistry*, 401(2):581–597, 2011.
- [16] F. E. Round and R. M. Crawford. The lines of evolution of the Bacil-lariophyta. I. Origin. *Proceedings of the Royal Society of London B: Biological Sciences*, 211(1183):237–260, 1981.
- [17] L. K. Medlin, W. H. C. F. Kooistra, and A. M. M. Schmid. A review of the evolution of the diatoms - a total approach using molecules,

- morphology and geology. In *The origin and early evolution of the diatoms: fossil, molecular and biogeographical approaches*, pages 13–35. A. Witkowski and J. Sieminska, Polish Academy of Sciences, Krakow, Poland, 2000.
- [18] L. K. Medlin and I. Kaczmarska. Evolution of the diatoms: V. Morphological and cytological support for the major clades and a taxonomic revision. *Phycologia*, 43(3):245–270, 2004.
- [19] D. M. Williams and P. Kociolek. Pursuit of a natural classification of diatoms: history, monophyly and the rejection of paraphyletic taxa. *European Journal of Phycology*, 42(3):313–319, 2007.
- [20] P. A. Sims, D. G. Mann, and L. K. Medlin. Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia*, 45(4):361–402, 2006.
- [21] A. J. Edgar and N. W. Edward. On the artificial culture of marine plankton organisms. *Journal of the Marine Biological Association*, 8(5):421–474, 1910.
- [22] D. Werner. *The Biology of Diatoms*. University of California Press, Berkley and Los Angeles, California, 1977.
- [23] Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, and A. Darzins. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *The Plant Journal*, 54(4):621–639, 2008.
- [24] J. A. Raven and A. M. Waite. The evolution of silicification in diatoms: inescapable sinking and sinking as escape? *New Phytology*, 162(1):45–61, 2004.
- [25] C. Wilhelm, C. Bchel, J. Fisahn, R. Goss, T. Jakob, J. Laroche, J. Lavaud, M. Lohr, U. Riebesell, K. Stehfest, K. Valentin, and P. G. Kroth. The regulation of carbon and nutrient assimilation in diatoms is significantly different from green algae. *Protist*, 157(2):91–124, 2006.

- [26] A. Moustafa, B. Beszteri, U. G. Maier, C. Bowler, K. Valentin, and D. Bhattacharya. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science*, 324(5935):1724–1726, 2009.
- [27] M. H. Sibley and L. E. Anderson. Light/Dark modulation of enzyme activity in developing barley leaves. *Plant Physiology*, 91(4):1620–1624, 1989.
- [28] L. E. Anderson. Light inactivation of transaldolase in pea leaf chloroplasts. *Biochemical and Biophysical Research Communications*, 99(4):1199–1202, 1981.
- [29] L. E. Anderson. Light/dark modulation of enzyme activity in plants. *Advances in Botanical Research*, 12:1–46, 1986.
- [30] P. Schurmann and J.-P. Jacquot. Plant thioredoxin systems revisited. *Annual Review of Plant Physiology and Plant Molecular Biology*, 51:371–400, 2000.
- [31] A. Wingler, P. J. Lea, P. W. Quick, and R. C. Leegood. Photorespiration: Metabolic Pathways and Their Role in Stress Protection. *Philosophical Transactions: Biological Sciences*, 355(1402):1517–1529, 2000.
- [32] M. Lohr and C. Wilhelm. Algae displaying the diadinoxanthin cycle also possess the violaxanthin cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 96(15):8784–8789, 1999.
- [33] B. Demmig, K. Winter, A. Krüger, and F. C. Czygan. Photoinhibition and Zeaxanthin Formation in Intact Leaves : A Possible Role of the Xanthophyll Cycle in the Dissipation of Excess Light Energy. *Plant Physiology*, 84(2):218–224, 1987.
- [34] B. Demmig, A. M. Gilmore, and W. W. Adams 3rd. Carotenoids 3: in vivo function of carotenoids in higher plants. *FASEB Journal*, 10(4):403–412, 1996.

- [35] P. Müller, X. P. Li, and K. K. Niyogi. Non-Photochemical Quenching. A Response to Excess Light Energy. Xiao-Ping Li. *Plant Physiology*, 125(4):1558–1566, 2001.
- [36] D. Latowski, S. Schaller-Laudel, M. Olchawa-Pajor, R. Goss, and K. Strzalka. Violaxanthin and diadinoxanthin cycles as an important photoprotective mechanism in photosynthesis. *Russian Journal of Plant Physiology*, 58(6):952, 2011.
- [37] A. E. Allen, B. B. Ward, and B. Song. Characterization of diatom (Bacillariophyceae) nitrate reductase genes and their detection in marine phytoplankton communities. *Journal of Phycology*, 41(1):95–104, 2005.
- [38] C. Bowler, A. Vardi, and A. E. Allen. Oceanographic and biogeochemical insights from diatom genomes. *Annual Review Marine Science*, 2:333–365, 2010.
- [39] B. J. Miflin and P. J. Lea. Ammonia assimilation. In *The Biochemistry of Plants: A Comprehensive Treatise*, pages 169–202. Academic Press, New York, 1980.
- [40] D. L. Nelson, A. L. Lehninger, and M. M. Cox. *Lehninger principles of biochemistry*. Macmillan, 2008.
- [41] R. B. Helling. Why does *Escherichia coli* have two primary pathways for synthesis of glutamate? *Journal of Bacteriology*, 176(13):4664–4668, 1994.
- [42] R. B. Helling. Pathway choice in glutamate synthesis in *Escherichia coli*. *Journal of Bacteriology*, 180(17):4571–4575, 1998.
- [43] C. Martin and A. M. Smith. Starch biosynthesis. *The Plant Cell*, 7(7):971–85, 1995.
- [44] C. Masterson, C. Wood, and D. R. Thomas. L-acetylcarnitine, a substrate for chloroplast fatty acid synthesis. *Plant, Cell and Environment*, 13(8):755–765, 1990.

- [45] Y. Li-Beisson, B. Shorosh, and F. Beisson. Acyl-Lipid Metabolism. *The Arabidopsis Book / American Society of Plant Biologists*, 11:e0161, 2013.
- [46] P. G. Kroth, A. Chiovitti, A. Gruber, V. Martin-Jezequel, T. Mock, M. S. Parker, M. S. Stanley, A. Kaplan, L. Caron, T. Weber, U. Maheswari, V. E. Armbrust, and C. Bowler. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricoratum* deduced from comparative whole genome analysis. *PLoS One*, 3(1):e1426, 2008.
- [47] M. Fabris, M. Matthijs, S. Rombauts, W. Vyverman, A. Goossens, and G. J. E. Baart. The metabolic blueprint of *Phaeodactylum tricoratum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *Plant Journal*, 70(6):1004–1014, 2012.
- [48] C. Leboulanger, L. Oriol, H. Jupin, and C. Desolas-gros. Diel variability of glycolate in the eastern tropical Atlantic Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, 44:2131–2139, 1997.
- [49] M. S. Parker, V. E. Armbrust, J. Piovia-Scott, and R. G. Keil. Induction of photorespiration by light in the centric diatom *Thalassiosira weissflogii* (bacillariophyceae): molecular characterization and physiological consequence. *Journal of Phycology*, 40:557–567, 2004.
- [50] W. J. Bruin, E. B. Nelson, and N. E. Tolbert. Glycolate pathway in green algae. *Plant Physiology*, 46(3):386–91, 1970.
- [51] J. Beardall, D. Mukerji, H. E. Glover, and I. Morris. The path of carbon in photosynthesis by marine phytoplankton. *Journal of Phycology*, 12:409–417, 1976.
- [52] M. A. Bromke. Amino Acid Biosynthesis Pathways in Diatoms. *Metabolites*, 3:294–311, 2013.
- [53] Y. Zheng, A. H. Quinn, and G. Sriram. Experimental evidence and isotopomer analysis of mixotrophic glucose metabolism in the marine

- diatom *Phaeodactylum tricornutum*. *Microbial Cell Factories*, 12:109–126, 2013.
- [54] A. Huang, L. Liu, C. Yang, and G. Wang. *Phaeodactylum tricornutum* photorespiration takes part in glycerol metabolism and is important for nitrogen-limited response. *Biotechnology for Biofuels*, 8:73–89, 2015.
- [55] S. R. Smith, R. M. Abbriano, and M. Hildebrand. Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. *Algal Research*, 1(1):2–16, 2012.
- [56] M. F. Liaud, C. Lichtl, K. Apt, W. Martin, and R. Cerff. Compartment-specific isoforms of TPI and GAPDH are imported into diatom mitochondria as a fusion protein: evidence in favor of a mitochondrial origin of the eukaryotic glycolytic pathway. *Molecular Biology and Evolution*, 17(2):213–23, 2000.
- [57] N. Entner and M. Doudoroff. Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*. *Journal of Biological Chemistry*, 196(2):853–862, 1952.
- [58] T. Conwat. The Entner-Doudorodd pathway: history, physiology and molecular biology. *FEMS Microbiology Reviews*, 103(1):1–28, 1992.
- [59] G. Gottschalk. *Bacterial metabolism*. Springer-Verlag, New York, 2nd edition, 1986.
- [60] L. Meile, L. M. Rohr, T. A. Geissmann, M. Herensperger, and M. Teuber. Characterization of the D-xylulose-5-phosphate/D-fructose-6-phosphate phosphoketolase gene (xfp) from *Bifidobacterium lactis*. *Journal of Bacteriology*, 183(9):2929–2936, 2001.
- [61] B. Sánchez, M. Zúñiga, F. González-Candelas, C. G. de los Reyes-Gavilán, and A. Margolles. Bacterial and Eukaryotic Phosphoketolases: Phylogeny, Distribution and Evolution. *Journal of Molecular Microbiology and Biotechnology*, 18(1):37–51, 2010.

- [62] V. E. Armbrust, J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kröger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, and D. S. Rokhsar. The Genome of the Diatom *Thalassiosira Pseudonana*: Ecology, Evolution, and Metabolism. *Science*, 306(5693):79–86, 2004.
- [63] C. Bowler, A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, and Florian. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*, 456(7219):239–244, 2008.
- [64] T. P. Mommsen and P. J. Walsh. Evolution of urea synthesis in vertebrates: the piscine connection. *Science*, 243:72–75, 1989.
- [65] A. E. Allen, C. L. Dupont, M. Obornik, A. Horak, A. Nunes-Nesi, J. P. McCrow, H. Zheng, D. A. Johnson, H. Hu, A. R. Fernie, and C. Bowler. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature*, 473(7346):203–207, 2011.
- [66] A. Beattie, E. L. Hirst, and E. Percival. Studies on the metabolism of the Chrysophyceae. Comparative structural investigations on leucosin (chrysolaminarin) separated from diatoms and laminarin from the brown algae. *Biochemical Journal*, 79(3):531–537, 1961.
- [67] C. W. Ford and E. Percival. The carbohydrates of *Phaeodactylum tri-cornutum*. Part I. Preliminary examination of the organism, and characterisation of low molecular weight material and of a glucan. *Journal of the Chemical Society*, 0:7035–7041, 1965.

- [68] B. S. Paulsen and S. Myklestad. Structural studies of the reserve glucan produced by the marine diatom *Skeletonema costatum* (grev.) Cleve. *Carbohydrate Research*, 62(2):386–388, 1978.
- [69] M. J. McConville, A. Bacic, and A. E. Clarke. Structural studies of chrysolaminaran from the ice diatom *Stauroneis amphioxys* (Gregory). *Carbohydrate Research*, 153(2):330–333, 1986.
- [70] G. Michel, T. Tonon, D. Scornet, J. M. Cock, and B. Kloareg. Central and storage carbon metabolism of the brown alga *Ectocarpus siliculosus*: insights into the origin and evolution of storage carbohydrates in Eukaryotes. *New Phytologist*, 188(1):67–81, 2010.
- [71] M. S. Chauton, P. Winge, T. Brembu, O. Vadstein, and A. M. Bones. Gene regulation of carbon fixation, storage, and utilization in the diatom *Phaeodactylum tricorutum* acclimated to light/dark cycles. *Plant Physiology*, 161(2):1034–1048, 2013.
- [72] O. Emanuelsson, S. Brunak, G. von Heijne, and H. Nielsen. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, 2(4):953–971, 2007.
- [73] T. G. Tornabene, G. Holzer, S. Lien, and N. Burris. Lipid composition of the nitrogen starved green alga *Neochloris oleoabundans*. *Enzyme and Microbial Technology*, 5(6):435–440, 1983.
- [74] T. Tonon, D. Harvey, T. R. Larson, and I. A. Graham. Long chain polyunsaturated fatty acid production and partitioning to triacylglycerols in four microalgae. *Phytochemistry*, 61(1):15–24, 2002.
- [75] C. Y. Chen, K. L. Yeh, R. Aisyah, D. J. Lee, and J. S. Chang. Cultivation, photobioreactor design and harvesting of microalgae for biodiesel production: A critical review. *Bioresource Technology*, 102(1):71–81, 2011.

- [76] S. Xue, Z. Su, and W. Cong. Growth of *Spirulina platensis* enhanced under intermittent illumination. *Journal of Biotechnology*, 151(3):271–277, 2011.
- [77] E. Albers, J. Mayers, and O. Svensson. Mixotrophy in microalgae diverse metabolic modes for utilisation of organic carbon in relation to photosynthesis. In *International Study Group for Systems Biology Meeting, Oct 4 - Oct 7, Jena, Germany*, 2016.
- [78] J. Zhan, J. Rong, and Q. Wang. Mixotrophic cultivation, a preferable microalgae cultivation mode for biomass/bioenergy production, and bioremediation, advances and prospect. *International Journal of Hydrogen Energy*, 42(12):8505–8517, 2017.
- [79] X. Liu, S. Duan, A. Li, N. Xu, Z. Cai, and Z. Hu. Effects of organic carbon sources on growth, photosynthesis, and respiration of *Phaeodactylum tricornerutum*. *Journal of Applied Phycology*, 21:239–246, 2009.
- [80] M. Garci, J. M. F. Sevilla, F. G. A. Fernández, E. M. Grima, and F. G. Camacho. Mixotrophic growth of *Phaeodactylum tricornerutum* on glycerol: growth rate and fatty acid profile. *Journal of Applied Phycology*, 12:239–248, 2004.
- [81] M. D. C. C. García, A. S. Miròn, J. M. F. Sevilla, E. M. Grima, and F. G. Camacho. Mixotrophic growth of the microalga *Phaeodactylum tricornerutum*: Influence of different nitrogen and organic carbon sources on productivity and biomass composition. *Process Biochemistry*, 40(1): 297–305, 2005.
- [82] M. D. C. C. García, J. M. F. Sevilla, A. S. Miròn, F. G. Camacho, A. C. Gómez, and E. M. Grima. Mixotrophic growth of *Phaeodactylum tricornerutum* on fructose and glycerol in fed-batch and semi-continuous modes. *Bioresource Technology*, 147:569–576, 2013.
- [83] J. C. Lewin. Heterotrophy in Diatoms. *Journal of general microbiology*, 9:305–318, 1953.

- [84] J. C. Lewin. The taxonomic position of *Phaeodactylum tricorutum*. *Journal of General Microbiology*, 18:427–482, 1958.
- [85] S. Scala, N. Carels, A. Falciatore, M. L. Chiusano, and C. Bowler. Genome properties of the diatom *Phaeodactylum tricorutum*. *Plant Physiology*, 129(3):993–1002, 2002.
- [86] A. Montsant, K. Jabbari, U. Maheswari, and C. Bowler. Comparative Genomics of the Pennate Diatom *Phaeodactylum tricorutum*. *Plant Physiology*, 137:500–513, 2005.
- [87] E. T. Yu, F. J. Zendejas, P. D. Lane, S. Gaucher, B. A. Simmons, and T. W. Lane. Triacylglycerol accumulation and profiling in the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricorutum* (Bacillariophyceae) during starvation. *Journal of Applied Phycology*, 21(6):669, 2009.
- [88] T. Lebeau and J. M. Robert. Diatom cultivation and biotechnologically relevant products. Part I: cultivation at various scales. *Applied Microbiology Biotechnology*, 60(6):612–623, 2003.
- [89] R. Heinrich and S. Schuster. *The regulation of cellular systems*. Chapman and Hall, London, England, 1996.
- [90] K. B. Douglas. The Virtual Human: Towards a Global Systems Biology of Multiscale, Distributed Biochemical Network Models. *IUBNB Life*, 59(11):689–695, 2007.
- [91] T. Pfeiffer, I. Sanchez-Valdenebro, J. C. Nuno, F. Montero, and S. Schuster. Metatool: for studying metabolic networks. *Bioinformatics*, 15(3):251–257, 1999.
- [92] S. Schuster and C. Hilgetag. On Elementary Flux Modes in Biochemical Systems at Steady State. *Journal of Biological Systems*, 2:165–182, 1994.

- [93] C. T. Trinh, A. Wlaschin, and F. Sreenc. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Applied Environmental Microbiology*, 81(5):813–826, 2009.
- [94] J. C. Liao, S. Y. Hou, and Y. P. Chao. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnology and Bioengineering*, 52(1):129–140, 1996.
- [95] C. T. Trinh, P. Unrean, and F. Sreenc. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Applied Environmental Microbiology*, 74(12):3634–3643, 2008.
- [96] S. J. Van Dien and M. E. Lidstrom. Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph *Methylobacterium extorquens* AM1, with application to reconstruction of C3 and C4 metabolism. *Biotechnology and Bioengineering*, 78(3):296–312, 2002.
- [97] M. G. Poolman, D. A. Fell, and C. A. Raines. Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. *European Journal Biochemistry*, 270(3):430–439, 2003.
- [98] D. A. Fell and R. J. Small. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochemical Journal*, 238(3):781–6, 1986.
- [99] A. Varma and B. O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology*, 60(10):3724–3731, 1994.
- [100] H.-G. Holzhütter. The generalized flux-minimization method and its application to metabolic networks affected by enzyme deficiencies. *Biosystems*, 83(2-3):98–107, 2006.

- [101] M. G. Poolman, L. Miguët, L. J. Sweetlove, and D. A. Fell. A genome-scale metabolic model of *Arabidopsis* and some of its properties. *Plant Physiology*, 151(3):1570–1581, 2009.
- [102] M. R. Watson. A discrete model of bacterial metabolism. *Computer Applications in the Biosciences*, 2(1):23–27, 1986.
- [103] A. Varma and B. O. Palsson. Metabolic capabilities of *Escherichia coli*: I. synthesis of biosynthetic precursors and cofactors. *Journal of Theoretical Biology*, 165(4):477–502, 1993.
- [104] A. Varma and B. O. Palsson. Metabolic Capabilities of *Escherichia coli*: II. Optimal Growth Patterns. *Journal of Theoretical Biology*, 165:503–522, 1993.
- [105] J. S. Edwards and B. O. Palsson. Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics*, 1(1):1, 2000.
- [106] N. R. Boyle and J. A. Morgan. Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC Systems Biology*, 3(1):4, 2009.
- [107] A. R. Zomorodi, P. F. Suthers, S. Ranganathan, and C. D. Maranas. Mathematical optimization applications in metabolic networks. *Metabolic Engineering*, 14(6):672–686, 2012.
- [108] M. G. Poolman, S. Kundu, R. Shaw, and D. A. Fell. Responses to light intensity in a genome-scale model of rice metabolism. *Plant Physiology*, 162(2):1060–1072, 2013.
- [109] J. Kim, M. Fabris, G. Baart, M. K. Kim, A. Goossens, W. Vyverman, P. G. Falkowski, and D. S. Lun. Flux balance analysis of primary metabolism in the diatom *Phaeodactylum tricornutum*. *Plant Journal*, 85(1):161–176, 2016.

- [110] J. Levering, J. Broddrick, C. L. Dupont, G. Peers, K. Beeri, J. Mayers, A. A. Gallina, A. E. Allen, B. O. Palsson, and K. Zengler. Genome-Scale Model Reveals Metabolic Basis of Biomass Partitioning in a Model Diatom. *PLoS ONE*, 11(5):1–22, 2016.
- [111] K. Raman and N. Chandra. Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics*, 10(4):435, 2009.
- [112] E. P. Gianchandani, A. K. Chavali, and J. A. Papin. The application of flux balance analysis in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2(3):372–382, 2010.
- [113] D. Singh, R. Carlson, D. Fell, and M. Poolman. Modelling metabolism of the diatom *Phaeodactylum tricornutum*. *Biochemical Society Transactions*, 43(6):1182–1186, 2015.
- [114] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 27:29–34, 1999.
- [115] M. Kanehisa and S. Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.
- [116] R. Caspi, T. Altman, R. Billington, K. Dreher, H. Foerster, C. A. Fulcher, T. A. Holland, I. M. Keseler, A. Kothari, A. Kubo, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, D. Weerasinghe, P. Zhang, and P. D. Karp. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, 42(Database issue):D459–D471, 2014.
- [117] R. Caspi, R. Billington, L. Ferrer, H. Foerster, C. A. Fulcher, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, and P. D. Karp. The MetaCyc database of metabolic pathways and enzymes and the BioCyc

- collection of pathway/genome databases. *Nucleic Acids Research*, 44 (Database issue):D471–480, 2016.
- [118] P. D. Karp, M. Latendresse, and R. Caspi. The Pathway Tools Pathway Prediction Algorithm. *Stand Genomic Sciences*, 5(3):424–429, 2011.
- [119] P. D. Karp, M. Riley, M. Saier, I. T. Paulsen, S. M. Paley, and A. Pellegrini-Toole. The EcoCyc and MetaCyc databases. *Nucleic Acids Research*, 28(1):56–59, 2000.
- [120] R. Caspi, H. Foerster, C. A. Fulcher, R. Hopkinson, J. Ingraham, P. Kaipa, M. Krummenacker, S. Paley, J. Pick, S. Y. Rhee, C. Tissier, P. Zhang, and P. D. Karp. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res*, 34(Database issue):D511–D516, 2006.
- [121] M. G. Poolman, B. K. Bonde, A. Gevorgyan, H. H. Patel, and D. A. Fell. Challenges to be faced in the reconstruction of metabolic networks from public databases. *IEE Proceedings Systems Biology*, 153(5):379–384, 2006.
- [122] K. A. Hunt, J. P. Folsom, R. L. Taffs, and R. P. Carlson. Complete enumeration of elementary flux modes through scalable demand-based subnetwork definition. *Bioinformatics*, 30(11):1569–78, 2014.
- [123] A. Gevorgyan, M. G. Poolman, and D. A. Fell. Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics*, 24(19):2245–2251, 2008.
- [124] D. Garfinkel. A Machine-Independent Language for the Simulation of Complex Chemical and Biochemical Systems. *Computers and Biomedical Research*, 2:31–44, 1968.
- [125] D. J. M. Park and B. E. Wright. Metasim, a General Purpose Metabolic Simulator for Studying Cellular Transformations. *Computer Programs in Biomedicine*, 3(1):10–26, 1973.

- [126] H. M. Sauro. SCAMP: a general-purpose metabolic simulator and metabolic control analysis program. *Computer Applications in the Biosciences*, 9(4):441–450, 1993.
- [127] E. M. Chance, A. R. Curtis, I. P. Jones, and C. R. Kirby. *FACSIMILE: a Computer Program for Flow and Chemistry Simulation, and General Initial Value Problems*. Harwell, 1977.
- [128] P. Mendes. Gepasi - a software package for modeling the dynamics, steady-states and control of biochemical and other systems. *Computer Applications in the Biosciences*, 9(5):563–571, 1993.
- [129] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI - a COmplex PATHway Simulator. *Bioinformatics*, 22(24):3067–3074, 2006.
- [130] S. Klamt, J. Saez-Rodriguez, and E. D. Gilles. Structural and functional analysis of cellular networks with CellNetAnalyze. *BMC Systems Biology*, 1(1):2, 2007.
- [131] J. Schellenberger, R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, and B. O. Palsson. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, 6(9):1290–1307, 2011.
- [132] M. G. Poolman. ScrumPy: metabolic modelling with Python. *IEE Proceedings - Systems Biology*, 153(5):375–378, 2006.
- [133] M. G. Poolman, D. A. Fell, and S. Thomas. Modelling photosynthesis and its control. *Journal of Experimental Botany*, 51(suppl_1):319–328, 2000.
- [134] M. G. Poolman, H. Ölçer, J. C. Lloyd, C. A. Raines, and D. A. Fell. Computer modelling and experimental evidence for two steady states in the photosynthetic Calvin cycle. *European Journal of Biochemistry*, 268(10):2810–2816, 2001.

- [135] H. Hartman. *Genome-scale metabolic modelling of Salmonella and Lactobacillus species*. PhD thesis, Oxford Brookes University, 2013.
- [136] M. Lutz and D. Ascher. *Learning Python*. O’Reilly and Associates, 1st edition, 1999.
- [137] M. Lutz. *Programming Python*. O’Reilly and Associates, 2nd edition, 2001.
- [138] V. Villanova. *Identification des mécanismes de la mixotrophie en Phaeodactylum tricornutum*. PhD thesis, University Grenoble Alpes, 2016.
- [139] P. C. Hinkle. P/O ratios of mitochondrial oxidative phosphorylation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1706(1):1–11, 2005.
- [140] A. M. J. Kliphuis, A. J. Klok, D. E. Martens, P. P. Lamers, M. Janssen, and R. H. Wijffels. Metabolic modeling of *Chlamydomonas reinhardtii*: energy requirements for photoautotrophic growth and maintenance. *Journal of Applied Phycology*, 24(2):253–266, 2012.
- [141] R. J. Geider, B. A. Osborne, and J. A. Raven. Growth, Photosynthesis and Maintenance metabolic cost in the diatom *Phaeodactylum tricornutum* at very low light levels. *Journal of Phycology*, 22(1):39–48, 1986.
- [142] M. D. Bennet, I. J. Leitch, H. J. Price, and J. S. Johnston. Comparisons with *Caenorhabditis* (100 Mb) and *Drosophila* (175 Mb) Using Flow Cytometry Show Genome Size in Arabidopsis to be 157 Mb and thus 25% Larger than the Arabidopsis Genome Initiative Estimate of 125 Mb. *Annals of Botany*, 91(5):547–557, 2005.
- [143] N. A. Eckardt. Sequencing the rice genome. *The Plant Cell*, 12(11): 2011–2018, 2000.

- [144] R. L. Chang, L. Ghamsari, A. Manichaikul, E. F. Y. Hom, S. Balaji, W. Fu, Y. Shen, T. Hao, B. O. Palsson, K. Salehi-Ashtiani, and J. A. Papin. Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Molecular Systems Biology*, 7(1):518–531, 2011.
- [145] J. Nogales, S. Gudmundsson, E. M. Knight, B. O. Palsson, and I. Thiele. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 109:2678–2683, 2012.
- [146] H. Knoop, M. Gründel, Y. Zilliges, R. Lehmann, S. Hoffmann, W. Lockau, and R. Steuer. Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of *Synechocystis* sp. PCC 6803. *PLoS Computational Biology*, 9(6):1–15, 2013.
- [147] B. A. Osborne and R. J. Geider. Photon requirement for growth of the diatom *Phaeodactylum tricornutum* Bohlin (Bacillariophyceae). *Plant, Cell and Environment*, 10(2):141–149, 1987.
- [148] H. A. Barker. Photosynthesis in diatoms. *Archives of Microbiology*, 6(1):141–156, 1935.
- [149] J. Myers. On the Algae: Thoughts about Physiology and Measurements of Efficiency. *Environmental Science Research*, 19:1–16, 1980.
- [150] A. M. Johnston and J. A. Raven. Effect of aeration rates on growth rates and natural abundance $^{13}\text{C}/^{12}\text{C}$ ratio of *Phaeodactylum tricornutum*. *Marine Ecology Progress Series*, 87(3):295–300, 1992.
- [151] A. Contreras, F. C. García, E. Molina, and J. C. Merchuk. Interaction between CO₂-mass transfer, light availability, and hydrodynamic stress in the growth of *Phaeodactylum tricornutum* in a concentric tube airlift photobioreactor. *Biotechnology and Bioengineering*, 60(3):317–325, 1998.

- [152] I. Kudo, M. Miyamoto, Y. Noiri, and Y. Maita. Combined effects of temperature and iron on the growth and physiology of the marine diatom *Phaeodactylum tricorneratum* (Bacillariophyceae). *Journal of Phycology*, 36(6):1096–1102, 2000.
- [153] Z. K. Yang, Y. F. Niu, Y. H. Ma, J. Xue, M. H. Zhang, W. D. Yang, J. S. Liu, S. H. Lu, Y. Guan, and H. Y. Li. Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. *Biotechnology for Biofuels*, 6(1):67–81, 2013.
- [154] D. Heineke, B. Riens, H. Grosse, P. Hoferichter, U. Peter, U.-I. Flügge, and H. W. Heldt. Redox Transfer across the Inner Chloroplast Envelope Membrane. *Plant Physiology*, 95(4):1131–1137, 1991.
- [155] C. Giersch. Capacity of the malate/oxaloacetate shuttle for transfer of reducing equivalents across the envelope of leaf chloroplasts. *Archives of Biochemistry and Biophysics*, 219(2):379–387, 1982.
- [156] M. Linka, J. Aziz, and P. W. Andreas. Functional Characterization of the Plastidic Phosphate Translocator Gene Family from the Thermo-Acidophilic Red Alga *Galdieria Sulphuraria* Reveals Specific Adaptations of Primary Carbon Partitioning in Green Plants and Red Algae. *Plant Physiology*, 148(3):1487–1496, 2008.
- [157] L. J. Sweetlove, K. F. Beard, A. Nunes-Nesi, A. R. Fernie, and G. R. Ratcliffe. Not just a circle: flux modes in the plant TCA cycle. *Trends in Plant Science*, 15(8):462–470, 2010.
- [158] J. Prihoda, A. Tanaka, W. B. M. de Paula, J. F. Allen, L. Tirichine, and C. Bowler. Chloroplast-mitochondria cross-talk in diatoms. *Journal of Experimental Botany*, 63(4):1543–1557, 2012.
- [159] H. Abida, L. J. Dolch, C. Mei, V. Villanova, M. Conte, M. A. Block, G. Finazzi, O. Bastien, Tirichine, C. Bowler, F. Rebeille, D. Petroutsos, J. Jouhet, and E. Marechal. Membrane glycerolipid remodeling triggered by nitrogen and phosphorus starvation in *Phaeodactylum tricorneratum*. *Plant Physiology*, 167(1):118–136, 2015.

- [160] H. A. Krebs, R. Hems, P. Lund, D. Halliday, and W. W. C. Read. Sources of ammonia for mammalian urea synthesis. *Biochemical Journal*, 176:733–737, 1978.
- [161] K. L. Manchester. Glutamate dehydrogenase: A reappraisal. *Biochemical Education*, 13(3):131–133, 1985.
- [162] T. Tercé-Laforgue, F. Dubois, S. Ferrario-Méry, M.-A. Pou de Crezenzo, R. Sangwan, and B. Hirel. Glutamate Dehydrogenase of Tobacco Is Mainly Induced in the Cytosol of Phloem Companion Cells When Ammonia Is Provided Either Externally or Released during Photorespiration. *Plant Physiology*, 136(4):4308–4317, 2004.
- [163] M. S. P. Mtolera. Some properties of glutamate dehydrogenase from the marine red alga *Gracilaria sordida* (Harv.) W. Nelson. *Western Indian Ocean Journal of Marine Science*, 2(2):179–186, 2003.
- [164] F. Dubois, T. Tercé-Laforgue, M. B. Gonzalez-Moro, J. M. Estavillo, R. Sangwan, A. Gallais, and B. Hirel. Glutamate dehydrogenase in plants: is there a new story for an old enzyme? *Plant Physiology and Biochemistry*, 41(6–7):565–576, 2003.
- [165] O. E. Bläsing, Y. Gibon, M. Günther, M. Höhne, R. Morcuende, D. Osuna, O. Thimm, B. Usadel, W. R. Scheible, and M. Stitt. Sugars and Circadian Regulation Make Major Contributions to the Global Regulation of Diurnal Gene Expression in Arabidopsis. *The Plant Cell*, 17(12):3257–3281, 2005.
- [166] C. Wilhelm and D. Selmar. Energy dissipation is an essential mechanism to sustain the viability of plants: The physiological limits of improved photosynthesis. *Journal of Plant Physiology*, 168(2):79–87, 2011.
- [167] J. P. Zehr and P. G. Falkowski. Pathway of ammonium assimilation in a marine diatom determined with the radiotracer ^{13}N . *Journal of Phycology*, 24(4):588–591, 1988.

- [168] M. G. Redinbaugh and W. H. Campbell. Glutamine synthetase and ferredoxin dependent glutamate synthase expression in the maize (*Zea mays*) root primary response to nitrate. Evidence for an organ specific response. *Plant Physiology*, 101(4):1249–1255, 1993.
- [169] K. A. Loulakakis and K. A. Roubelakis-Angelakis. Molecular cloning and characterization of cDNAs coding for ferredoxindependent glutamate synthase from *Vitis vinifera*. *Physiologia Plantarum*, 101(1):220–228, 1997.
- [170] M. F. Suárez, C. Avila, F. Gallardo, F. R. Cantón, A. García Gutiérrez, M. G. Claros, and F. M. Cánovas. Molecular and enzymatic analysis of ammonium assimilation in woody plants. *Journal of Experimental Botany*, 53(370):891, 2002.
- [171] M. Haimovich-Dayana, N. Garfinkel, D. Ewe, Y. Marcus, A. Gruber, H. Wagner, P. G. Kroth, and A. Kaplan. The role of C4 metabolism in the marine diatom *Phaeodactylum tricornutum*. *New Phytologist*, 197(1):177–185, 2013.
- [172] P. Dzeja and T. Andre. Adenylate Kinase and AMP Signaling Networks: Metabolic Monitoring, Signal Communication and Body Energy Sensing. *International Journal of Molecular Sciences*, 10(4):1729–1772, 2009.
- [173] C. H. Foyer and B. Halliwell. The presence of glutathione and glutathione reductase in chloroplasts: A proposed role in ascorbic acid metabolism. *Planta*, 133(1):21–25, 1976.
- [174] C. H. Foyer and G. Noctor. Ascorbate and Glutathione: The Heart of the Redox Hub. *Plant Physiology*, 155(1):2–18, 2011.
- [175] A. Schumann, R. Goss, T. Jakob, and C. Wilhelm. Investigation of the quenching efficiency of diatoxanthin in cells of *Phaeodactylum tricornutum* (Bacillariophyceae) with different pool sizes of xanthophyll cycle pigments. *Phycologia*, 46(1):113–117, 2007.

- [176] P. D. Tortell. Evolutionary and ecological perspectives on carbon acquisition in phytoplankton. *Limnology and Oceanography*, 45(3):744–750, 2000.
- [177] J. S. Paul and B. E. Volcani. Photorespiration in diatoms. *Archives of Microbiology*, 110(2):247–252, 1976.
- [178] C. Xu, Z. Jiang, and B. Huang. Nitrogen Deficiency-induced Protein Changes in Immature and Mature Leaves of Creeping Bentgrass. *JASHS*, 136(6):399–407, 2011.
- [179] J. Soutourina, S. Blanquet, and P. Plateau. Role of D-cysteine desulhydrase in the adaptation of *Escherichia coli* to D-cysteine. *Journal of Biological Chemistry*, 276(44):40864–40872, 2001.
- [180] A. Riemenschneider, R. Wegele, A. Schmidt, and J. Papenbrock. Isolation and characterization of a D-cysteine desulhydrase protein from *Arabidopsis thaliana*. *FEBS Journal*, 272(5):1291–1304, 2005.
- [181] D. M. Orcutt and G. W. Patterson. Effect of light intensity upon lipid composition of *Nitzschia closterium* (*Cylindrotheca fusiformis*). *Lipids*, 9(12):1000–1003, 1974.
- [182] A. Sukenik, Y. Carmeli, and T. Berner. Regulation of fatty acid composition by irradiance level in the eustigmatophyte *nannochloropsis* sp. *Journal of Phycology*, 25(4):686–692, 1989.
- [183] M. R. Brown, G. A. Dunstan, S. J. Norwood, and K. A. Miller. Effects of harvest stage and light on the biochemical composition of the diatom *Thalassiosira pseudonana*. *Journal of Phycology*, 32:64–73, 1996.
- [184] S. V. Khotimchenko and I. M. Yakovleva. Lipid composition of the red alga *Tichocarpus crinitus* exposed to different levels of photon irradiance. *Phytochemistry*, 66(1):73–79, 2005.
- [185] B. W. de Jong, S. Shi, V. Siewers, and J. Nielsen. Improved production of fatty acid ethyl esters in *Saccharomyces cerevisiae* through

- up-regulation of the ethanol degradation pathway and expression of the heterologous phosphoketolase pathway. *Microbial Cell Factories*, 13(1):13–39, 2014.
- [186] G. Panagiotou, M. R. Andersen, T. Grotkjae, T. B. Regueira, G. Hofmann, J. Nielsen, and L. Olsson. Systems Analysis Unfolds the Relationship between the Phosphoketolase Pathway and Growth in *Aspergillus nidulans*. *PLoS ONE*, 3:1–8, 2008.
- [187] M. Papini, I. Nookaew, V. Siewers, and J. Nielsen. Physiological characterization of recombinant *Saccharomyces cerevisiae* expressing the *Aspergillus nidulans* phosphoketolase pathway: validation of activity through ^{13}C -based metabolic flux analysis. *Applied Microbiology and Biotechnology*, 95(4):1001–1010, 2012.
- [188] K. Kocharin, V. Siewers, and J. Nielsen. Improved polyhydroxybutyrate production by *Saccharomyces cerevisiae* through the use of the phosphoketolase pathway. *Biotechnology and Bioengineering*, 110(8):2216–2224, 2013.
- [189] A. Bar-Even. Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways. *Biochemistry*, 55(28):3851–3863, 2016.
- [190] V. Villanova, A. E. Fortunato, D. Singh, D. D. Bo, M. Conte, T. Obata, J. Jouhet, A. R. Fernie, E. Marechal, A. Falciatore, J. Pagliardini, A. Le Monnier, M. Poolman, G. Curien, D. Petroutsos, and G. Finazzi. Investigating mixotrophic metabolism in the model diatom *Phaeodactylum tricornutum*. *Philosophical Transactions of the Royal Society B*, 372(1728), 2017.
- [191] A. Domínguez-Ferrerías, M. J. Soto, R. Pérez-Arnedo, J. Olivares, and J. Sanjuán. Importance of Trehalose Biosynthesis for *Sinorhizobium meliloti* Osmotolerance and Nodulation of Alfalfa Roots. *Journal of Bacteriology*, 191(24):7490–7499, 2009.

- [192] O. M. Tekolo, J. Mckenzie, A. Botha, and B. A. Prior. The osmotic stress tolerance of basidiomycetous yeasts. *FEMS Yeast Research*, 10(4):482–491, 2010.
- [193] M. Sugawara, E. J. Cytryn, and M. J. Sadowsky. Functional role of *Bradyrhizobium japonicum* trehalose biosynthesis and metabolism genes during physiological stress and nodulation. *Applied and Environmental Microbiology*, 76(4):1071–1081, 2010.
- [194] L. D. Zhu, Z. H. Li, and E. Hiltunen. Strategies for Lipid Production Improvement in Microalgae as a Biodiesel Feedstock. *BioMed Research International*, 2016(1):1–8, 2016.
- [195] A. Norici, A. M. Bazzoni, A. Pugnetti, J. A. Raven, and M. Giordano. Impact of irradiance on the C allocation in the coastal marine diatom *Skeletonema marinoi* Sarno and Zingone. *Plant, Cell and Environment*, 34(10):1666–1677, 2011.
- [196] H. P. Dong, Y. L. Dong, L. Cui, S. Balamurugan, J. Gao, S. H. Lu, and T. Jiang. High light stress triggers distinct proteomic responses in the marine diatom *Thalassiosira pseudonana*. *BMC Genomics*, 17(1):994–1008, 2016.
- [197] Q. KaiXian and M. A. Borowitzka. Light and nitrogen deficiency effects on the growth and composition of *Phaeodactylum tricornutum*. *Applied Biochemistry and Biotechnology*, 38(1):93–103, 1993.
- [198] F. Almada-Calvo. *Effect of temperature, dissolved inorganic carbon and light intensity on the growth rates of two microalgae species in monocultures and co-cultures*. PhD thesis, The University of Texas at Austin, 2014.
- [199] P. G. Roessler. Effects of silicon deficiency on lipid composition and metabolism in the diatom *Cyclotella Cryptica*. *Journal of Phycology*, 24(3):394–400, 1988.

- [200] I. Khozin-Goldberg and Z. Cohen. The effect of phosphate starvation on the lipid and fatty acid composition of the fresh water eustigmatophyte *Monodus subterraneus*. *Phytochemistry*, 67(7):696–701, 2006.
- [201] I. A. Guschina and J. L. Harwood. Lipids and lipid metabolism in eukaryotic algae. *Progress in Lipid Research*, 45(2):160–186, 2006.
- [202] A. E. Solovchenko, I. Khozin-Goldberg, S. Didi-Cohen, Z. Cohen, and M. N. Merzlyak. Effects of light intensity and nitrogen starvation on growth, total fatty acids and arachidonic acid in the green microalga *Parietochloris incisa*. *Journal of Applied Phycology*, 20(3):245–251, 2008.
- [203] G. Breuer, P. P. Lamers, D. E. Martens, R. B. Draaisma, and R. H. Wijffels. The impact of nitrogen starvation on the dynamics of triacylglycerol accumulation in nine microalgae strains. *Bioresource Technology*, 124:217–226, 2012.
- [204] G. Benvenuti, R. Bosma, M. Cuaresma, M. Janssen, M. J. Barbosa, and R. H. Wijffels. Selecting microalgae with high lipid productivity and photosynthetic activity under nitrogen starvation. *Journal of Applied Phycology*, 27(4):1425–1431, 2015.
- [205] R. D. Gardner, K. E. Cooksey, F. Mus, R. Macur, K. Moll, E. Eustance, R. P. Carlson, R. Gerlach, M. W. Fields, and B. M. Peyton. Use of sodium bicarbonate to stimulate triacylglycerol accumulation in the chlorophyte *Scenedesmus* sp. and the diatom *Phaeodactylum tri-cornutum*. *Journal of Applied Phycology*, 24:1311–1320, 2012.
- [206] R. D. Gardner, E. J. Lohman, K. E. Cooksey, R. Gerlach, and B. M. Peyton. Cellular Cycling, Carbon Utilization, and Photosynthetic Oxygen Production during Bicarbonate-Induced Triacylglycerol Accumulation in a *Scenedesmus* sp. *Energies*, 6:6060–6076, 2013.
- [207] F. Guihéneuf and D. B. Stengel. LC-PUFA-Enriched Oil Production by Microalgae: Accumulation of Lipid and Triacylglycerols Containing n-3 LC-PUFA Is Triggered by Nitrogen Limitation and Inorganic Carbon

- Availability in the Marine Haptophyte *Pavlova lutheri*. *Marine Drugs*, 11(11):4246–4266., 2013.
- [208] J. P. Heiden, K. Bischof, and S. Trimborn. Light Intensity Modulates the Response of Two Antarctic Diatom Species to Ocean Acidification. *Frontiers in Marine Science*, 3:260–277, 2016.
- [209] B. A. Templeton and M. A. Savageau. Transport of Biosynthetic Intermediates: Homoserine and Threonine Uptake in *Escherichia coli*. *Journal of Bacteriology*, 117(3):1002–1009, 1974.
- [210] E. Grafahrend-Belau, F. Schreiber, D. Koschützki, and B. H. Junker. Flux Balance Analysis of Barley Seeds: A Computational Approach to Study Systemic Properties of Central Metabolism. *Plant Physiology*, 149(1):585–598, 2009.
- [211] C. Y. M. Cheung. *Genome Scale Metabolic Models of Plant Tissues*. PhD thesis, Oxford University, 2013.
- [212] M. W. Covert, C. H. Schilling, and B. O. Palsson. Regulation of Gene Expression in Flux Balance Models of Metabolism. *Journal of Theoretical Biology*, 213(1):73–88, 2001.
- [213] S. Chandrasekaran and N. D. Price. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, 107(41):17845–17850, 2010.
- [214] C. Colijn, A. Brandes, J. Zucker, D. S. Lun, B. Weiner, M. R. Farhat, T.-Y. Cheng, B. D. Moody, M. Murray, and J. E. Galagan. Interpreting Expression Data with Metabolic Flux Models: Predicting *Mycobacterium tuberculosis* Mycolic Acid Production. *PLOS Computational Biology*, 5(8):1–14, 2009.
- [215] S. A. Becker and P. O. Bernhard. Context-Specific Metabolic Networks Are Consistent with Experiments. *PLOS Computational Biology*, 4(5): 1–10, 2008.

- [216] R. J. P. van Berlo, D. de Ridder, J. M. Daran, P. A. S. Daran-Lapujade, B. Teusink, and M. J. T. Reinders. Predicting Metabolic Fluxes Using Gene Expression Differences As Constraints. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(1):206–216, 2011.
- [217] N. Töpfer, S. Kleessen, and Z. Nikoloski. Integration of metabolomics data into metabolic networks. *Frontiers in Plant Science*, 6:49–62, 2015.
- [218] L. Jerby, T. Shlomi, and E. Ruppin. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Molecular Systems Biology*, 6(1):401–410, 2010.
- [219] B. J. Schmidt, A. Ebrahim, T. O. Metz, J. N. Adkins, B. O. Pals-son, and D. R. Hyduke. GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics*, 29(22):2900–2908, 2013.
- [220] A. Hoppe, S. Hoffmann, and H.-G. Holzhütter. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Systems Biology*, 1(1):23–35, 2007.
- [221] M. M. Reboloso-Fuentes, A. Navarro-Pérez, J. J. Ramos-Miras, and J. L. Guil-Guerrero. Biomass nutrient profiles of the microalga *Phaeodactylum tricornutum*. *Journal of Food Biochemistry*, 25(1):57–76, 2001.

Appendix A

Models and Additional materials

Content of the attached CD.

A.1 Model Construction

Method to generate GSM from PGDB database or given list of reactions is in python file: BuildOrg.py. List of reactions for specific compartments is present in python file: Compartments.py. Directories Model and PhaeoModelFromDiatomCyc consists of Model I and Model II respectively in ScrumPy format. Phaeo.spy file is the main module to be called while loading the model object. It imports other modules spread in mutiple files as described in section 3.2.1. Sub-directory Tools consist of python module to generate general LP object (python file: BuildLP.py) and biomass composition in experimentally observed proportion to constraint flux in biomass transporters.

A.1.1 Deleted reactions and metabolites

List of deleted reactions and metabolites. Python file (Unwanted.py).

A.1.2 Corrected reactions and metabolites

List of reactions corrected for reversibility and directionality, and renamed metabolites to avoid naming inconsistencies. Python file (Substitutes.py).

A.2 Python Modules for Model Analysis

A.2.1 ModelProperty

Methods to validate the models for energy and redox conservation as described in section 3.2.3 and general model properties such as number of reactions, metabolites, gene association. Python file (ModelProperty.py)

A.2.2 LightScan

Methods to generate LP and perform Lightscan analysis. Python file (LightScanLP.py and LightScan.py respectively)

A.2.3 LipidScan

Methods to generate LP and perform Lipidscan analysis. Python file (LipidScanLP.py and LipidScan.py respectively)

A.2.4 MixoScan

Methods to generate LP and perform glycerol scan analysis. Python file (MixoScanLP.py and MixoScan.py respectively)

Appendix B

Experimental Data

All the experimental data in this section has been provided by the industrial/experimental partners of AccliPhot, Fermentalg and CNRS.

B.1 Electron Transfer Rate

With each photon absorbed by the reaction center one electron is released into the electron transport chain. Therefore, the electron transfer rate (ETR) is also used as a measure to get an estimate of absorbed light. The ETR value for three different incident light intensities (75, 170, 340 $\mu\text{E m}^{-2} \text{s}^{-1}$), in mixotrophic and phototrophic condition, were measured by our experimental partner, CNRS. This provided ETR measure was then extrapolated to zero get an estimate of absorbed light for range of incident light intensity, as shown in figure B.1, used in the experimental study.

B.2 Estimation of Growth Rate

Phototrophic axenic cultures of *P. tricornutum* were grown in 250 ml flask, shaking at 100 rpm, in artificial seawater ESAW at 20° C. The cultures, volume of 0.05 litre, were continuously illuminated with the light intensity of 40 $\mu\text{mol Photon m}^{-2} \text{s}^{-1}$. The illuminated surface area was roughly estimated to be 0.017 m^2 , based on the dimension of the flask and height of culture. The

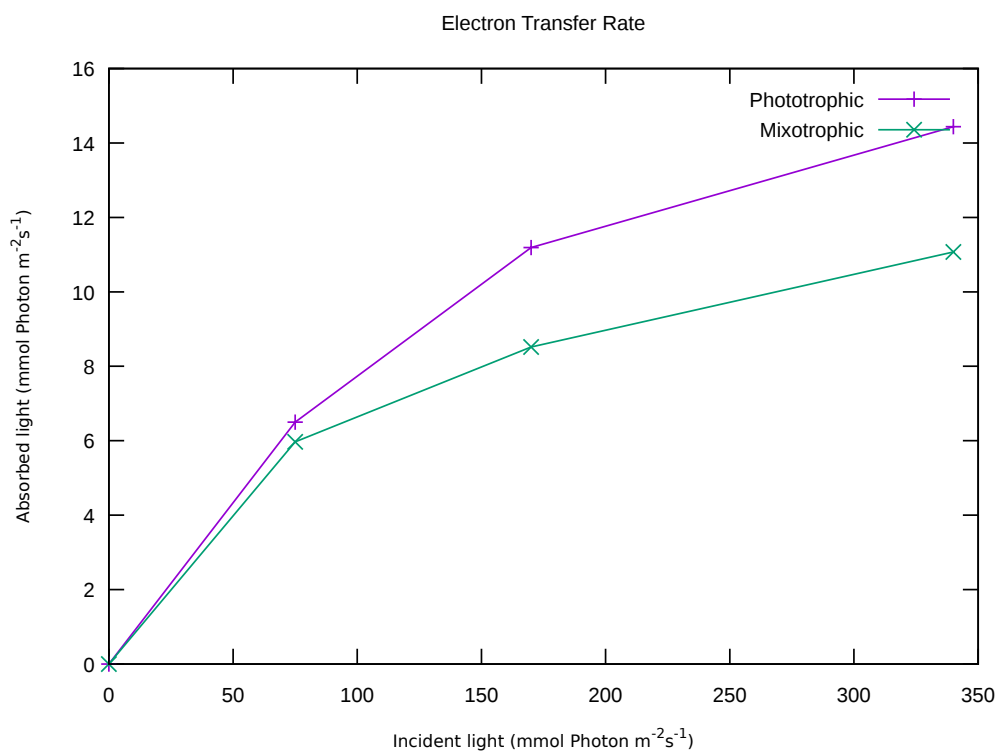


Figure B.1: Electron Transfer Rate data provided by CNRS, France is denoted by ‘*’. This data was extrapolated to zero in order to obtain absorbed photon flux at lower incident light intensity.

biomass concentration at the end of 10 days was 0.33 gDW/l. The growth rate of the culture was estimated to be 0.01 hr^{-1} , as shown in figure B.2, based on the cell count.

B.3 Estimation of Light Absorbed and Glycerol Consumed

P. tricornutum culture was grown in same condition as described in section B.2, i.e phototrophic condition, except that the culture was exposed to the light intensity of $60 \mu\text{mol Photon m}^{-2} \text{ s}^{-1}$. Under similar condition, another culture of *P. tricornutum* was grown and 4.5 g/l of glycerol was added to the

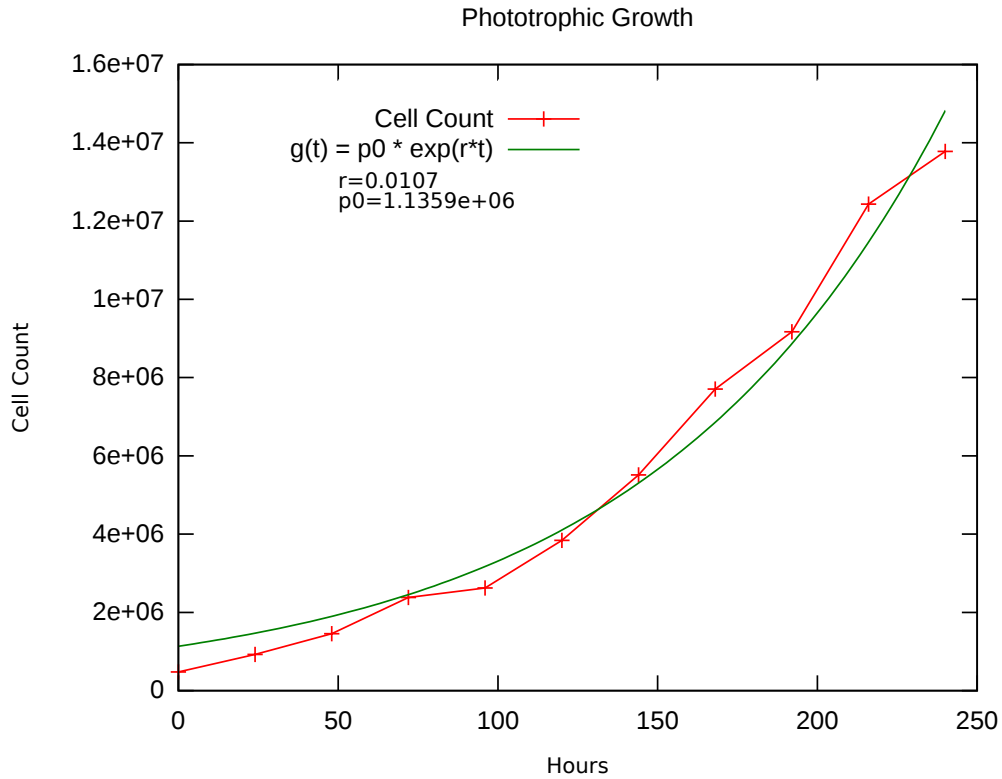


Figure B.2: The cell count data fitted to the exponential function to obtain the growth rate.

medium, i.e mixotrophic condition. The biomass concentration, at this point, was 0.57 gDW/l and 0.53 gDW/l for mixotrophic and phototrophic culture respectively. These experimental parameters were used to get an estimate of light absorbed per amount of biomass based on eq.3.2.2 and was 10 mmol Photon gDW⁻¹ hr⁻¹ and 12 mmol Photon gDW⁻¹ hr⁻¹ for mixotrophic and phototrophic conditions respectively.

The effect of glycerol on growth was visible at 120 hr. At this time, 0.3 g/l glycerol was consumed by culture (transcriptomics and metabolomics analysis in Villanova [138]) and glycerol consumption rate per amount of biomass was estimated to be 0.1 mmol gDW⁻¹ hr⁻¹ as follows:

$$\text{glycerol consumption} = \frac{\text{glycerol consumed} \times \text{growth rate}}{\text{molecular weight} \times \text{biomass concentration}}$$

Appendix C

Biomass Composition

Table C.1: Biomass composition of *P.tricornutum* [122]. The growth rate of 0.01 hr^{-1} , as estimated in B.2, was multiplied to this biomass composition and used as biomass constraints to the FBA problem.

Biomass Components	Phototrophic (mmol gDW ⁻¹)	Biomass Components	Phototrophic (mmol gDW ⁻¹)
Alanine	0.3812	Ascorbate	0.0152
Arginine	0.1414	Riboflavin	0.0001
Aspartic Acid	0.1785	Niacinamide	0.0059
Cysteine	0.0528	Tocopherol	9.287e-05
Glutamic acid	0.1668	Chlorophyll	0.0336
Glycine	0.3407	DAMP	0.0524
Histidine	0.0644	DCMP	0.0663
Isoleucine	0.1371	DGMP	0.0587
Leucine	0.2863	GMP	0.0561
Lysine	0.1249	UMP	0.0535
Methionine	0.0616	CMP	0.0630
Phenylalanine	0.0895	TMP	0.0539
Proline	0.1805	AMP	0.0499
Serine	0.3194	Chrysolaminaran	0.8881
Threonine	0.2046	TAG	0.1709
Tyrosine	0.0573	Glutamine	0.1121
Valine	0.2319	Asparagine	0.1149
Tryptophan	0.0274		

Table C.2: Biomass composition of *P. tricornutum*, provided by Valeria Villanova, for phototrophic and mixotrophic conditions [190]. The growth rate of 0.019 hr^{-1} , as in section B.3, was multiplied to the biomass compositions and used as biomass constraints to the FBA problem. Note: here zero represents value that were below threshold and not determined in this experimental analysis. The biomass component that were not measured were regarded to be same as in Hunt et al. [122] or Reboloso-Fuentes et al. [221].

Biomass Components	Mixotrophic (mmol gDW ⁻¹)	Phototrophic (mmol gDW ⁻¹)	Biomass Components	Mixotrophic (mmol gDW ⁻¹)	Phototrophic (mmol gDW ⁻¹)
Alanine	0.3132	0.3194	TAG	0.0421	0.0283
Arginine	0.1289	0.1137	AMP	0.0067	0.0067[221]
Aspartic Acid	0.2997	0.2875	TMP	0.0072	0.0072[221]
Cysteine	0.0	0.0	CMP	0.0084	0.0084[221]
Glutamic acid	0.2989	0.3355	UMP	0.0071	0.0071[221]
Glycine	0.2613	0.2871	GMP	0.0075	0.0075[221]
Histidine	0.0410	0.0438	DGMP	0.0078	0.0078[221]
Isoleucine	0.1251	0.1405	DCMP	0.0088	0.0088[221]
Leucine	0.2178	0.2447	DAMP	0.0070	0.0070[221]
Lysine	0.1280	0.1268	Tocopherol	9.287e-05	9.287e-05[122]
Methionine	0.0680	0.0410	Niacinamide	0.0059	0.0059[122]
Phenylalanine	0.1100	0.1222	Ribiflavin	0.0001	0.0001[122]
Proline	0.1253	0.1621	Ascorbate	0.0152	0.0152[122]
Serine	0.1751	0.1896	Chrysolaminaran	0.6744	0.4357
Threonine	0.1545	0.1641	Chlorophyll	0.0067	0.0112
Tyrosine	0.0	0.0	Palmitate	0.7722	0.5197
Valine	0.1732	0.1895	Trehalose	0.1775	0.1147
Tryptophan	0.0	0.0	Mannitol	0.1667	0.1077
Asparagine	0.0	0.0	Carotene	0.0056	0.0074
Glutamine	0.0	0.0	Mannose	0.0503	0.0325

Appendix D

Published Papers

Dipali Singh, Ross Carlson, David Fell and Mark Poolman, 'Modelling metabolism of the diatom *Phaeodactylum tricornutum*', *Biochemical Society Transactions* (2015) Volume 43, part 6
<http://www.biochemsoctrans.org/content/43/6/1182.article-info>
<https://doi.org/10.1042/BST20150152>

Valeria Villanova, Antonio Emidio Fortunato, Dipali Singh, Davide Dal Bo, Melissa Conte, Toshihiro Obata, Juliette Jouhet, Alisdair R. Fernie, Eric Marechal, Angela Falciatore, Julien Pagliardini, Adeline Le Monnier, Mark Poolman, Gilles Curien, Dimitris Petroustos and Giovanni Finazzi, '*Investigating mixotrophic metabolism in the model diatom Phaeodactylum tricornutum*', *Phil. Trans. R. Soc. B* 372: 20160404. <http://dx.doi.org/10.1098/rstb.2016.0404>