# Genome Scale Metabolic Modelling of *Arabidopsis thaliana* and *Chlamydomonas reinhardtii*

**Kailash Adhikari**

A thesis submitted in partial fulfillment 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

**June 2017**

# Abstract

Recent advances in genome sequencing technology have enabled the elucidation of complete genome sequence for plants and algae. Genome scale metabolic models constructed from genome annotations represent the entire metabolic characteristics of the organism and can be used to integrate other data and study metabolic capabilities of the organisms under different conditions. This framework can help develop new insight about operating characteristic of the organism and propose new hypotheses that can be tested experimentally.

In order to advance our understanding of photosynthetic metabolism in plants and algae, GSMs of *A. thaliana* and *C. reinhardtii* have been constructed using annotations from their respective BioCyc databases. They satisfy all theoretical considerations and are able to represent known biological behaviors and thus can be used in subsequent investigations.

In collaboration with experimental partners, the arabidopsis model was used to study the knock-out phenotype of the Calvin Cycle enzymes. This correctly predicted the viability of single knockouts of 4 Calvin cycle enzymes. Alternate metabolic routes that make such change possible were identified using flux balance analysis. The analysis demonstrated a complementary role of SBPase and FBPase in the Calvin cycle and further proposed a novel role of transaldolase in daytime metabolism under knockout conditions.

Both models, were used to investigate likely coordinated changes in metabolic networks to dissipate excess energy under high light conditions. Methods that use correlation coefficient and mixed integer linear programming have been developed for this purpose. Proteomics data obtained under high light conditions was integrated in the model to propose energy dissipating modes that are more likely to occur *in vivo*. Further, removal of reactions involved in energy dissipation mechanisms showed improve biomass yield.

# Acknowledgements

Last few years have been a great learning experience for me not only in the scientific arena, but also on a personal level. I am thankful to everyone who has made this possible. First and foremost my deepest gratitude goes to my supervisors Dr. Mark Poolman and Prof. David Fell for their endless support, encouragement, guidance and teaching throughout my PhD.

I would like to thank all members of the Cell Systems Modelling Group, Dr. Hassan Hartman, Dipali Singh, Noah Mesfin and Elzbieta Al-Saidi for always being there for discussions and help when needed. I am also thankful to all the visiting members of the group Dr. Agris Pentjuss, Huili Yuan, Rupert Norman, Sisse Mortensen, Andreas and Maite for sharing a joyful times in the lab. I sincerely appreciate Dr. Nicole Pearcy for sharing her experience and advises on thesis writing.

I am very thankful to all the members of the AccliPhot consortium for great scientific collaboration and educational meetings. I am especially grateful to Prof. Samuel Zeeman and Martina Zenella for valuable experimental support for the knockout investigation. I am grateful to Prof. Guillaume Conge and Brieue Urbain for sharing experimental data and helping me use them in the models. I am also deeply thankful to Dr. Chris Baker for providing me the proteomics data. I also acknowledge the financial support from the AccliPhot Marie Curie Fellowship.

I would also like to thank the Oxford Brookes University and the staff of the Department of Biological and Medical Sciences for their academic and administrative support. In particular, I would like to acknowledge Prof. John Runions, Prof.Chris Hawes, Prof.Susan Brookes, Prof.David Evans, Jill Organ and Philip Voysey.

Finally, I am very grateful to my dearest parents and caring family members for their boundless love and support over the years.

# Contents

**Abstract** **i**

**Acknowledgements** **ii**

**Contents** **ii**

**Nomenclature** **viii**

URLs . . . . . viii

Abbreviations . . . . . viii

**1 Introduction to Plant Metabolism** **1**

1.1 Project Background . . . . . 1

1.2 Motivation . . . . . 2

1.3 Introduction of organisms . . . . . 4

1.4 Structure of a plant cell . . . . . 5

1.4.1 Photosynthesis : Chloroplastic metabolism . . . . . 7

1.4.2 Mitochondrial metabolism . . . . . 14

1.4.3 Cytosolic metabolism . . . . . 18

1.4.4 Interaction between compartments . . . . . 21

1.4.5 Discussion . . . . . 27

1.5 Aims and Structure of the thesis . . . . . 28

**2 Mathematical modelling of metabolism - General Methodology** **29**

2.1 Introduction . . . . . 29

2.1.1 Fundamental concepts . . . . . 30

2.1.2 The steady-state assumption . . . 31
2.1.3 Structural Modelling . . . 33
2.1.4 Null Space Analysis . . . 33
2.1.5 Elementary Modes Analysis . . . 37
2.1.6 Linear Programming/Flux Balance Analysis . . . 42
2.1.7 Mixed integer linear programming . . . 44
2.1.8 Integration of proteomics data . . . 45
2.2 Metabolic Modelling Tools . . . 47
2.2.1 Python Programming Language . . . 47
2.2.2 ScrumPy . . . 48

**3 Construction of genome scale metabolic models 52**
3.1 Introduction . . . 52
3.1.1 Approaches for construction of models . . . 53
3.1.2 Inconsistencies in models . . . 53
3.1.3 GSMs of *A. thaliana* and *C. reinhardtii* . . . 55
3.2 Methods - Construction and Validation of GSM . . . 56
3.2.1 Resources for genome annotations . . . 56
3.2.2 Manually added reactions . . . 60
3.2.3 Construction of GSMs . . . 62
3.2.4 Handling badly defined metabolites and reactions . . . 65
3.2.5 Definition of inputs and outputs . . . 66
3.2.6 Atomic Balance . . . 67
3.2.7 Identifying energetically inconsistent energy producing cycles . . . 67
3.2.8 Identifying stoichiometric inconsistencies . . . 68
3.2.9 Identifying missing reactions . . . 69
3.2.10 Ability of the model to account for growth . . . 69
3.3 Results . . . 70
3.3.1 GSM of *A. thaliana* . . . 70
3.3.2 GSM of *C. reinhardtii* . . . 74
3.4 Determination of biomass composition of Chlamydomonas . . . 75
3.4.1 Method and results . . . 75

3.4.2 Comparison of biomass composition . . . . . 77
3.4.3 Calculating photon stoichiometry . . . . . 78
3.4.4 Calculation of total acetate consumed . . . . . 79
3.5 Discussion and further developments . . . . . 79

**4 Photosynthetic properties of GSMs of *A. thaliana* and *C. reinhardtii* 81**
4.1 Introduction . . . . . 81
4.1.1 Models to study photosynthetic properties . . . . . 81
4.2 Methodology . . . . . 83
4.2.1 General photosynthetic responses . . . . . 83
4.2.2 Responses under changing light conditions - light scan analysis . . . . . 84
4.3 Results from the Arabidopsis model . . . . . 85
4.3.1 Heterotrophic metabolism . . . . . 85
4.3.2 Phototrophic metabolism . . . . . 85
4.3.3 Discussion . . . . . 95
4.4 Results from the GSM of *C. reinhardtii* . . . . . 97
4.4.1 Phototrophic metabolism . . . . . 97
4.4.2 Mixotrophic metabolism . . . . . 100
4.4.3 Discussion . . . . . 101
4.5 Conclusion . . . . . 103

**5 Effects of enzyme knockouts on GSM of *A. thaliana* 105**
5.1 Introduction . . . . . 105
5.2 Method . . . . . 108
5.2.1 Investigating knockout properties with the GSM . . . . 108
5.2.2 Elementary modes analysis . . . . . 109
5.3 Results . . . . . 109
5.3.1 Elementary modes analysis of Calvin cycle to study effect of deregulation . . . . . 119
5.4 Discussion . . . . . 119
5.4.1 Newer properties of the enzymes . . . . . 123

5.4.2 Prediction of phenotype from FBA studies . . . . . 124
5.4.3 Deregulation of SBPase and FBPase could be lethal to plants . . . . . 124
5.4.4 Conclusion . . . . . 125

**6 Energy dissipation mechanisms of Arabidopsis and Chalmydomonas 126**
6.1 Introduction . . . . . 126
6.1.1 Biochemical basis of energy dissipation mechanisms . . 127
6.1.2 Metabolic cycles for energy dissipation . . . . . 128
6.2 Methodology . . . . . 131
6.2.1 Identifying a photon absorbing core model . . . . . 131
6.2.2 Energy dissipating nodes in correlation tree . . . . . 132
6.2.3 Computing energy dissipating modes using MILP . . . 132
6.2.4 Integration of proteomics data . . . . . 133
6.2.5 Light scan analysis . . . . . 134
6.3 Results from GSM of *A. thaliana* . . . . . 135
6.3.1 Energy dissipating nodes - Correlation Tree Analysis . 135
6.3.2 Alternate energy dissipating modes - MILP . . . . . 140
6.3.3 Light Scan Analysis - increase in biomass components . 140
6.4 Results from Chlamydomonas model . . . . . 143
6.4.1 Energy dissipating nodes - Correlation Tree Analysis . 143
6.4.2 Energy dissipating modes - MILP . . . . . 144
6.5 Discussion . . . . . 144
6.5.1 Experimental validation of results with proteomics data 147
6.5.2 Broad applicability of the method . . . . . 149
6.5.3 Possibilities of *in vivo* activity . . . . . 149
6.5.4 Increase in biomass composition after removing reactions involved in energy dissipation . . . . . 150

**7 General summary 152**
7.1 Overview of results obtained . . . . . 152
7.2 Scope of the model and methods developed during this study . 155

7.3 Limitations of the model and future developments . . . . . 156
7.4 Conclusion . . . . . 157

**Appendix A Knockout investigation - experimental work 187**

**Appendix B Energy dissipating reactions 189**

**Appendix C Model files and Python code 196**
C.1 Contents of the attached CD. . . . . 196

# URLs

| | |
|---|---|
| AccliPhot | http://www.accliphot.eu/ |
| BioCyc | http://biocyc.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 |
| SUBA | http://suba.live/ |

# Abbreviations

Table 1: Summary of abbreviations and BioCyc identifiers of reactions and metabolites.

| **Abbreviation** | **Common name** | **BioCyc IDs** |
|---|---|---|
| *Reactions* | | |
| aKGdh | 2-oxoglutarate dehydrogenase | 2OXOGLUTARATEDEH-RXN |
| 2PGH | phosphopyruvate hydratase | 2PGADEHYDRAT-RXN |
| 3PGAR | Phosphoglyceromutase | 3PGAREARR-RXN |
| 3DQdh | dehydroquinate dehydratase | 3-DEHYDROQUINATE-DEHYDRATASE-RXN |
| 6PGLac | 6-phosphogluconolactonase | 6PGLUCONOLACT-RXN |
| 6PF | 6-phosphofructokinase | 6PFRUCTPHOS-RXN |
| ACN | aconitase | ACONITATEDEHYDR-RXN |
| ASPsyn | glutamic aspartic transaminase | ASPAMINOTRANS-RXN |
| ASPTran | aspartate transaminase | ASPAMINOTRANS-RXN |

| | | |
|---|---|---|
| CarbPSyn | carbamoyl phosphate synthase | 6.3.4.16-RXN |
| CarbK | carbamate kinase | CARBAMATE-KINASE-RXN |
| CarboDht | carbonate dehydrtase | RXN0-5224 |
| ChorSyn | chorismate synthase | CHORISMATE-SYNTHASE-RXN |
| CitSyn | citrate synthase | CITSYN-RXN |
| DAHPsyn | DHAP synthase | DAHPSYN-RXN |
| FBPase | fructose-bisphosphatase | F16BDEPHOS-RXN |
| FBPAld | fructose-bisphosphate aldolase | F16ALDOLASE-RXN |
| FUMhyd | fumarate hydratase | FUMHYDR-RXN |
| G1PAt | glucose-1-phosphate adenylyltransferase | GLUC1PADENYLTRANS-RXN |
| G6Pdh | glucose-6-phosphate dehydrogenase | GLU6PDEHYDROG-RXN |
| GAPdh | glyceraldehyde-3-phosphate dehydrogenase | 1.2.1.13-RXN |
| GAPdhC | glyceraldehyde-3-phosphate dehydrogenase(p) | GAPOXNPHOSPHN-RXN |
| GlyPhos | glycogen phosphorylase | RXN0-5184 |
| GltmSyn | glutamine synthetase | GLUTAMINESYN-RXN |
| GluthDh | glutathione dehydrogenase | 1.8.5.1-RXN |
| GluthRd | glutathione reductase | GLUTATHIONE-REDUCT-NADPH-RXN |
| ICdh | isocitrate dehydrogenase(NAD) | ISOCITRATE-DEHYDROGENASE-NAD |
| LacDh | lactaldehyde dehydrogenase | LACTALDDEHYDROG-RXN |
| MALdh | malate dehydrogenase | MALATE-DEH-RXN |
| PEPkin | phosphoenolpyruvate kinase | PEPDEPHOS-RXN |
| PEPCk | phosphoenolpyruvic carboxylase | PEPCARBOX-RXN |
| PGI | glucose-6-phosphate isomerase | PGLUCISOM-RXN |
| PGM | phosphoglucomutase | PHOSPHOGLUCMUT-RXN |
| PGKin | phosphoglycerate kinase | PHOSGLYPHOS-RXN |
| PGluDh | 6-phosphogluconate dehydrogenase | 6PGLUCONDEHYDROG-RXN |
| PRK | phosphoribulokinase | PHOSPHORIBULOKINASE-RXN |
| PPDK | pyruvate orthophosphate dikinase | PYRUVATEORTHOPHOSPHATE-DIKINASE-RXN |
| PPI | inorganic diphosphatase | INORGPYROPHOSPHAT-RXN |

| | | |
|---|---|---|
| PyrShiSyn | pyruvoylshikimate phosphate synthase | 2.5.1.19-RXN |
| PYRdh | pyruvate dehydrogenase | PYRUVDEH-RXN |
| Rib5PEpi | ribulose-phosphate 3-epimerase | RIBULP3EPIM-RXN |
| Rub5PIso | ribose-5-phosphate isomerase | RIBOSE-5P |
| RuBisCO | ribulose-bisphosphate carboxylase | RIBULOSE-BISPHOSPHATE-CARBOXYLASE-RXN |
| SBPase | sedoheptulose-bisphosphatase | SEDOHEPTULOSE-BISPHOSPHATASE-RXN |
| SBPAld | sedoheptulose-bisphosphatase aldolase | SEDOBISALDOL-RXN |
| SClig | succinyl-CoA synthetase | SUCCCOASYN-RXN |
| SHIkin | shikimate kinase | SHIKIMATE-KINASE-RXN |
| SHIdh | shikimate dehydrogenase | SHIKIMATE-5-DEHYDROGENASE-RXN |
| StSyn | starch synthase | GLYCOGENSYN-RXN |
| SUCdh | succinate dehydrogenase | Complex-II |
| TransketI | transketolase | 1TRANSKETO-RXN |
| TransketII | glycoaldehyde transferase | 2TRANSKETO-RXN |
| Transald | transaldolase | TRANSALDOL-RXN |
| TriPIsom | triosephosphate isomerase | TRIOSEPISOMERIZATION-RXN |
| UDPgpyrophos | UDP-glucose pyrophosphorylase | GLUC1PURIDYLTRANS-RXN |
| VDE | violaxanthin de-epoxidase | RXN-13185 |
| ZeaxEp | zeaxanthin epoxidase | RXN-13193 |

| *Metabolites* | | |
|---|---|---|
| $\alpha$KG | $\alpha$-ketoglutarate | 2-KETOGLUTARATE |
| 2-PGA | 2-phosphoglycerate | 2-PG |
| 3-PGA | 3-phosphoglycerate | G3P |
| 6PG | 6-phospho-D-gluconate | CPD-2961 |
| AcCoA | acetyl-coenzyme A | ACETYL-COA |
| Acet | acetate | ACET |
| Ac-P | acetylphosphate | ACETYL-P |

| | | |
|---|---|---|
| ADP-GMH | ADP-L-glycero-D-emphmanno-heptose | ADP-L-GLYCERO-D-MANNO-HEPTOSE |
| ATP | adenosine-triphosphate | ATP |
| ADP | adenosine-diphosphate | ADP |
| AMP | adenosine-monophosphate | AMP |
| ASP | aspartate | L-ASPARTATE |
| Cit | citrate | CIT |
| CisAcon | cis-aconitate | CIS-ACONITATE |
| CMP-KDO | CMP-3-deoxy-D-*manno*-octulosonate | CMP-KDO |
| DHAP | dihydroxyacetone phosphate | DIHYDROXY-ACETONE-PHOSPHATE |
| DPG | 1,3-diphosphoglycerate | DPG |
| E4P | D-erythrose-4-phosphate | ERYTHROSE-4P |
| F6P | fructose-6-phosphate | FRUCTOSE-6P |
| FBP | fructose-1,6-bisphosphate | FRUCTOSE-16-DIPHOSPHATE |
| Fum | fumarate | FUM |
| G6P | glucose-6-phosphate | GLC-6-P |
| GAP | glyceraldehyde-3-phosphate | GAP |
| GLT | glutamate | GLT |
| GLN | glutamine | GLN |
| GL6P | D-glucono-$\Delta$-lactone-6-phosphate | D-6-P-GLUCONO-DELTA-LACTONE |
| Glyox | glyoxylate | GLYOX |
| GlycAld | glycolaldehyde | GLYCOLALDEHYDE |
| Glycol | glycolate | GLYCOLLATE |
| IsoCit | iso-citrate | THREO-DS-ISO-CITRATE |
| KDP | 2-keto-3-deoxy-6-phospho-gluconate | 2-KETO-3-DEOXY-6-P-GLUCONATE |
| Mal | malate | MAL |
| M-DAP | *meso*-diaminopimelate | MESO-DIAMINOPIMELATE |
| $NH_3$ | ammonia | AMMONIA |
| NADPH | nicotinamide adenine dinucleotide phosphate | NADPH |
| NAD | nicotinamide adenine dinucleotide | NAD |

| | | |
|---|---|---|
| NADH | nicotinamide adenine dinucleotide | NADH |
| NADP | nicotinamide adenine dinucleotide phosphate | NADP |
| OAA | oxaloacetate | OXALACETIC_ACID |
| PEP | phosphoenol pyruvate | PHOSPHO-ENOL-PYRUVATE |
| Pyr | pyruvate | PYRUVATE |
| R5P | D-ribose-5-phosphate | RIBOSE-5P |
| Rb5P | D-ribulose-5-phosphate | RIBULOSE-5P |
| RuBP | ribulose-1,5-bisphosphate | D-RIBULOSE-15-P2 |
| SBP | sedoheptulose-1,7-bisphosphate | D-SEDOHEPTULOSE-1-7-P2 |
| S7P | sedoheptulose-7-phosphate | D-SEDOHEPTULOSE-7-P |
| SucCoA | succinyl-coenzyme A | SUC-COA |
| Suc | succinate | SUC |
| UDP-Glc | UDP-glucose | UDP-GLUCOSE |
| UDP-Gal | UDP-Galactose | UDP-GALACTOSE |
| UnDe-P | undecaprenyl phosphate | UNDECAPRENYL-P |
| UDP-GA | UDP-*N*-acetyl-D-glucosamine | UDP-N-ACETYL-D-GLUCOSAMINE |
| X5P | D-xylulose-5-phosphate | XYLULOSE-5-PHOSPHATE |
| $H_i^+$ | intracellular proton | PROTON |
| $O_2$ | oxygen | O2 |
| $CO_2$ | carbon-dioxide | CO2 |

Table 2: Summary of reaction numbering in network diagrams.

| **r number** | **Reaction** |
|---|---|
| r1 | RuBisCO |
| r2 | PGKin |
| r3 | G3Pdh |
| r4 | TriPIsom |
| r5 | FBPAld |
| r6 | FBPase |
| r7 | TransketI |

| | |
|---|---|
| r8 | SBPAld |
| r9 | SBPase |
| r10 | Transald |
| r11 | TransketII |
| r12 | Rub5PIso |
| r13 | Rib5PEpi |
| r14 | PRK |
| r15 | StSyn |
| r16 | GlyPhos |
| r17 | PGM |
| r18 | PGI |
| r19 | lumped reaction of G6Pdh, 6PGLac and PGluDh. |
| r21 | 3PGAR |
| r22 | 2PGH |
| r23 | DAHPsyn |
| r24 | 3DQdh |
| r25 | SHIdh |
| r26 | SHIkin |
| r27 | PyrShiSyn |
| r28 | ChorSyn |
| r29 | G6P_tx |
| r30 | DHAP_tx |
| r31 | GAP_tx |
| r32 | G3P_tx |
| r33 | PEP_tx |
| r34 | TriPIsom |
| r38 | 3PGAR |
| r39 | 2PGH |
| r40 | PEPkin |
| r41 | PEPCk |
| r42 | ASPsyn |
| r43 | FBPAld |
| r44 | 6PF |
| r45 | FBPase |

| | |
|---|---|
| r46 | PGI |
| r49 | UDPgpyrophos |
| r53 | SucSyn |
| r54 | PYRdh |
| r55 | Maldh |
| r56 | CitSyn |
| r57 | ICdh |
| r58 | aKGdh |
| r60 | SClig |
| r61 | ComplexII |
| I | ComplexI |
| III | ComplexIII |
| IV | ComplexIV |
| V | ComplexV |
| l1 | LightCyc |
| l2 | LightNonCyc |

Table 3: Other commonly used abbreviations.

| | |
|---|---|
| ETC | Electron transport chain |
| FBA | Flux Balance Analysis |
| GSM | Genome scale metabolic model |
| KO | Knockout |
| LP | Linear programming |
| MILP | Mixed integer linear programming |
| NPQ | Non-photochemical quenching |
| OPPP | Oxidative pentose phosphate pathway |
| PSI | Photosytem I |
| PSII | Photosystem II |
| RCC | Reaction correlation coefficient |
| TCA | Tricarboxylic acid cycle |
| XC | Xanthophyll cycle |

# Chapter 1

# Introduction to Plant Metabolism

## 1.1 Project Background

The research carried out in this thesis is part of the AccliPhot consortium (`http://www.accliphot.eu/`), the Marie Curie Initial Training Network funded by the European Commission (grant agreement number PITN-GA-2012-316427). The multidisciplinary consortium is composed of eleven theoretical, experimental and industrial partners across Europe. The main research aim of AccliPhot is to investigate and understand short-term acclimation mechanisms to changes in light conditions in the photosynthetic organisms *Arabidopsis thaliana* a multicellular flowering plant, *Chlamydomonas reinhardtii* a unicellular algae and *Phaeodactylum tricornutum*, a unicellular diatom. Using the resources and expertise available in the consortium, experimental and theoretical methods have been combined to investigate metabolism and whole-organism behavior, in particular growth and biomass yield. The aim is to employ the findings to optimise and upscale biotechnological exploitation of photosynthetic microalgae for the production of biofuels and high-value commodities.

The contribution of this project to the overall aim of AccliPhot is the reconstruction of genome scale metabolic models (GSMs) of *A. thaliana* and *C. reinhardtii* and their analysis to develop a systematic understanding of photosynthetic

metabolism of these organisms. The models have been used to study potential responses to changing light conditions and, especially under high light, to identify potential new metabolic routes used for the dissipation of energy. Moreover, the models have been used to integrate various experimental data generated by the experimental partners in the consortium and the predictions made by the model have been a good starting point to design new experiments with more certainty of what to expect. Overall the aim is to use the models to gain new insight into the metabolic characteristics of these organisms and ultimately help to improve the photosynthetic productivity of useful biochemicals.

## 1.2 Motivation

Life on Earth is mainly powered by energy from the sun, captured by plants, algae and bacteria, through the process of photosynthesis, which uses sunlight, water and $CO_2$ from the environment to produce sugar and other organic compounds, and ultimately provides food for many living organisms, supporting their life on Earth. Also, it has been reported that one quarter of all synthetic medicines comes directly from plants and plant derivatives and more then 60% of cancer therapeutics available in the market or under test are natural product based (Brower, 2008).

With the growing population of the world, the demand for food and medicines rises every year. The Food and Agriculture Organization of the United Nations estimates that the world crop production increased by 160% over last 65 years. They predict that in order to ensure food security and improve nutrition to a growing population, around 1.4% more grain needs to be made each year, representing 30% more by 2030 and around 70% more by 2050 [http://www.fao.org/faostat/]. If agricultural production is to be increased at that rate, it will require the doubling of productivity per hectare, because of the limited arable land (Edgerton, 2009). Thus in order to improve crop yield and efficiently use arable land for a sustainable future it is vitally important that we throughly understand photosynthesis.

An in-depth knowledge of photosynthetic biochemistry will help to develop metabolic engineering strategies to improve crop productivity and nutrient-use efficiency (Sweetlove et al., 2016). Recent advances in plant science research have made it possible to use molecular genetic technology to manipulate biochemical path-

ways to achieve desired functions (Andrews and Whitney, 2003; Raines, 2006). Improving the photosynthetic rate correlates with biomass formation and thus has a positive influence on crop productivity (Kruger and Volin, 2006).

Over the last decade, various scientific organizations have been investing heavily to discover plants with improved characteristics. Various collaborative research and bio-engineering projects are being undertaken to help achieve this goal (Ort et al., 2015). The ready availability of data pertaining to plant genotype has stimulated a number of modelling efforts with the aim of investigating the feasibility of increasing productivity for industrial and pharmaceutical process. In this regard, systems biology and in particular the use of metabolic and regulatory network models, is a promising tool to broaden our knowledge and further explore the potential of biotechnological use of plants.

Systems biology is an interdisciplinary field of study that is focused on studying interaction between and within biological systems using mathematical and computational approaches. It is based on the assumption that properties of a system cannot be understood by focusing on one aspect of their highly interacting components (Joyard and McCormick, 2010). Thus the interface between mathematics, statistics, computer science and biology is used to understand the synergy between molecular constituents (DNA, RNA, proteins, enzymes, metabolites and ions). One of the aims of system biology is to use computational tools and mathematical models to extract useful information from readily available high-throughput datasets. The information can be used to develop new understanding of structure and dynamics of biological networks (Joyard and McCormick, 2010). The ultimate goal of plant systems biology in particular is to develop computational models capable of representing the entire biological organisation of plants (Gutiérrez et al., 2005). Such models can be used to further develop our understanding of molecular, cellular and physiological properties of plant species.

*A. thaliana* is an extensively studied organism and was the first plant to have its full genome sequenced (Arabidopsis Genome Initiative, 2000). This provided the foundation to collectively study the structure and dynamics of plant genomes and systematically identify the genes responsible for growth, development and environmental responses (Bevan and Walsh, 2005). A successful application of

predictive modelling in this model plant can bring forward a system-biological interpretation of plant biochemistry and the knowledge can guide metabolic engineering strategies to employ plants as biotechnological factories. With the advent of high throughput data, the trend of constructing large-scale models of different plant species is getting popular with remarkable success in discovering newer insight (Poolman et al., 2009; Cheung et al., 2013; Lakshmanan et al., 2016)

Within the scope of this thesis, we will examine photosynthesis properties of plants and algae using genome scale metabolic models (GSMs) and analyse responses of the network to further develop our understanding of photosynthetic biochemistry. The detailed discussion on GSMs and the mathematical concepts used to analyse them will be presented in following chapters. This chapter will give an overview of general biochemistry of plant metabolism and set a background for the rest of the investigations carried out in this thesis.

## 1.3 Introduction of organisms

### *Arabidopsis thalaiana*

*A. thaliana*, is a multicellular flowering plant and belongs to family Brassicaceae. It is a popular model organism for cellular and molecular plant biology research worldwide because of its many unique properties. It has a relatively small genome size compared to other higher plants and 114.5 Mb out of 125 Mb of its total genome has been sequenced (Arabidopsis Genome Initiative, 2000). It has a short life span, of 8 weeks from germination to maturing its seeds, which are produced in large amounts. The plant can grow in most environmental conditions and although it has no agronomic value, it resembles many economically important crop plants such as *Brassica Napsus*. Many scientists are striving to fully characterize the functions of its approximately 25000 genes with the goal of providing access to deeper understanding of plant development and environmental responses (Arabidopsis Genome Initiative, 2000; Bevan and Walsh, 2005).

### *Chlamydomonas reinhardtii*

*C. reinhardtii* is a unicellular green algae, belonging to family Chlamydomonadaceae, and is a model organism for the study of photosynthesis (Merchant et al., 2007). It can grow on phototrophic or mixotrophic, using other organic compounds like acetate as source of carbon, conditions. Mixotrophic growth of *C. reinhardtii*, on acetate, is faster (Boyle and Morgan, 2009; de Oliveira Dal'Molin et al., 2011; Chapman et al., 2015) compared to its phototrophic growth. Many scientific projects today are focused on algal research with the aim of identifying sustainable and profitable sources of energy - biofuels. Algae based biochemical products such as triacyl glycerols, methane and ethanol are all examples of potential biofuel precursors (Rittmann, 2008). Algae, along with other marine species, fix almost half the inorganic carbon in the atmosphere (Field and Randerson, 1998). Besides that, algal metabolism is being explored for photoproduction of hydrogen as an alternate biofuel (Baltz et al., 2014). More recently *C. reinhardtii* has been identified as a potential host for producing high value compounds for industrial biotechnology through synthetic biology approaches (Scaife et al., 2015). So a better understanding of metabolic behavior of algae becomes important on the grounds of both economical and environmental prospectives.

## 1.4 Structure of a plant cell

Plants have a complex cellular architecture, divided into various compartments, each of which carry out specific functions necessary for the organism's survival and growth. Two unique features in the structure of plant cells that distinguish them from those of other eukaryotic organisms are the presence of the cell wall and the chloroplast (Figure 1.1). The cell wall surrounds the cell membrane and is made up of structural components such as cellulose, lignin and pectin. It functions to provide support and rigidity to the cell, thus maintaining its structure. All cellular components are membrane bound and are located within the cytoplasm inside the cell wall (Figure 1.1). Each sub-cellular reaction space in the cell is enveloped by at least one membrane which forms a diffusion barrier preventing the uncontrolled exchange of intermediates (Linka and Weber, 2010). The chloroplast (see Figure 1.1 for detailed structure) is the site for photosynthesis. Mitochondria (see Figure 1.1 for detailed structure) house an electron transport chain that drives

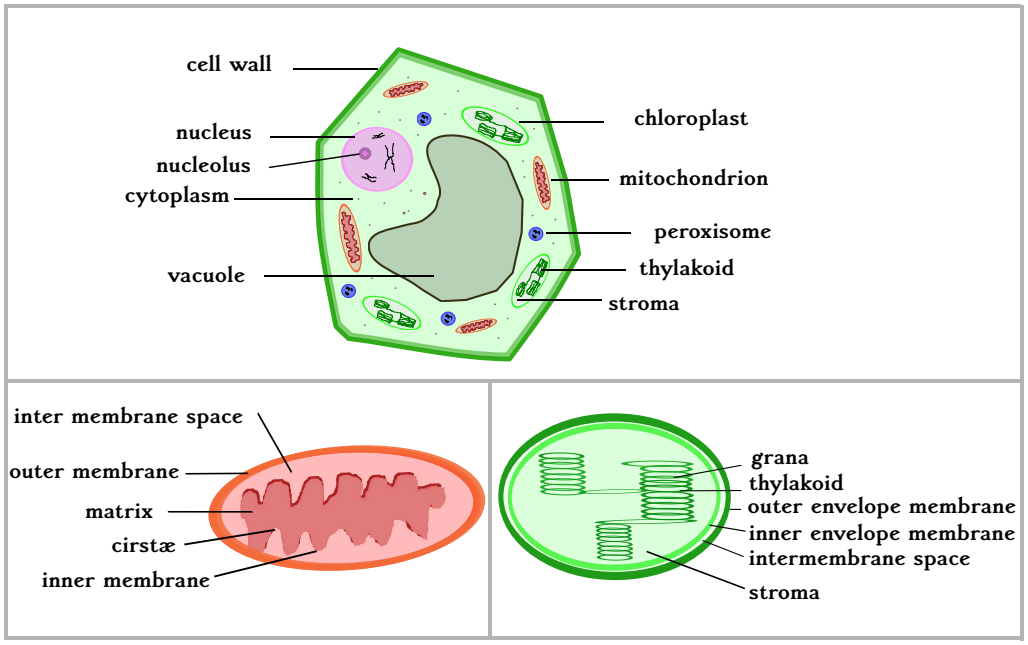



Figure 1.1: Schematic of a plant cell [top] showing the main organelles and compartments. The cell wall and the chloroplast are the unique feature in plant cells compared to other eukaryotic organisms. All the cellular components are located within the cytoplasm. The chloroplast is the site for photosynthesis. Mitochondria are important for production of ATP by oxidative phosphorylation. The vacuole is the largest compartment and helps to maintain the cellular rigidity. In the mitochondria [more detailed structure at bottom left], the outer membrane encloses the entire organelle while the inner membrane encloses the matrix where ATP is produced with the aid of ATP synthase contained in the membrane. Redox reactions of the oxidative phosphorylation also take place in the inner membrane. The cristae in the inner membrane expand the surface area of the inner membrane, enhancing its ability to produce ATP. Similarly, in the chloroplast [bottom right], the outer membrane is permeable to small organic molecules, whereas the inner membrane is less permeable and has specific transport proteins. The inner matrix, the stroma, contains grana of thylakoids where photosynthesis takes place. The thylakoids enclose the thylakoid lumen.

the synthesis of ATP by oxidative phosphorylation. The nucleus contains most of the genetic information organised as DNA and controls the activity of the cell by regulating protein synthesis within the cell. The nucleolus located inside the nucleus is the site for the synthesis of RNA. The vacuole is the largest compartment in the cell and contributes to the rigidity of the plant by developing hydrostatic pressure, storage of nutrients and breakdown of complex molecules. Metabolic interactions between the cellular compartments are required for a cell to operate as a functional unit. Metabolites are exchanged between the compartments for the supply of energy components or metabolic precursors. Thus in order to understand the complete functioning of the cell, it is important to know both how the different metabolic processes are compartmentalised and how they are linked together and controlled. The metabolic reactions in five main compartments and the underlying interactions between them are discussed below.

### 1.4.1 Photosynthesis : Chloroplastic metabolism

Photosynthesis is the process in which plants and algae synthesise organic compounds utilizing light, inorganic substances, and water. It is composed of two main components, light harvesting and carbon fixation, both of which take place in the double membrane-bound organelle, the chloroplast (Figure 1.1)

#### Electron transport chain

The electron transport chain (ETC) is a processes that transfer electrons across the membrane, from thylakoid lumen to chloroplast stroma, coupled with transfer of protons thus creating an electrochemical gradient, the proton motive force (PMF), that drives the synthesis of ATP. There are four main protein complexes in the thylakoid membrane, Photosystems I and II (PSI and PSII), Cytochrome b6f (cyt b6f) and ATP synthase (Figure 1.2). Light is absorbed at the PSI and PSII, at wavelengths 700 nm and 680 nm, respectively.

PSII upon absorption of a photon, transfers an electron, generated from water, to an electron acceptor called plastoquinone (PQ) creating a pool of electrons (PQ pool). PQ carries the electrons from PSII to the cytochrome b6f complex and itself gets reduced, to plastoquinol ($PQH_2$). The $PQH_2$ thus formed transfers its electrons through an intermediate electron transfer complex, cytochrome b6f, to

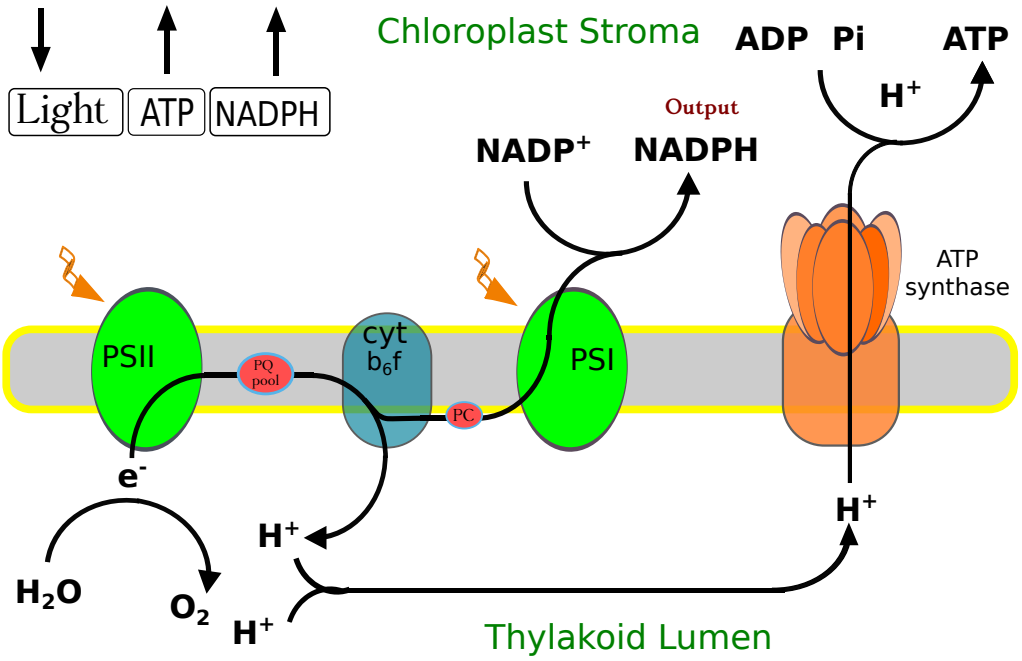



Figure 1.2: Simplified schematic representation of the photosystems located in the thylakoid membrane and the electron transport chain. Light absorbed in the PSII is used to transfer electrons, generated from oxidation of water, through cytochrome b6f to PSI thus reducing $NADP^+$ to NADPH. The concomitant release of proton in the lumen creates a gradient across the thylakoid membrane which is used to drive the synthesis of ATP.

the PSI. In the cytochrome b6f, electrons are passed one at a time from plastoquinol (PQ) to plastocyanin (PC), a water soluble electron carrier found in the thylakoid lumen, also leading to transport of a proton from stroma to lumen. The final stage of the light reactions are catalysed in PSI where a special pair of chlorophyll A molecules upon excitation transfer an electron through a chlorophyll and a bound quinone to a set of iron-sulfur (4Fe-4S) clusters. From here the electrons are transferred to ferredoxin (Fd), a water soluble mobile electron carrier located in the stroma. Further, the transfer of electrons from reduced ferredoxin to $NADP^+$ is catalysed by a flavoprotein ferredoxin-$NADP^+$ reductase (FADH2). This complex contains a tightly bound FAD which accepts the electrons one at a time from Fd. The FADH2 then transfers a hydride to NADP forming NADPH. The uptake of a proton by $NADP^+$ contributes to the pH gradient across the membrane. The proton motive force thus generated is used to drive the synthesis of ATP by the ATP synthase complex.

The phosphorylation of ADP to ATP is called photophosphorylation, which happens in a cyclic and a noncyclic manner. In the cyclic phosphorylation, electrons travels in cyclic manner between cyt b6f complex and the PSI, the process of which pumps the protons across the thylakoid membrane producing a pH gradient which then drives the synthesis of ATP (Figure 1.2, bold lines). The cyclic process does not involve PSII and only generates ATP. In non-cyclic phosphorylation, electrons travel in non-cyclic manner involving both PSI and PSII and generate ATP and NADPH.

### Carbon assimilation

The process of converting inorganic carbon from atmospheric $CO_2$ to organic carbohydrate derivatives by using the ATP and NADPH generated from the light reactions, is referred to as $CO_2$ assimilation or $CO_2$ fixation. The process occurs in the chloroplast stroma. The Calvin cycle, more correctly known as the Calvin-Benson-Bassham Cycle, in honor of Melvin Calvin and his collaborators Andrew Benson and James Bassham for its discovery (Bassham et al., 1950; Benson, 2002; Bassham, 2003), is the primary pathway to fix the external carbon and make it available to rest of the metabolism. The process is also referred to as the reductive pentose phosphate pathway (RPPP) because reduction happens and pentoses are

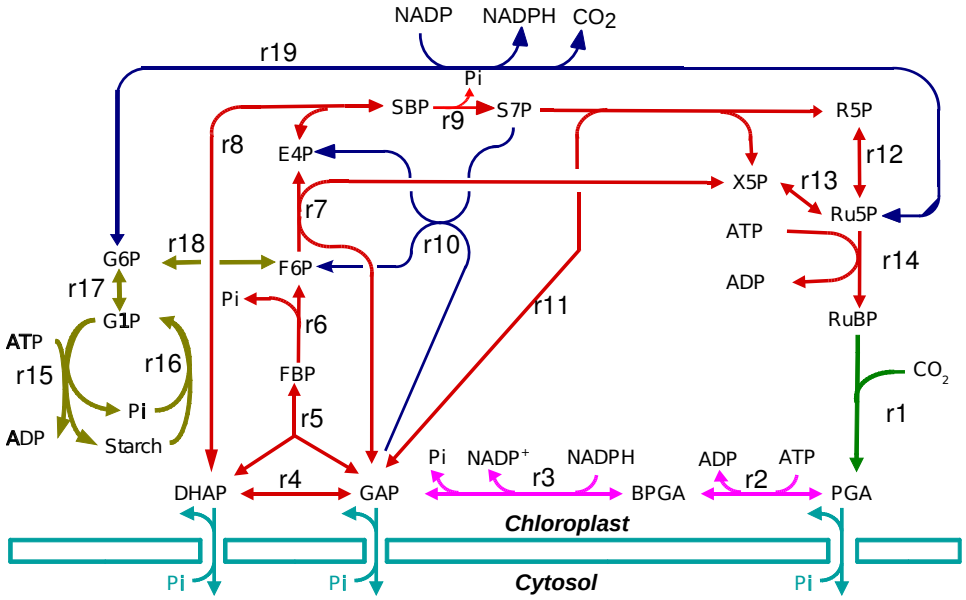



Figure 1.3: Different functional limbs of the Calvin cycle along with OPPP reactions are shown in different colors. Assimilation limb shown in green; r1 RuBisCO. Reduction shown in purple; r2 PGKin, r3 G3Pdh. Regeneration in red; r4 TriPIsom, r5 FBPAld, r6 FBPase, r7 TransketI, r8 SBPAld, r9 SBPase, r11 TransketII, r12 Rub5PIso, r13 Rib5PEpi, r14 PRK, Storage in olive; r15 StSyn, r16 GlyPhos, r17 PGM, r18 PGI. OPPP in blue; r10 Transald and r19 lumped reaction of G6Pdh, 6PGLac and PGluDh. The exchange of triose-phosphate molecules are represented in light blue color. Activity of four of the reactions in the cycle, r3, r6, r9, r14 is controlled by the Thioredoxin system represented in Figure 1.4. Figure adapted from (Poolman et al., 2003).

formed during the process (Heldt and Piechulla, 2011*a*). Based on function of reactions in the cycle, the whole process can be divided into four sections, referred to as limbs; assimilation, reduction, regeneration and storage (Figure 1.3) (Heldt and Piechulla, 2011*a*; Salisbury and Ross, 1992).

During the assimilation process, the enzyme ribulose1-5-bisphosphate carboxylase-oxygenase (RuBisCO, r1 in Figure 1.3) catalyses the carboxylation of $CO_2$ acceptor molecule ribulose-1,5-bisphosphate (RuBP), a five carbon compound, to form two molecules of 3-phosphoglycerate (3-PGA), a three carbon compound.

The PGA generated from $CO_2$ assimilation is converted to the triose phosphate molecules GAP and DHAP in a two step process using ATP and NADPH in the reductive limb of the cycle (r2, r3 in Figure 1.3.) The first step catalysed by phosphoglycerate kinase (PGK) attaches the phosphoryl group of ATP molecule to form BPGA. BPGA is then reduced by NADPH in the reaction catalysed by G3Pdh yielding GAP and $NADP^+$İn the regenerative phase, triose phosphate molecules are used to regenerate RuBP by rearrangement of carbon skeletons in a series of 10 reactions catalysed by the enzymes triose-phosphate isomerase (TriPIsom), fructose-bisphosphate aldolase (FBPAld), fructose-bisphosphatase (FBPase), transketolase (TransketI), sedoheptulose-bisphosphatase aldolase (SBPAld), sedoheptulose-bisphosphatase (SBPase), glycoaldehyde transferase or transketolase-II (TransketII), ribose-5-phosphate isomerase (Rub5PIso), ribulose-phosphate 3-epimerase (Rib5PEpi) and phosphoribulokinase (PRK) (Farquhar et al., 2001; Heldt and Piechulla, 2011*a*) (Figure 1.3).

The stoichiometry of the Calvin cycle requires five out of six triose phosphates formed by photosynthesis to regenerate 3 molecules of RuBP while the remaining one is utilised to produce starch, sucrose and other end products (Heldt and Piechulla, 2011*a*). The storage limb of the cycle (r15 -r18 in Figure 1.3) is involved in synthesis of the starch during the day and its degradation at night.

Various intermediates produced during the Calvin cycle are used in different parts of metabolism: the shikimate pathway uses the E4P to produce amino acids and lignin; G3P is used in the synthesis of secondary metabolites such as isoprenoid, and R5P is used for thiamine metabolism and cell wall biosynthesis. Nonetheless, it

is extremely important that a balance is maintained between the amount of carbon leaving the cycle and that retained to regenerate RuBP, for continual functioning of the cycle. The excessive production of other carbon derivatives (E4P, R5P) will inhibit photosynthesis as RuBP cannot be regenerated, while a low rate of end product generation will lead to problems such as accumulation of phosophorylated intermediates or depletion of free phosphate (Heldt and Piechulla, 2011*a*). Thus, there is a regulatory mechanism that controls the carbon flux in the Calvin cycle by controlling the activity of its different enzymes (Buchanan, 1980, 1991; Scheibe, 1991; Jacquot et al., 1997).

### Oxidative pentose phosphate pathway

Besides the reductive pentose phosphate pathway, the chloroplast also contains the enzymes of the oxidative pentose phosphate pathway (OPPP) which operates in the absence of light and oxidises a hexose phosphate to a pentose phosphate (Heldt and Piechulla, 2011*a*). The main function of the OPPP is to oxidise glucose phosphate and reduce NADP to generate NADPH, $CO_2$ and pentose phosphate. The pathway comprises reactions catalysed by G6Pdh, 6PGLac and PGluDh (shown as a single lumped reaction in Figure 1.3). Another important reaction of the pathway is Transald (r10 in Figure 1.3) Along with the Transald, all the OPPP reactions are active both in plants and algae under dark conditions (Buchanan and Balmer, 2005; Michelet et al., 2013).

Degradation of starch serves as an alternate source of carbon and energy during dark, when OPPP reactions are active. The starch is stored as granules in the chloroplast and contains two major sub-classes of alpha-glucans: amylose and amylopectin. Amylose contains fewer glucose molecues and has an 1,4-glycosidic linkage while amylopectin contains more glucose molecules and has additional 1,6-glycosidic linkage, thus is more branched. In order to use starch as a carbohydrate source overnight, the plant has to cleave these linkages and split starch into maltose and glucose. These species can then be transported into the rest of metabolism.

### Thioredoxin regulation

Plants and other photosynthetic organisms have to continuously adapt their metabolism to changing light conditions. As the enzymes of both the reductive and

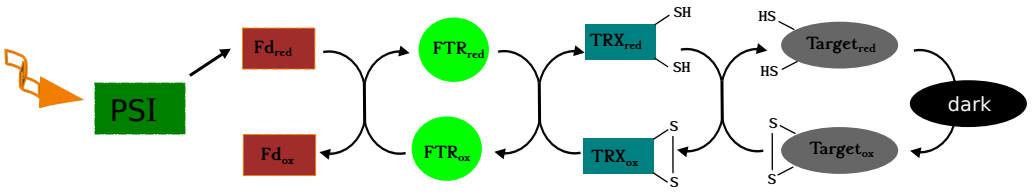



Figure 1.4: Schematic representation of the mechanism of action of thioredoxin (TRX). Light controls the activity state of thioredoxin system via reduction of ferredoxin (Fd) and via production of oxidants in chloroplasts. Chloroplast proteins are reduced by ferredoxin-dependent thioredoxin (FTR). The thioredoxin reductase reduces thioredoxin and donates electrons to free TRXs. The reduced TRX then catalyses the reduction of disulfide (s-s) within oxidised target proteins. Figure adapted from Michelet et al. (2013).

the oxidative pentose phosphate pathways are located in the chloroplast stroma, a concurrent operation of both of these pathways would represent a futile cycle consuming 3 molecule of ATP in each turn (Heldt and Piechulla, 2011*a*). Thus, a regulatory system, which prevents such waste of energy, operates in the stroma, activating RPPP in light and OPPP in the dark. The regulation is brought about by the modification of cysteine residues controlled by small disulfide oxido-reductases called thioredoxins (TRXs) (Michelet et al., 2013; Nikkanen and Rintamäki, 2014) The thioredoxin systems are an important part of the redox network, connecting light signals to chloroplast functioning (Nikkanen and Rintamäki, 2014). Figure 1.4 shows the mechanism of redox activation of the thioredoxin system in the plant chloroplast. The overall system consists of three types of soluble proteins located in the stroma; ferredoxin (Fd), ferredoxin/thioredoxin reductase (FTR) and thioredoxin (TRX). Upon availability of light, ferredoxin is reduced by the photosynthetic electron transport chain at PSI. The disulphide bonds in thioredoxin are reduced in a reaction catalysed by the enzyme ferredoxin-thioredoxin reductase when they receive the electrons. Thus reduced thioredoxin donates electrons for the reduction of disulphide bonds in four of the light-active Calvin cycle enzymes. These reductive reactions are accompanied by conformational changes that increase the activity of the enzyme. Under dark conditions, all these enzymes are inactivated as a result of re-oxidation of the disulphide bonds, thus stopping the $CO_2$ assimilation. The mechanism of light-dependent regulation of carbon

assimilation enzymes is considered as the best characterized mechanism of redox signalling in photosynthetic organisms (Foyer and Noctor, 2005).

Various biochemical measurements and enzymology experiments have established that Calvin cycle enzymes SBPase, FBPase, PRK, RuBisCO and G3Pdh along with G6pdh from OPPP are subject to regulation by light-dependent post-translational redox modification by thioredoxin (Buchanan and Balmer, 2005), light dependent change in pH and $Mg^{2+}$ feed back regulation by their product and allosteric regulation by other metabolites (Heldt and Piechulla, 2011*a*; Stitt and Sonnewald, 1995). The activity is mediated through reducing power produced by photosynthetic light reactions and is transferred by ferredoxin and thioredoxin to a disulphide bond in these enzymes thus activating them under light conditions and inactivating them under dark (Raines et al., 2000; Buchanan, 1991; Buchanan and Balmer, 2005).

### 1.4.2 Mitochondrial metabolism

Mitochondria provide the chemical energy from the oxidative degradation of organic and amino acids in plant cells (Mackenzie and McIntosh, 1999; Logan, 2006). They are composed of a smooth outer membrane surrounding an inner membrane with folds -the cristae, which provide increased surface area (Figure 1.1) The most important processes of mitochondria are the Tricarboxylic Acid (TCA) cycle and the electron transport chain.

#### Tricarboxylic Acid Cycle

The TCA cycle plays an important role in the complete oxidation of pyruvate to $CO_2$, generating reducing equivalents and TCA intermediates that serve as substrates for various biosynthetic pathways (Laloi, 1999; Haferkamp, 2007). Figure 1.5 represents the reactions and associated cofactors in the TCA cycle. Pyruvate enters the mitochondrial matrix through a pyruvate carrier. The cycle starts with the oxidation of pyruvate, catalysed by pyruvate dehydrogenase (PYRdh), located in the mitochondrial matrix, to form Acetyl CoA (AcCoA) with the concomitant release of $CO_2$ and reduction of NAD to NADH. The AcCoA thus produced is oxidised through a series of nine reactions forming a cycle known as the TCA cycle.

The first step of the TCA cycle is the conversion of AcCoA to citrate by citrate

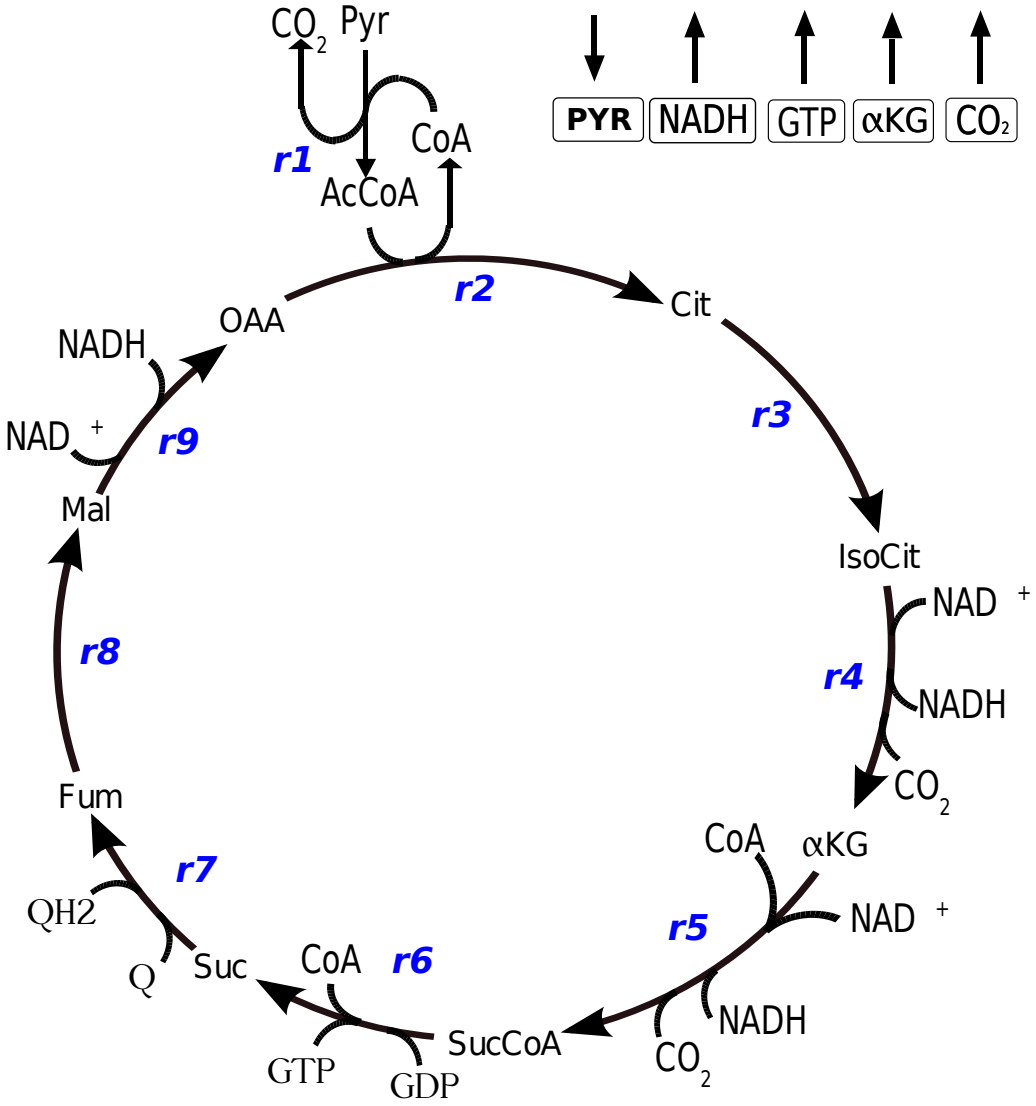



Figure 1.5: Representation of the TCA cycle reactions in the mitochondria. Pyruvate is transported from the chloroplast and oxidised to the acetyl CoA which in a series of eight reactions cycles to generate ATP in conjunction with electron transport chain and provide precursors or other pathways. r1, PYRdh; r2, CitSyn; r3, ACN; r4, ICdh; r5, aKGdh; r6, SClig; r7, SUCdh; r8, FUMhyd; r9, MALdh

synthase (CitSyn) (Fernie et al., 2004). Aconitase (ACN) then converts citrate to isocitrate via the bound intermediate cis-aconitate. ACN is also reported to play a crucial role in regulating carbon metabolism (Carrari et al., 2003) which is in agreement with the suggestion that the TCA cycle has a role in the illuminated leaf. The isocitrate thus produced is oxidatively decarboxylated to 2-oxoglutarate by either $NAD^+$ or $NADP^+$ dependent isocitrate dehydogenases (ICdh), generating $CO_2$ and NADH or NADPH respectively. Another oxidative decarboxylation reaction that catalyses the 2-oxoglutarate dehydrogenase complex (aKGdh), converts 2-oxoglutarate to succinyl CoA, producing more $CO_2$ and NADH. Succinyl CoA ligase (SClig) subsequently couples the synthesis of ATP from ADP and Pi with the concomitant conversion of Succinyl-CoA to succinate. Succinate dehydrogenase (SUCdh), also referred to as Complex II, catalyzes the oxidation of succinate to fumarate with simultaneous reduction of ubiquinone (Q) to ubiquinol (QH2). Further, the reaction catalyzed by fumarase (FUMhyd) converts fumarate to malate which is then oxidized to oxoloacetate (OAA) by $NAD^+$ dependent malate dehydrogenase (MALdh) to complete the cycle.

The MALdh is an important reaction not only for NADH oxidation but also as a component of the malate-aspartate and malate-OAA shuttles for the exchange of substrate, and reducing equivalents across the mitochondrial membrane.

### Oxidative Phosphorylation and the ETC

Oxidative phosphorylation or the mitochondrial electron transport chain (ETC) oxidises NADH generated from the TCA cycle and release electrons that are carried through a chain of electron carriers embedded in the inner membrane, leading to the formation of ATP. Figure 1.6 represents the oxidative phosphorylation and the flow of electrons across the mitochondrial matrix. NADH dehydrogenase, (represented as Complex I in Figure 1.6) is the first protein in the chain and catalyses the transfer of electron from NADH to ubiquinone. Similarly, the succinate dehydrogenase, represented as (Complex II), catalyses the transfer of electrons from succinate to quinones. The reduced quinone (Q) serves as a carrier of electrons which are passed to cytochrome bc1 complex (Complex III). Here, the electrons are passed to cytochrome C that delivers the electrons to cytochrome oxidase (COX) (Complex IV), which completes the chain by transferring them from reduced cytochrome

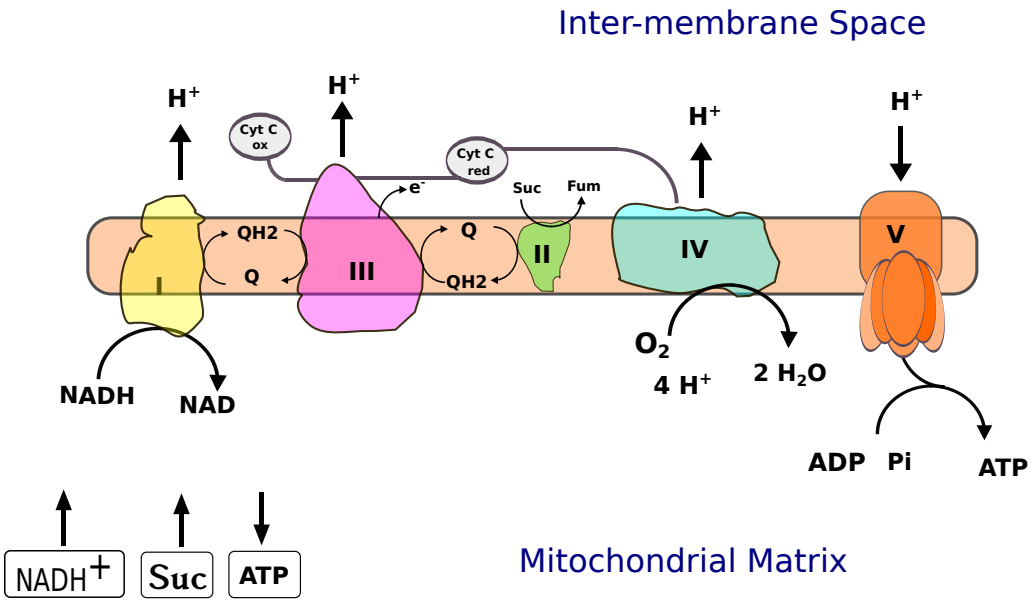



Figure 1.6: Representation of the flow of electrons through the mitochondrial ETC and the formation of ATP. The NADH and succinate generated in the TCA cycle are oxidised to generate ATP. Q,ubiquinone; $QH_2$, ubiquinol; Cyt C ox, cytochrome C oxidized; Cyt C red, cytochrome C reduced

C to $O_2$. A concomitant flow of protons from the matrix accompanies the flow of electrons through these complexes which creates an electrical gradient across the membrane. As the membrane is impermeable to protons, the proton-specific channel of the ATPase complex is the only route through which the protons can re-enter the matrix. Such movement of protein across the membrane is catalysed by the ATP synthase (Complex V) and results in the synthesis of ATP (Saraste, 1999).

### 1.4.3 Cytosolic metabolism

The plant cytosol is the intracellular space that functions to support other compartments in the cell and also acts as a link between intercompartmental transports (Ito et al., 2014). A wide range of biochemical interactions such as synthesis and degradation of various amino acids, sucrose and cellulose take place in the cytosol. The glycolytic pathway plays a major role in utilisation of carbohydrates and is present at least in part in all living organisms (Heldt and Piechulla, 2011*a*). Each cell has two sets of glycolytic enzymes, one localised in the cytosol and other in the plastid, each encoded by different genes. Some enzymes which take part in cytosolic glycolysis have a plastidic isoform and are involved in the Calvin cycle. The net effect of glycolysis is the oxidation of one molecule of glucose to two molecule of pyruvate with the generation of ATP and NADH (Plaxton, 1996). The pathway also produces five other important precursor metabolites: glucose-6P; fructose-6P; glycerone phosphate; 3-phospho-D-glycerate and phosphoenol pyruvate. These are used in other parts of metabolism. Figure 1.7 depicts the glycolytic pathway. The breakdown of glucose to pyruvate happens in two phases. The first phase is involved in phosphorylation of glucose released either from sucrose or starch to glucose-6-phosphate. The enzyme phosophoglucose isomerase (PGI) then converts G6P to F6P which is further phosophorylated by phosphofructokinase (PFK) to form fructose-bis-phosphate (FBP) (r2 1.7). FBP is then converted to DHAP and GAP by fructose bisphosphate aldolase (FBPAld). In the second phase two molecule of GAP are oxidised and phosophrylated by phosphorylating NAD-dependent GAP dehydrogenase (GAPdh) to form two molecule of 1,3-bisphosophoglycerate (BPGA) coupled by formation of NADH molecules (r5 1.7). In the following reaction, catalysed by the enzyme PGK, phosphoglycerate mutase (PGM), phosphopyruvate hydratase (PPH) and pyruvate kinase (PyrKin),

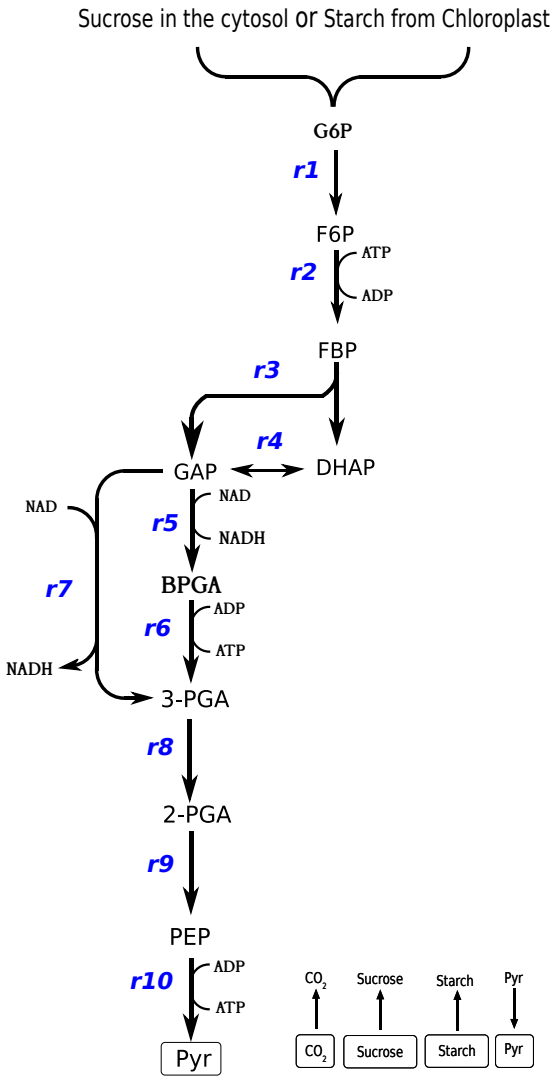



Figure 1.7: Glycolysis in the cytosol, converts the glucose, coming either form degradation of starch in the chloroplast or sucrose in the cytosol, to pyruvate with net production of ATP, in a series of intermediate reactions. r1, PGI; r2, PFK; r3, FBPAld; r4, TPI; r5, GAPdh; r6, PGK; r7, NP-GAPdh; r8, PGM; r9, PPH; r10, PyrKin

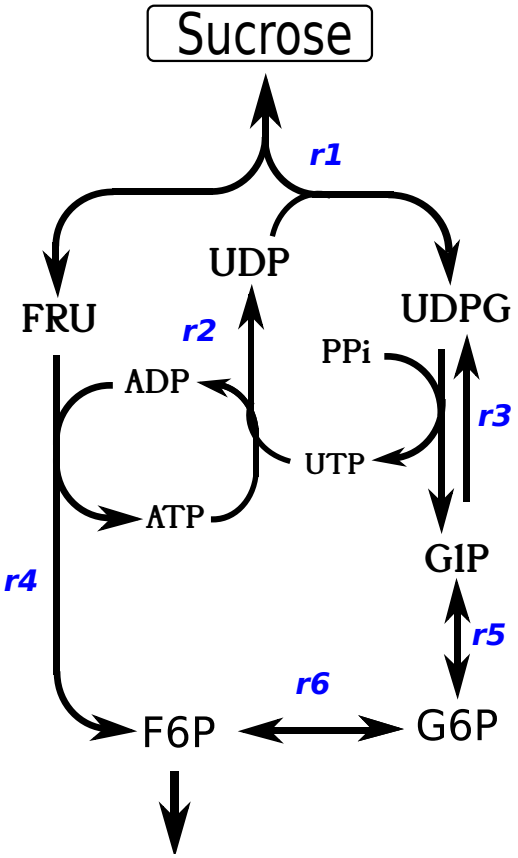



Figure 1.8: Representation of the synthesis and degradation of sucrose. r1, SUCSyn; r2, NDPK, r3, UDPPyPh; r4, HK; r5, PGM; r6, PGI

two molecules of BPG are converted to two molecules of pyruvate also leading to the formation of four molecules of ATP.

### Sucrose synthesis and degradation

Sucrose is one of the major products resulting from carbon fixation during photosynthesis. The triose phosphates generated from the Calvin cycle are transported from the chloroplast into the cytoplasm by TPT tranlocators and transformed to hexose posphate via the gluconeogenic pathway. These hexoses are used in sucrose biosynthesis in the cytoplasm. Like in case of starch biosynthesis, G1P reacts with a nucloside, UTP instead of ATP, to form UDP-glucose by the enzyme UDP-glucose pyrophosphatase (r3 in Figure 1.8). In the next step the conden-

sation between UDP-glucose and F6P by the enzyme sucrose synthase results in the formation of sucrose (r1 in Figure 1.8) However, the reaction is reversible and is also involved in the degradation of sucrose by catalysing the formation of UDP-glucose and fructose from UDP and sucrose. The enzyme occur mostly in non-photosynthetic organelles, such as the amyloplast supporting starch synthesis. Sucrose synthase also supports the synthesis of cellulose by delivering the glucose molecule in the form of UDP-glucose to cellulose synthase in the plasma membrane. Sucrose is also oxidised by the glycolytic enzymes to generate energy equivalents, reductants, and building blocks for anabolism, thus serving as a starting point for glycolysis and end point of gluconeogenesis in higher plants (Sung et al., 1988).

### 1.4.4 Interaction between compartments

The interplay of the metabolic processes between different compartments requires transfer of substances across the cellular membrane as well as between cells. Metabolism in every compartment depends upon these interactions for supply of energy, redox equivalents and metabolic precursors. Thus the metabolite transporters contribute to control the fluxes of solutes between compartments and are an essential part of metabolic networks (Philippar and Soll, 2007; Lunn, 2007). Protein structure predictions on *A. thaliana* have identified 1800 genes that could encode membrane proteins with transport functions (Schwacke et al., 2003).

There are various ways in which transfer of compounds takes place between compartments such as using specific translocators, channels, pores, via vesicles transport or by non specific diffusion through membranes (in the case of $CO_2$ and $O_2$) (Heldt and Piechulla, 2011*b*). When a molecule is moved across the membrane without help from any other metabolites, it is called *uniport* transport while the counter-exchange of molecules is called *antiport* transport. A simultaneous transport of two substrates in the same direction is called *symport*. A transport system which involves charge transfer along with uniport, antiport or symport transport is known as electrogenic transport. Involvement of photochemical reactions in the transport process is termed as primary active transport (Heldt and Piechulla, 2011*b*). The following section will provide an overview of transport systems related to photosynthesis focusing mainly on the mechanisms which connect the metabolic responses investigated within the scope of this thesis.

### Chloroplastic transporters

The core reactions of photosynthesis occur exclusively in the chloroplast. However the efficient and complete operation of photosynthesis critically depends on the presence of transport proteins that connect the chloroplast with other compartments. The stoichiometric coupling through transmembrane flux across compartments is vital for photosynthesis (Weber and Linka, 2011). The selectively-permeable chloroplast inner membrane (Figure 1.1) represents the interface between chloroplast and cytosol.

**Triose phosphate transport** Triose-phosphate translocators (TPT) located in the chloroplast inner membrane mediate the transport of TPs between the chloroplast and the cytosol. They catalyse the antiport transport of TPs or PGA with strict one-to-one stoichiometry for counter-exchange of Pi (Flugge and Heldt, 1984; Flügge et al., 2011). The net effect of each transport step is equivalent to three reduced carbon atoms, with no net transport of phosphate (Brautigam and Weber, 2011). Dissecting the physiological role of this transport protein is complex since TPs can either be exported to cytosol or stored inside the chloroplast in the form of transitory starch. Inorganic phosphates released from TPs during these biosyntheses are made available to chloroplast for continual operation of the photosynthetic machinery (Brautigam and Weber, 2011). This kind of exchange also provides a regulatory link between photosynthetic rates and cytosolic carbon metabolism (Brautigam and Weber, 2011). For example, decrease in the level of Pi in the cytosol indicates limited synthesis of sucrose, which means the Pi is not available for counter exchange with TPs. This in-turn promotes the reduction of triose phosphate, in the chloroplast itself, to be used for synthesis of starch.

Breakdown of transitory starch is critical to support metabolism during the night. Starch is broken down mainly to maltose and to lesser extent to glucose, both of which are exported to cytosol to be used for synthesis of sucrose (Weise et al., 2004). The export of maltose is mediated by a maltose exporter and glucose by the chloroplastic glucose transporter.

**Hexose phosphate transport** Another important transport system of the chloroplast is the G6P transport, mediated by the glucose-6-phosphate translocators (GPT). The expression of the GPT gene is mainly restricted to heterotrophic

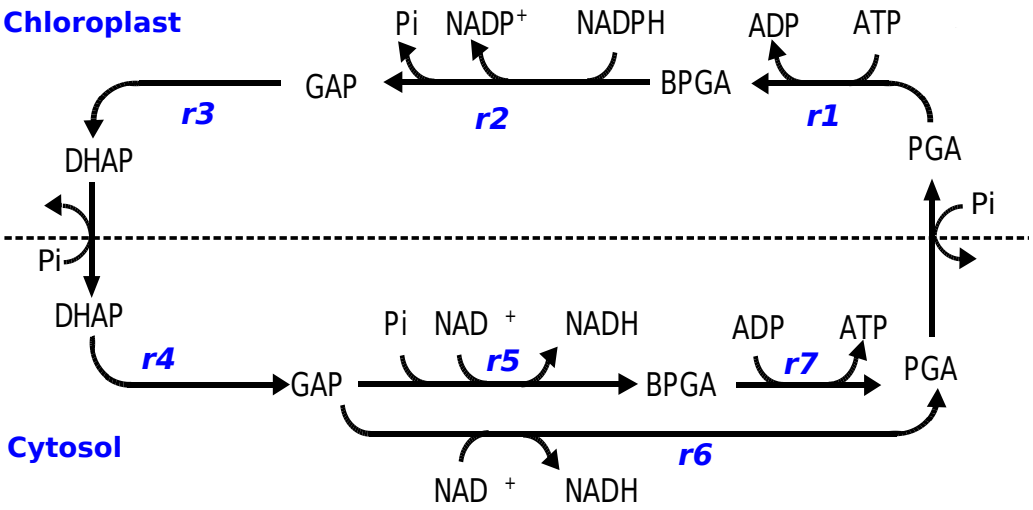



Figure 1.9: The TP-PGA shuttle between the chloroplast and the cytosol. TPs generated in the chloroplast are exported to cytosol are oxidised back to PGA with the net effect of one molecule of ATP and NADH released into the cytosol. r1, BPG; r2, GAPdh; r3, TPI, r4, TPIcy; r5, GAPdhcy; r6, NAD-GAPdh; r7, BPGdh

tissue (Kammerer et al., 1998) where its main function is to import G6P into the amyloplast. Transgenic experiments conducted by Niewiadomski et al. (2005) suggest that the import of G6P into the amyloplast is crucial for pollen maturation and gametophyte development. GPTs are highly expressed in stamens during plant development or in senescent leaves during flowering (Knappe et al., 2003; Niewiadomski et al., 2005). However, in-vivo experiment of transcript analysis, performed on samples from maize, pea and potato leaves, suggest that the GPT genes are not expressed in photosynthetic tissues (Kammerer et al., 1998). Nonetheless, expression has been noted, to a lesser extent, in guard cells and mesophyll cells. The guard cells lack FBPase activity and therefore rely on the availability of hexose phosophates transported for starch biosynthesis.

### Shuttle systems in the chloroplast

The reducing equivalents generated in the chloroplast are essential for enzymatic reactions in other compartments and must be transported. Due to their large size and charge, they cannot penetrate the selectively permeable inner chloroplast membrane. Two redox shuttle mechanisms, TP/3-phosphoglyceric acid (3-PGA)

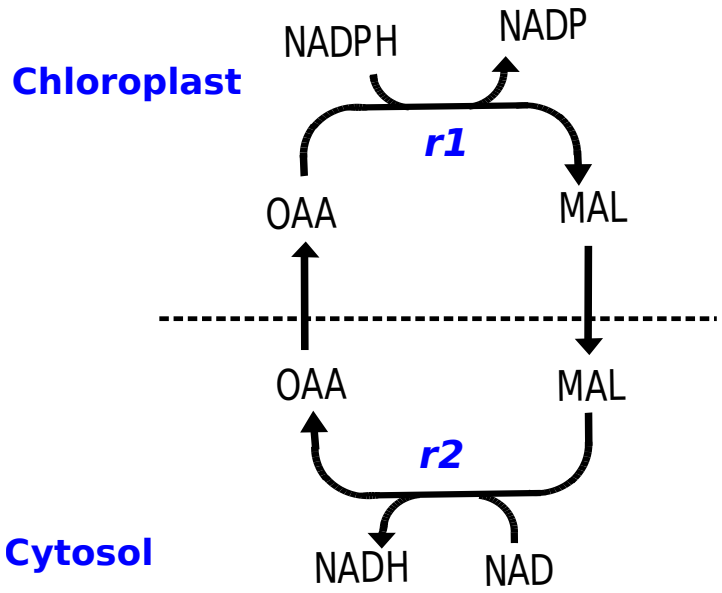



Figure 1.10: Representation of MAL-OAA shuttle between the chloroplast and the cytosol. OAA is converted to MAL by NADH dependent MALdh in the chloroplast and transported to the cytosol. MAL in the cytosol is converted back to OAA by NAD dependent MALdh. The export of MAL is accompanied by concomitant import of OAA in the chloroplast thus acting as a shuttle system. r1, NADP-MALdh; r2, NAD-MALdh

shuttle and the malate/oxaloacetate (OAA) shuttle, export these reductants out of the chloroplast and operate by the redox gradient between the chloroplast and the cytosol (Heineke et al., 1991). To maintain a proper balance of reductants and metabolites, the shuttle systems are finely controlled in response to stromal redox state (Taniguchi and Miyake, 2012). Moreover, as the energy components are not always consumed and produced evenly in the chloroplast, the redox shuttle also helps to maintain a balance in the ATP-NADPH ratio (Kramer and Evans, 2010).

The TPT of higher plants can exchange the TPs from 3PGA with net movement of one reducing equivalent. At first PGA is converted to triose phosphate, GAP and then to DHAP, by the use of ATP and NADPH (r1, r2, r3 in Figure 1.9). TP is exported to the cytosol by TPT and oxidised back to PGA, by the coupling of NAD dependent GAPdh and PGK, generating ATP and NADH. Thus for every molecule of TPs exported, one molecule of ATP and NADH each is released into the cytosol. However, if the conversion of GAP to PGA happens by the NAD-dependent non-phosphorylating GAPdh, only one molecule of NADH is released. This shuttle also maintains the redox level between the NADPH-NADP ratio in the stroma and the NADH-NAD ratio in the cytosol primarily by limiting the oxidation of TPs in the cytosol (Heineke et al., 1991).

The MAL-OAA shuttle is mediated by the dicarboxylate translocator DiT1 (Taniguchi et al., 2002). It exchanges OAA for malate and catalyses the net movement of one reducing equivalent without net C4 acid transport (Renne et al., 2003). As shown in Figure 1.10, OAA in the chloroplast is reduced to MAL by the reaction catalysed by the NADP-dependent malate dehydrogenase(NADP-MALdh). In the cytosol, MAL is converted back to OAA by a similar reaction catalysed by NAD dependent MDH (NAD-MALdh). The shuttle is initiated by the export of malate to the cytosol via malate transporters that are also linked to the OAA transporters such that any export of MAL results in a concomitant import of OAA, creating a MAL-OAA antiport system (Neuhaus and Wagner, 2000). The shuttle also functions as a safety valve, the so called malate valve, to export excess reducing equivalents from the chloroplast. The consumption of NADPH and regeneration of the electron acceptor NADP during the shuttle lowers redox status in the stroma. The valve also plays a role in carbon-nitrogen metabolism. Reducing equivalents exported by this valve can be used in oxidative electron transport in the mitochon-

dria and photorespiratory metabolism in the peroxisome (Scheibe, 2004; Noguchi and Yoshida, 2008; Kinoshita et al., 2011).

### ATP transport

The ATP-ADP translocators, are located in the chloroplast membrane (Heber and Santarius, 1970; Knappe et al., 2003). However, the activity of the ATP-ADP translocator on the chloroplast membrane, is very low, suited only for importing ATP into the chloroplast (Neuhaus and Wagner, 2000) and does not play a major role during the day. The maximum rate of ATP import, mainly observed in young leaves, has been found to be 10-fold lower than that of other metabolites such as Pi (Flugge and Heldt, 1991). The capacity for ATP transport in the chloroplast envelope have been reported to be 100-fold lower than that for TPs and that the knockout of the plastidial ATP transporter genes (NTTs) was found to have no effect on photosynthetic metabolism (Reiser et al., 2004). On the other hand, import of energy in the form of ATP is deemed necessary, to energise anabolic and catabolic reactions such as starch and fatty acid synthesis, in storage plastids and amyloplasts, at night (Neuhaus et al., 1993; Neuhaus and Schulte, 1996). The import of ATP into the stroma is also regulated by the distribution of ATP-ADP gradients in the cell and the ratio is higher in cytosol than in stroma both in photosynthesising and non-photosynthesising leaves (Heineke et al., 1991).

### Mitochondrial transport

The TCA cycle operating in the mitochondria also provides reducing equivalents and carbon skeletons to other pathways. Regular exchange of molecules between the mitochondrial matrix and cytosol takes place via the mitochondrial membrane (Figure 1.1). Metabolites are transported by diffusion from the outer membrane while the inner membrane contains specific transport proteins (Laloi, 1999; del Arco and Satrústegui, 2005). The most abundant transporter in the mitochondrial membrane is the ADP-ATP translocator which catalyses the export of ATP generated in the mitochondrial matrix through counter-exchange, electrogenic transport, with cytosolic ADP (Klingenberg, 2008). Due to the membrane potential generated by electron transport of the respiratory chain, ADP is preferentially transported inward while ATP is pumped outward. In this way, the mitochondrial ATP is supplied to other compartments (Klingenberg, 2008). Also, due to this

phenomena, the ATP to ADP ratio in the cytosol is higher than in the mitochiondrial matrix (Stitt et al., 1982; Heldt and Piechulla, 2011*a*).

Moreover, ATP synthesis via oxidative phosphorylation is dependent on the uptake of Pi into the mitochondrial matrix (Rausch and Bucher, 2002). A high level of Pi accumulation in the matrix is maintained by the mitochondrial phosphate carrier which uses the pH gradient across the membrane, generated by the mitochondrial electron transport chain, to facilitate the Pi influx into the membrane either by a Pi-OH antiport or Pi-$H^+$ symport.

A transport protein, dicarboxylate/tricarboxylate carrier mediates the transport of dicarboxylates (ketoglutarate, oxaloacetate, malate, and succinate) and tricarboxylates (citrate, isocitrate, aconitate) by a counter-exchange mechanism. Similarly, the export of the redox equivalents from the mitochondria is mediated by the shuttles catalysed by mitochondrial membrane transport proteins. Examples of such shuttle include the MAL-CIT shuttle, MAL-ASP exchange involving MAL-2KG and GLT-ASP translocator.

### 1.4.5 Discussion

Eukaryotic organism are complex system to study due to their multi compartment cellular architecture and interactions between them. The advent of modern high-throughput molecular biology techniques such as genome sequencing, gene expression profiling etc. have made organism specific genome annotations and biochemical information widely available through public databases. Development of highly efficient computational tools and techniques along with mathematical algorithms have made it possible to derive biologically important information from such experimental datasets and systematically study them to understand the structural and dynamic properties of the organisms. In order to fully understand behaviors resulting from the combined activity of multiple pathways, operating in different compartments of a cell, a system wide study of their characteristic is essential. Reconstruction of metabolic models, that can represent the interactions in an individual component, or entire biological networks, provide a platform to analyse existing data and use the knowledge to predict new functions. These models can be used to simulate metabolic or physiological responses under specific environmen-

tal conditions or genetic manipulation leading to development of new hypotheses and providing a guide to design new experiments. Although the reconstruction of such metabolic models is increasingly popular, the quest to fully understand the underlying capabilities of biological systems has not been achieved. The work in this thesis is aimed to develop further insights into plant metabolic network with special attention to their photosynthetic metabolism. The following chapters will describe the concept of metabolic modelling, construction of the models and their use to investigate various aspect of photosynthesis. The overall structure of the thesis is outlined below.

## 1.5 Aims and Structure of the thesis

The overall structure of the thesis can be outlined as follows:

- Chapter 2 introduces the general concepts of metabolic modelling and methodologies that will be used in later chapters.
- Chapter 3 describes the process of reconstruction of the GSMs of *A. thaliana* and *C. reinhardtii* and also presents the literature review of already existing models.
- Chapter 4 describes various photosynthetic properties of the GSMs whose constructions are discussed in Chapter 3.
- Chapter 5 describes the analysis of *A. thaliana* GSM, constructed in Chapter 3 and using the techniques discussed in Chapter 2, to investigate the properties of Calvin cycle enzymes and the effect of knocking out these reactions from the model.
- Chapter 6 describes the analysis of the *A. thaliana* GSM, constructed in Chapter 3 and using the techniques discussed in Chapter 2, to investigate energy dissipation mechanism under high light conditions.
- Chapter 7 presents the overall discussions of the results presented in this thesis and future perspectives.

# Chapter 2

# Mathematical modelling of metabolism - General Methodology

## 2.1 Introduction

Metabolic systems consist of a number of biochemical reactions and associated metabolites that are consumed and produced. Metabolic models can represent a network of such reactions in mathematical notation as a set of variables and equations (Heinrich and Schuster, 1996). Once a model is created, it can be used as a framework to apply certain logical and mathematical reasoning to simulate real life situations, to study and to understand the functional behavior of an organism. In this regard, a model can serve as a *in silico* lab, allowing us to perform complex investigations in a short time compared to expensive *in vivo* experiments. The predictions thus made from the model can be used to generate new hypotheses and design new experiments with more information on what to expect.

Based on the particular properties of the system represented by a metabolic model it can be classified as either *s*tructural or *k*inetic. Structural models are defined in terms of stoichiometry and thermodynamics of the system. Information such as a list of reactions, number of routes, optimal stoichiometries, network flux values can be obtained from such models. In addition, kinetic models include kinetic proper-

ties of enzymes and can generate time courses, steady-states, sensitivity analyses or control distribution.

### 2.1.1 Fundamental concepts

A model consists of a list of selected metabolic reactions and associated metabolites. These metabolites are categorised as *internal* or *external* based on the modelling considerations. Internal metabolites are produced and consumed by intracellular metabolism of the system under study while the external metabolites are assumed to have constant concentration maintained by the environment and are likely to be exchanged across the system boundary. Examples of external metabolites include:

- source metabolites such as $CO_2$ and sink metabolites such as biomass components
- metabolites whose internal concentration is high enough that any change in concentration is negligible. eg. $H_2O$, sometimes polymers such as starch

All information about the link of substrates with product can be represented using a matrix, termed as stoichiometric matrix, conventionally denoted as **N** (Figure 2.1). Thus the coefficients of all reactions in a system can be used to create the stoichiometric matrix **N** of dimension $m * n$ where each row, $m$, is a representation of a metabolite and each column, $n$, a reaction. An element in **N** is the stoichiometric coefficient of a metabolite in a particular reaction. Positive or negative value of the indicates that the metabolite coefficient is being produced or consumed by the corresponding reaction respectively, and a value 0 means that it is not involved in the reaction (Heinrich and Schuster, 1996). The stoichiometry matrix represents the structural architecture of a network and thus the knowledge of its properties is prerequisite for any mathematical analyses (Schilling et al., 1999).
The concentration of the internal metabolites is not constant. It is determined by their initial values and rates of their consumption and production by the reactions in the system. Consider metabolite A, as represented in the Figure 2.1, is produced by the transporter, $A_tx$ and consumed by the reaction $R_1$. Thus the rate of change of A can be expressed as:

$$dA/dt = V_{A_tx} - V_{R1} \tag{2.1.1}$$

where, $VA_tx$ and $V_R1$ are fluxes carried by reactions $A_tx$ and $R_1$. Similarly, the rate of change of all internal metabolites represented in Figure 2.1 can be represented as a set of ordinary differential equations (ODEs), determined simply by the stoichiometry of the network and the reaction rates, as shown below:

$$\begin{aligned} dA/dt &= V_{A_tx} - V_{R1} \\ dB/dt &= V_{R1} - V_{R2} - V_{R3} \\ dC/dt &= V_{R2} - V_{R4} \\ dD/dt &= V_{R3} - V_{R5} - V_{R6} \\ dE/dt &= V_{R4} + V_{R5} - V_{E_tx} \\ dF/dt &= V_{R6} \end{aligned} \tag{2.1.2}$$

Equation 2.1.2 can be more succinctly written as

$$\mathbf{dS/dt} = \mathbf{N}.\mathbf{v} \tag{2.1.3}$$

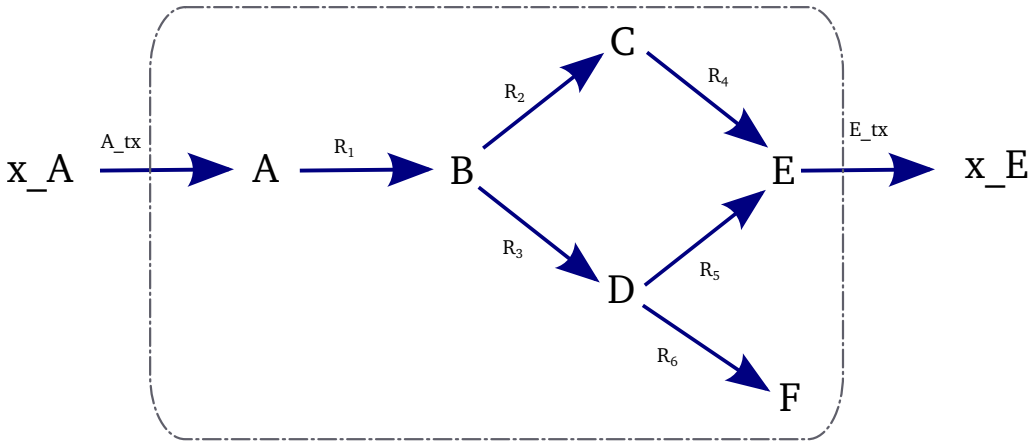
where **S** is the vector of the internal metabolite concentrations, **N** (shown in Figure 2.1) is the stoichiometric matrix and **v** is the vector of reaction rates.

### 2.1.2 The steady-state assumption

There is a constant exchange of fluxes from the source to the sink in all metabolic networks under steady state conditions, the concentration of the intermediates remain constant because their rates of formation equals their rates of degradation. This entails that the rate of production of all internal metabolites must always equal the rate of their consumption and that their total concentration is invariant to the time. The assumption is a base of most standard metabolic modelling applications. The rate of change of the metabolite A, as discussed in the example above, can thus be expressed as:

$$dA/dt = V_{A_tx} - V_{R1} = 0 \tag{2.1.4}$$

This equation represents the relationship between the two steady state fluxes, but is independent of their actual value, and is therefore also independent of their kinetic characteristics (Poolman et al., 2004). If all the vectors representing the reactions fluxes are represented as **v** and the vectors of zeros represented as **0**, the steady state of whole system, as shown in the Figure 2.1 can more concisely be




$$\mathbf{N} = \begin{array}{c} \begin{matrix} R_1 & R_2 & R_3 & R_4 & R_5 & R_6 & E_tx & A_tx \end{matrix} \\ \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \end{matrix}$$

Figure 2.1: Diagram in $A$ represents a simple hypothetical network. Metabolites with preffix x_ represent external metabolites. The matrix, below the diagram, represents the stoichiometric matrix ($\mathbf{N}$), excluding the external metabolites (x_A and x_E). Column in the matrix represents a reaction and each row represents the stoichiometric coefficient of a metabolite in the corresponding reaction. Consumption of a metabolite by a reactions is denoted by negative coefficient while production is denoted by positive coefficient, elements for metabolites not participating in the reaction are filled with 0.

written by modifying equation 2.1.3 as :

$$\mathbf{dS/dt} = \mathbf{Nv} = 0 \tag{2.1.5}$$

The equation holds true for any system capable of reaching steady state, and provides the foundation for the methods used in this thesis.

### 2.1.3 Structural Modelling

Structural models are constructed considering only the stoichiometry of the reactions and are designed to describe the network topology of the system rather than its kinetic behavior (Poolman et al., 2004). All kinetic models are capable of representing the structural properties, but the opposite is not true.

### 2.1.4 Null Space Analysis

The *null-space* of the stoichiometry matrix covers all possible steady state behavior of a network and allows identification of relationship between fluxes thus can be a base for most of the structural analysis. The equation 2.1.5 is a homogeneous system of linear equations and as the fluxes $v$ are variables, the equation is underdetermined, meaning there are infinitely many flux solutions that satisfy the equation. This space of possible solutions is called the null-space and can be derived from the stoichiometry matrix $\mathbf{N}$, either by applying Gaussian elimination (Cornish-Bowden and Hofmeyr JH, 2002) or Singular Value Decomposition (Famili and Palsson, 2003). The null-space can be defined mathematically using a kernel matrix ($\mathbf{K}$) where the columns are linearly independent vectors that together form a basis spanning the vector space (Schilling and Palsson, 1998) and is more formally written as :

$$\mathbf{N} \cdot \mathbf{K} \ = 0 \tag{2.1.6}$$

An example of kernel matrix for the example shown in Figure 2.1 is given by:

$$\mathbf{K} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \\ E_tx \\ A_tx \end{matrix}$$

Where each column represents a flux distribution and potential route through the network. The row order of the kernel matrix is same as the column order of (**N**). Although the null-space analysis does not take into account the thermodynammics and it is rather hard to integrate experimental flux observations, some very important steady state properties of a metabolic network can be derived from it. Some of its applications are discussed below.

### Dead Reactions

Reaction $R_6$ shown in the kernel matrix above is an example of a dead reaction. Mathematically, the rows with values of zeros in the kernel matrix correspond to dead reactions, which cannot carry a steady-state flux. One of the reasons for reactions to be dead is the presence of either orphan, or dead-end metabolites.

Orphan and dead-end metabolites are not calculated from K (null-space) but can be identified from the stoichiometric matrix directly. Metabolite $F$ as shown in Figure 2.1 is an example of an orphan metabolite. These are the metabolites which are involved with just one reaction in the network and are either consumed or produced only by the reaction. They cannot be balanced at the steady state. Mathematically, a row with single entry in the stoichiometric matrix represents the orphan metabolite. Dead-end metabolites are the metabolites that are associated with more then one reactions; however they are either only consumed or only produced by these reactions. Mathematically, rows with more then one entry, in stoichiometry matrix, with either only positive or only negative coefficients represent the dead-end metabolites, provided the reactions converted are all irreversible reactions; however they are only consumed or only produced. Any missing

reactions from the network or reversibility issues that contribute to disconnected network results in the orphan or dead-end metabolites.

### Enzyme Subsets

Reactions $R_3$ and $R_5$ as represented in the kernel matrix **K** are an example of enzyme subset (Pfeiffer et al., 1999). These are a set of reactions which will carry steady state flux in a fixed proportion. Mathematically, row vectors in the kernel matrix that are scalar multiple of each other (parallel vectors) correspond to enzyme subsets. If any of the reactions in a subset are removed, then the rest of the reactions in the subset will have a zero flux at steady state. Similarly, all reactions in a subset can be lumped together as a single equivalent reaction to reduce the dimensionality of a network.

It is important to note that null space analysis does not take into account reaction reversibility. For example even if reaction $R_3$ in Figure 2.1 is defined in opposite direction i.e

$$R3 : B \leftarrow D$$

the kernel matrix K would still be the same and the row vectors in K would still suggest that reactions are in same subset. However, irreversible reactions in a subset if defined in opposing direction cannot carry flux at steady state and thus are referred to as *inconsistent enzyme subsets*. Metabolic models should be devoid of such subsets and are recursively checked during the process of model curation.

### Reaction Correlation Coefficient

Reaction Correlation Coefficients (RCC) can be calculated from the null space **K** of the stoichiometric matrix and is an extension to the concept of enzyme subsets discussed in section 2.1.4 (Poolman et al., 2007). RCC is similar to the Pearson's (population) correlation coefficient between the fluxes carried by the pair of reaction $i$ and $j$ for all possible steady states in the model and is calculated from the row vectors in the null-space. The angles between the row vectors of the kernel matrix are unique provided the matrix is orthogonal i.e if all column vectors are perpendicular to each other and are of unit magnitude. Hence, the RCC is always calculated using the orthogonal kernel matrix where no row vector are zeros

(dead reactions removed). Mathematically, RCC ($\phi_{ij}$) between the fluxes carried by reactions in the kernel matrix $\mathbf{K}$ is calculated by the cosine angle ($cos(\theta_{ij}^{\mathbf{K}})$) between two row vectors $K_i$ and $K_j$ such that:

$$\phi_{ij} = cos(\theta_{ij}^{\mathbf{K}}) = \frac{\mathbf{K}_i \mathbf{K}_j^T}{\sqrt{\mathbf{K}_i \mathbf{K}_i^T}\sqrt{\mathbf{K}_j \mathbf{K}_j^T}} \tag{2.1.7}$$

where, $\mathbf{K}$ is an orthogonal kernel matrix with $m$ rows, $i, j \in 1...n$ and $\theta_{ij}^{\mathbf{K}}$ is the angle between rows $i$ and $j$ in $\mathbf{K}$. RCC is statistically identical to Pearson correlation coefficient and is in the range [-1, 1] where 0 represent no correlation between the reaction fluxes while 1 represent 100% correlation between the reaction pair.

The reaction correlation of a network can be visualised using the clustering of reactions into hierarchal trees. Many algorithms can be used for this purpose. Here, the WPGMA algorithm (Weighted Pair Group Method using Arithmetic Averaging) (Morgan and Ray, 1995) has been used, which groups objects (reactions) into clusters based on similarity measure (RCC) to generate dendogram. Here, the absolute value of the RCC is used for clustering purpose. A dendogram, can be visualised using phylogenetic tree viewing programs such as NJPlot (Perrière and Gouy, 1996) and are, referred to as *metabolic trees* in this thesis. A metabolic tree is generally composed of following components.

**Graph :** Collection of nodes joined by edges

**Tree :** An acyclic directed graph

**Node :** In a tree a node has 0 or 1 parent nodes and $\geq 0$ or more children

**Leaf node :** A node with no children.

**Internal node :** A non-leaf node.

**Sub-tree :** An internal node and all its descendants.

**Root node :** A node with no parent and and one or more children. Every tree has exactly one root

In a metabolic tree, each leaf node corresponds uniquely to a reaction, and all other nodes thus represent a collection of reactions with related metabolic functions. These metabolic trees are useful to interpret the properties of the metabolic

networks, identify disconnected subnetworks in the metabolic models and decompose the large network to smaller functional modules. Absolute values of RCC can also be used to study similarities between reactions by clustering them based on correlation of fluxes carried by them. The analysis of such correlated fluxes can help to identify specifically coupled reactions and functional modules in the metabolic network. We have used the technique to identify energy dissipating cycles under illuminated conditions in plants and algae, which will be discussed in more details in Chapter 6.

### Conserved Moieties

The conservation relationship between metabolites, also known as conserved moieties is a characteristic feature of biological networks where certain molecular subgroups are conserved (Heinrich and Schuster, 1996; Hofmeyr, 1986). Some components of conserved moieties are not synthesised or degraded by the metabolism described by the model. Conserved moieties are identified by examining the linear dependencies between metabolites in the *left null space* of stoichiometric matrix $\mathbf{N}$, which is calculated as the null-space of the transpose of $\mathbf{N}$. Mathematically, it can be expressed as:

$$\mathbf{K}^T \cdot \mathbf{N} \;=\; \mathbf{0}^T$$

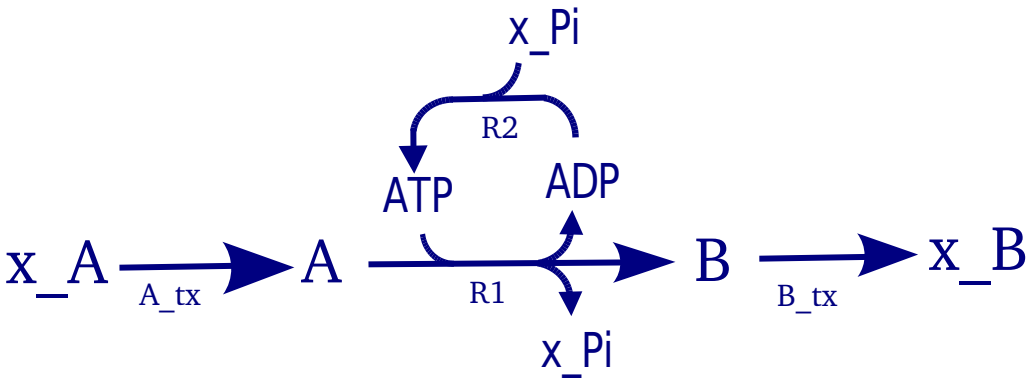
or

$$\mathbf{N}^T \cdot \mathbf{K} \;=\; \mathbf{0} \tag{2.1.8}$$

An example of such relationship is illustrated in Figure 2.2, where the sum of the concentration of metabolites ATP and ADP is constant. These metabolites are not consumed as product but are used as cofactors where one form is converted to the other form and the sum of their concentrations always remains constant in the system. This property can be derived by the analysis of left null space where conserved moieties have the same coefficient.

## 2.1.5 Elementary Modes Analysis

The null-space does not take into account thermodynamic constraints on reactions and it is rather hard to integrate experimental flux observations. These limitation




$$\mathbf{N} = \begin{array}{c} \begin{matrix} \mathbf{R_1} & \mathbf{R_2} & \mathbf{R_3} \end{matrix} \\ \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ -1 & 0 \\ 1 & 0 \end{bmatrix} \begin{matrix} A \\ B \\ ATP \\ ADP \end{matrix} \end{array}$$

$$\mathbf{K} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \begin{matrix} A \\ B \\ ATP \\ ADP \end{matrix}$$

Figure 2.2: A simple network diagram representing the conservation relationship, its associated stoichiometric matrix ($\mathbf{N}$) and left null space ($\mathbf{K}$). Metabolites with suffix $x_$ are considered external and are not shown in the matrix for simplicity. The left kernel, $\mathbf{K}$, shows the conservation relationship between metabolites ATP and ADP.

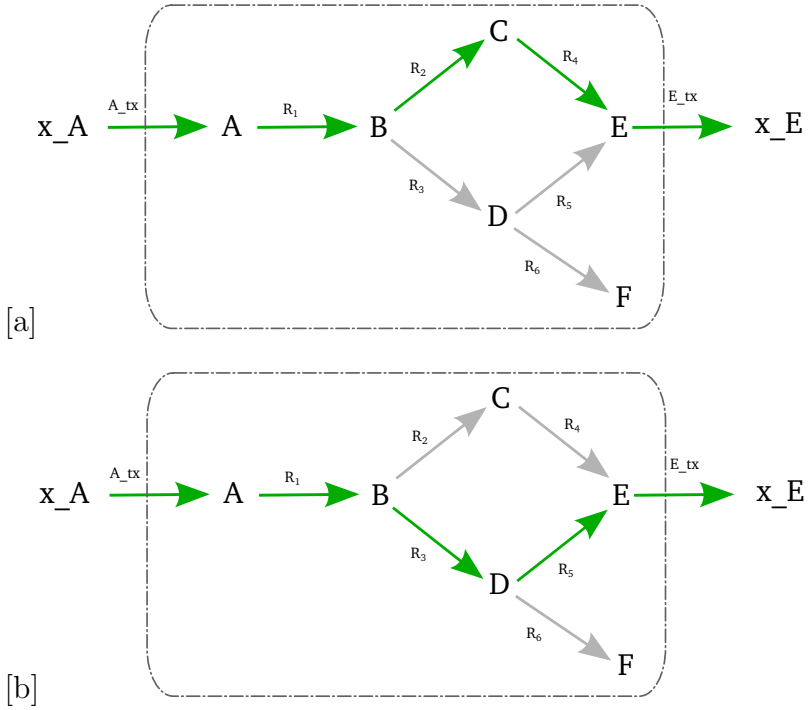



Figure 2.3: Elementary modes for a simple hypothetical network. Reactions involved in elementary mode are represented with green arrow while other reactions, in the network, that are not a part of elementary mode are presented in gray. Both of the elementary modes, [a] and [b], convert metabolite x_A to metabolite x_E and the mode cannot be further decomposed into smaller functional subset. Also, all the participating irreversible reactions are active in forward direction and thus are thermodynamically feasible.

of the null-space analysis were addressed by the introduction of elementary (flux) modes conceptualised by Schuster and Hilgetag (1994). The elementary mode can be defined as a steady-state flux distribution by a minimal set of reactions in a network such that they comply with the thermodynamic feasibility and cannot be further decomposed into smaller subsets (Schuster et al., 1999; Schurmann and Jacquot, 2000; Schuster et al., 2002). The thermodynamic criterion makes sure that all irreversible reactions in the mode proceed in the appropriate direction while the non-decomposability criterion makes sure that the elementary mode cannot be represented as a positive linear combination of any other flux modes in the network and thus act as a minimal functional units within the network.

Consider the representative network shown in Figure 2.1. The two elementary modes that occur in this network are represented in Figure 2.3 [a] and [b]. Both of these modes satisfy all the condition stated above. They satisfy the steady state assumption as all the internal metabolites are stoichiometrically balanced, they are thermodynamically feasible as all the reactions involved are irreversible and proceed in the appropriate direction and the modes cannot further be broken down into smaller network.

Several algorithms have been described to compute the elementary modes such as the canonical basis approach suggested by Schurmann and Jacquot (2000), the null-space approach proposed by Wagner (2004) and the binary approach discussed by Gagneur and Klamt (2004). Although all of these algorithm have their own significance, their application is limited to smaller or medium sized network due to combinatorial explosion in number of elementary modes in the larger (genome scale) networks. The number of elementary mode increases exponentially with increasing network size (Klamt and Stelling, 2002). The number of extreme pathway (Schilling et al., 2000) (discussed in more details below) for the genome scale model of human metabolism was found to be in the order of $10^{29}$ Yeung et al. (2007). Such explosion can to some extend be controlled by changing reversibility of reactions. Irreversible reactions reduce the number of elementary modes by imposing additional thermodynamic constraints thus reducing the solution space while the identification of enzyme subsets could reduce the computational load by representing the subset with just a single reaction. Optimisation methods such as genetic algorithms have also been used in conjunction with elementary mode

analysis (Kaleta et al., 2009) to address the problems of combinatorial explosion. In another slightly different approach, Sridharan et al. (2015) used the elementary mode with a network partition method that is based on cyclic interactions between reactions. Similarly, Pey et al. (2015) developed a method called TreeEFM, which uses the linear programming-based tree search method to list subsets of elementary modes.

Schilling et al. (2000) has proposed an alternative approach to compute all unique and minimal set of steady state flux distribution in the network, referred to as *extreme pathways*. Extreme pathways are the subsets of the elementary mode (Schuster and Hilgetag, 1994) with the additional feature of systemic independence meaning none of the extreme pathways is a non-negative linear combination of others. In other words coefficients of linear combinations in the extreme pathway cannot be negative Schilling et al. (2000). Extreme pathways are identical to elementary modes if all the reactions in the network are irreversible; however if the reactions are reversible, they are decoupled into two separate reactions with forward and backward direction and subsequently computed for the pathway for the network. This property of extreme pathways makes them computationally less expensive and thus they could be applied to relatively larger networks (Klamt and Stelling, 2003). However, special consideration is required if extreme pathways are to be used to study system properties as they do not produce a complete set of independent routes within the metabolic network, which limits the assessment of structural robustness and relative importance of reactions in the network (Klamt and Stelling, 2003).The difference between elementary modes and extreme pathways is discussed in more detail by Papin et al. (2003).

A number of methods primarily based on the concept of elementary mode analysis have been developed to address these issues. Kaleta et al. (2009) developed a method called EFMEvolver, that couples with a genetic algorithm to compute elementary modes on the genome scale level by targeting specific functional modes without having to compute the entire set of elementary modes.

### Applications of Elementary Modes Analysis

Elementary mode analysis allows the examination of the relationships between the source and sink metabolites in a metabolic network and, further, the flexibility (redundancy, robustness) of the network. It has thus been useful to investigate the properties of metabolic network and proves to be a useful tool for metabolic engineering (Papin et al., 2003; Gagneur and Klamt, 2004; Trinh et al., 2009). Moreover, the knowledge of the frequency distribution of a particular reaction or set of reactions in the elementary modes provides an idea of the importance of the reaction, for the system's performance, under the investigation conditions (Jorg Stelling et al., 2004). It has been applied to several biological problems such as study of biochemically relevant metabolic pathways (Carlson and Srienc, 2004), to study metabolic network properties (Stelling et al., 2002; Behre et al., 2008) and to optimise a microorganism with respect to the production of a certain metabolite (Trinh et al., 2008). Carlson and Srienc (2004) were able to find four unique pathways in *E. coli*, that most efficiently convert glucose and $O_2$ into new cells and maintain energy under any level of $O_2$ limitation. In the case of plant models, Poolman et al. (2003) applied EM analysis to a model of the Calvin cycle to study light/dark metabolism. Similar work was also done on a TCA cycle model of plant mitochondria to describe its physiological properties (Steuer et al., 2007).

## 2.1.6 Linear Programming/Flux Balance Analysis

Linear programming (LP) is a method to find an optimal solution to an underdetermined set of linear equation given a objective function $f$ and set of constraints $S$ (Orth et al., 2010). For a metabolic system at steady state, the equation Nv=0 forms the set of linear equation for the unknown variables V. Then define additional constrains or flux constrainsts on v and objective functions. Mathematically, it identifies optimal solution to the Equation 2.1.5. The objective function is:

$$\text{max or min} \quad f(V) \quad \text{subject to} \quad V \in S \tag{2.1.9}$$

Where, $f$ is the linear objective function and $S$ is the imposed constraint.

It assumes the system is at steady state and obeys the law of mass conservation (Fell and Small, 1986; Watson, 1986; Varma and Palsson, 1993). In general,

minimisation of total reaction flux or maximisation of overall biomass yield are used as objective functions (Holzhütter, 2006; Varma and Palsson, 1993) and the requirement to synthesize biomass components is used as constraints. The function $f$ is a linear function and the set of constraints, $S$, is described using linear inequalities or equations such as $Av \leq b\ v \geq 0$. The constraints of the inequalities specify a convex polytope over which the objective function is to be optimised. The solution given must satisfy the constraints of the problem and among all the solution that satisfies the constraints, LP will return the highest or lowest objective function value which is referred as *optimal value*. If $V^*$ is a solution to Equation 2.1.9 then $f(V^*)$ is an optimal value.
The objective function as minimisation will favor the flux solutions through those reactions in the network which are collectively contributing to minimise protein investment, and is based on the assumption that cells prefer to minimise the protein investment for growth. The main advantage of such constraint-based analysis is the ability to predict the system behavior using a minimal knowledge base (Smallbone and Simeonidis, 2009). Equation 2.1.10 shows the formulation of a general LP problem.

$$\begin{aligned} &\text{minimise|maximise} && : \ \mathbf{v}_{targs} \\ &\text{subject to} && \begin{cases} \mathbf{Nv} = \mathbf{0} \\ \mathbf{L}B \leq V \leq \mathbf{U}B \end{cases} \end{aligned} \tag{2.1.10}$$

where:

- $\mathbf{v}$ is the vector of reaction fluxes which can be minimised or maximised based on the analysis
- $\mathbf{Nv} = \mathbf{0}$ defines steady-state.
- $\mathbf{L}B \leq V \leq \mathbf{U}B$ defines lower (LB) and upper (UB) bounds for additional constraints such as thermodynamics constraints, demand for biomass production, limits on reaction rates etc.

### 2.1.7 Mixed integer linear programming

Apart from LP, a more advanced concept called Mixed Integer Linear Programming (MILP), can be used for optimisation, in which each variable is associated with both an integer (0 or 1) and a continuous value. The continuous value represent as before a valid solution to Nv=0 and the integer records whether the continuous value satisfies some criterion eg Vi = 0. Constraints can be applied to the integer variables, and optimisation is maximisation or minimisation of a function of them (eg. minimise to number of reactions with v > 0) One of the important application is to use the MILP iteratively to compute qualitatively different solutions to an FBA problem that have the same objective value (Lee et al., 2000). MILP problems however are more complex and computationally expensive compared to LP problems.

IBM-CPLEX is a software package for solving linear optimisation problem and offers libraries to solve linearly or quadratically constrained optimisation problems where the objective to be optimised can be expressed as a linear function or a convex quadratic function. A build-in function of the IBM-CPLEX called *populate* was used to enumerate alternative solution to a given mixed integer linear programming (MILP) problem Detailed discussion of the features of IBM-CPLEX is outside the scope of this thesis and interested readers are referred to the website `https://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/`

As in the case of LP, the first step is to define an objective and constraints to the MILP problem. The objective function is set to all reactions in the model so that the feasible region covers all intersecting modes. The constraints can then be set to one or more target reaction to carry only positive flux. The MILP problem is then solved. If a feasible solution is possible, the current solution is excluded from the feasible space and the problem solved repeatedly until all combinations of solutions are computed. This will generate multiple optimal solutions that satisfy the same constraints either with the same or different objective values. All these alternate solutions can be collected as a dataset for further analysis. Application of MILP is discussed in more detail in Section 6.2.3.

### Application of LP/FBA

Early application of LP to study metabolism was described by Fell and Small (1986) and Watson (1986). The term FBA was later coined (Varma and Palsson, 1993), when the technique was used to study synthesis of biosynthetic precursors in *E. coli*. Since then the concept has been used widely to address various biological questions. It have been used to identify targets for gene deletion (Beste et al., 2007; Jamshidi and Palsson, 2007), product yield for metabolic engineering (Ågren et al., 2013), and to study metabolism of plants and algae (Poolman et al., 2009; Boyle and Morgan, 2009). However, predictions of metabolic fluxes by the models using FBA are sensitive to the structure of the model, choice of objective functions and constraints (Schuster and Fell, 2007; Raman and Chandra, 2009; Yuan et al., 2016). It is also important to check if the predictions reflect the known behavior of the metabolic network by examining whether the pathways of known physiological importance are working as expected. But, this is not always feasible, especially when we are investigating lesser known pathways. To address these issues, data obtained from techniques such as $^{13}C$ Metabolic Flux Analysis (MFA) are being used to complement the FBA results (Masakapalli et al., 2010). Recent advances are also aiming to identify the range of the optimal solution space, instead of just a single solution, using techniques such as Flux Variability Analysis (Hay and Schwender, 2014). More recently, integration of OMICS data to study capacity of metabolic pathways is becoming increasingly successful (Töpfer et al., 2013).

## 2.1.8 Integration of proteomics data

Proteomics, transcriptomics and metabolomics data (omics) can be used to study condition dependent changes into metabolic activity of the organism using metabolic models as a platform. Proteomics and transcriptomics data can give important information about the hierarchal regulation of metabolic flux by representing the control over the maximum activity of enzymes (Yizhak et al., 2010). The metabolic fluxes are the end result of the interplay of gene expression, protein concentration, protein kinetics and metabolite concentrations and thus represent the metabolic phenotype of the organism (Winter and Kromer, 2013). Thus, integration of metabolic models with various genome and proteome data help to improve the prediction of metabolic flux distribution.

Several methods such as rFBA, PROM, eMAT, EFLUX, GIMME, tFBA, cited in the review (Machado and Herrgård, 2014), have been developed over the years each with their own features, but primarily based on the assumption that the rate of expression of genes or proteins is correlated with the flux carried by the reaction associated with the gene or the protein.

The method used in this thesis is aimed at qualitatively comparing proteomics data with metabolic activity and is based on the the assumption that reactions which are associated with proteins present in higher amounts, are more likely to be active *in vivo*. This is done by first identifying the genes associated with the reaction and respectively mapping the amount of protein to the gene. The method is simple to implement and does not require arbitrary user-defined threshold for expression of gene or protein and need not exclude any reaction from the solution space. However the only consideration with the method is to accurately map all the reactions under investigation with the respective gene. This was done based on the annotations derived from the BioCyc database through the PyoCyc module in ScrumPy (see section 2.2.2). In total 4 different categories, in which a reaction could be associated with a gene, were identified, which are discussed below using example from the Arabidopsis model:

1. One to one
   The reaction is associated with only one gene and it is unique to that reaction. (the gene is not associated with any other reaction) eg. SBPase - at3g55800, aspartate semialdehyde dehydrogenase - at1g14810

2. One to many
   The reaction is associated with more than one gene but none of these genes are associated with any other reactions in the model. eg. citrate synthase - at3g58740, at3g60100, gqt-2323, at2g44350, at3g58750, at2g42790 phosphoenolpyruvate kinase- at3g55650, at5g56350, at5g63680, at3g25960, at4g26390, at3g55810, at2g36580, at3g04050, at3g49160, at5g08570, at3g52990, gqt-436, at3g22960, at1g32440, at5g52920

3. Many to one
   The reaction is associated with single gene but the gene is also associated with other reaction or reactions. In other word more than one reactions are associated to a single gene. There is no unique relationship here. eg.

glucose-6-phosphatase, 3-phosphoglycerate phosphatase - at5g34850, chlorophyll synthetase reactions (rxn1f-66, rxn-7674, rxn-7673) - at3g51820,

4. Many to many
   The reaction is associated with more than one gene and any of those genes are also associated with some other reaction in the model. eg. 1 transketolase - at2g45290, at3g60750, gqt-437, transaldolase- at2g45290, at5g13420, at1g12230, gqt-2336

In a case where a reaction is associated with more then one gene, the gene for which its corresponding protein has maximum number of peptide counts was considered. More detailed discussion of its application is presented in Section 6.2.4.

## 2.2 Metabolic Modelling Tools

Computational techniques have are long been used to study biological systems and thus a variety of software and tools are available for metabolic modelling. Most of these are designed to simplify the process of converting specified properties of a metabolic model into mathematical objects and facilitate their analysis. Some of the software such as BIOSSIM (Garfinkel, 1968), METASIM (Park and Wright, 1973), FACSIMILE (Chance et al., 1977), and SCAMP (Sauro, 1993) were designed for kinetic model analysis and could only be used in batch mode. A software application called COPASI (Hoops et al., 2006) also supports GUI usage. METATOOL (Pfeiffer et al., 1999) is one of the earliest tool for structural analysis and has a script-based user interface. Other programs that are available specifically for structural analysis include CellNetAnalyzer (Klamt et al., 2007), Flux Analysis and Modeling Environment (FAME) (Maarleveld et al., 2014). Constraint-based reconstruction and analysis (COBRA) toolbox (Schellenberger et al., 2011) runs in MATLAB although some implementations use Python. A more extensive Python-based tool called ScrumPy (discussed in details below) is used in this thesis.

### 2.2.1 Python Programming Language

Python is a high-level, object-oriented, interpreted programming language and can be used for wide range of scientific and numeric computing. It is syntactically clear and easy to learn compared to other programming languages and is developed

under an open source license, making it free to use and distribute. It has large standard libraries such as *SciPy*, a collection of packages for mathematics, science, and engineering and *Pandas*, a data analysis and modeling library. The Python Package Index (PyPI) is a repository of software and hosts thousands of third-party modules for Python. Moreover, it can easily be integrated with other programming languages such as C, C++, Java and packages from their libraries. It supports multiple programming paradigms such as object-orientation (Lutz and Ascher, 1999), which allows collection of data into objects or class instances. The data collected in objects is referred to as attributes, objects also store functions that usually perform actions on the attributes. Objects can inherit properties from different classes and process differently depending on their data type or class, thus promoting a highly modular and structured management of information. More detailed documentation on Python (Lutz, 2001) can be found at Python official website `http://www.python.org/`

### 2.2.2 ScrumPy

ScrumPy is a metabolic modelling tool and is collection of programs written in Python for analysing metabolic models (Poolman, 2006). It is released under the GNU Public Licence, and is available to download from `http://mudshark.brookes.ac.uk/ScrumPy`. It has a number of packages and modules to build models and perform structural and kinetic analysis. The model descriptions in ScrumPy are intended to be simple and can readily be understood with basic knowledge of biochemistry and computing. In addition, the functions in ScrumPy can easily be modified or extended to suit the purpose of analysis. Pre-written standard modules or packages from Python's standard libraries such as SciPy, NumPy can also be used in conjunction with existing packages in ScrumPy.

The *PyoCyc* module in ScrumPy helps building the model by extracting information from BioCyc PGDBs (more details on Section 3.2.1) and mapping the annotations such as relationship between genes, proteins, enzymes, reactions and metabolites. All features in the ScrumPy can either be used interactively on the command line user interface, build on the standard Python console or in batch mode, which promotes re-usability of stored programmes. The deliberate absence of a conventional graphic user interface (GUI) for ScrumPy is to allow users a flex-

ibility to extend the programs to tailor their needs in specific scientific or research contexts.

Metabolic models in the ScrumPy are defined using simple, plain-text (ASCII) files, and are saved with file extension *.spy*. The model file usually have four distinct types of entries:

1. Directives : These are optional instructions to specify how the model description is intended to be processes. Some commonly used directives are *Structural(), to specify a structural model,* External ('metabolite id'), to specify external metabolite, *Include ('FileName')*, to load multiple modules which is a particularly useful feature in ScrumPy to modularise bigger models (see Figure 3.3).

2. Comment : Is single line statement precceded by hash (#), any text or characters written after the hash are completely disregarded while processing, but anything before the hash on the line is regarded a part of the model.

3. Reactions : Is mandatory content and describe the properties of reactions including their unique identifier, metabolites and corresponding stoichiometric coefficient on left hand side and right hand side, reaction directionality and kinetic rate law (replaced by tilde (∼) for structural analysis which specifies a default rate law with mass action kinetics with rate and equilibrium constants of one)

4. Initialisation : (only in kinetic analysis) to assign numerical values to parameters and initial metabolite concentrations

Once installed, ScrumPy can simply be run by typing ScrumPy in a shell. Presaved models can then be loaded by creating an instance of a model object with name of the model being the name of the file present in the same directory that ScrumPy is started from. If no name is provided, a GUI for selecting a model is launched. An example of a typical ScrumPy model is shown below:

```
DeQuote()              # remove quotes from identifiers
Structural()           # skip kinetic processing
External("WATER", "PROTON")  # treat water and protons as external
```

```
R 1:              # unique reaction identifier
   2A <> 1B       # stoichiometry of reaction
 ~    # end of reaction
```

Once loaded all the methods of the model object can be seen by *dir(m)*, where 'm' is the identifier that has been assigned to the model instance. ScrumPy commands for some of the common operations are listed below.

```
>>> m.sm        # returns stoichiometric matrix of the model
>>> m.sm.NullSpace() # returns nullspace of the stoichiometric matrix
>>> m.EnzSubsets()     # returns enzyme subsets
>>> m.DeadReactions()     # returns all dead reactions
>>> m.ElModes()     # computes elementary modes
>>> lp = m.GetLP()    # generates basic LP object assigned to lp
>>> ds = DataSets.DataSet() #generates a empty dataset assigned to ds
                            DataSets is a class of Data package
```

Most of the common model analysis tasks are available as attributes of the model and can be explored by typing:
>>> dir(m)

Similarly, all attributes of objects can be listed using the function *dir()*. For example dir(lp) will list all functions that can be used with *lp* instance. Besides all the built-in attributes of ScrumPy, analysis specific functions can also be written. ScrumPy modules with functions used for analysis presented in subsequent chapters can be found in Appendix. Following chapters will describe various analysis using the methodology discussed in this chapter.

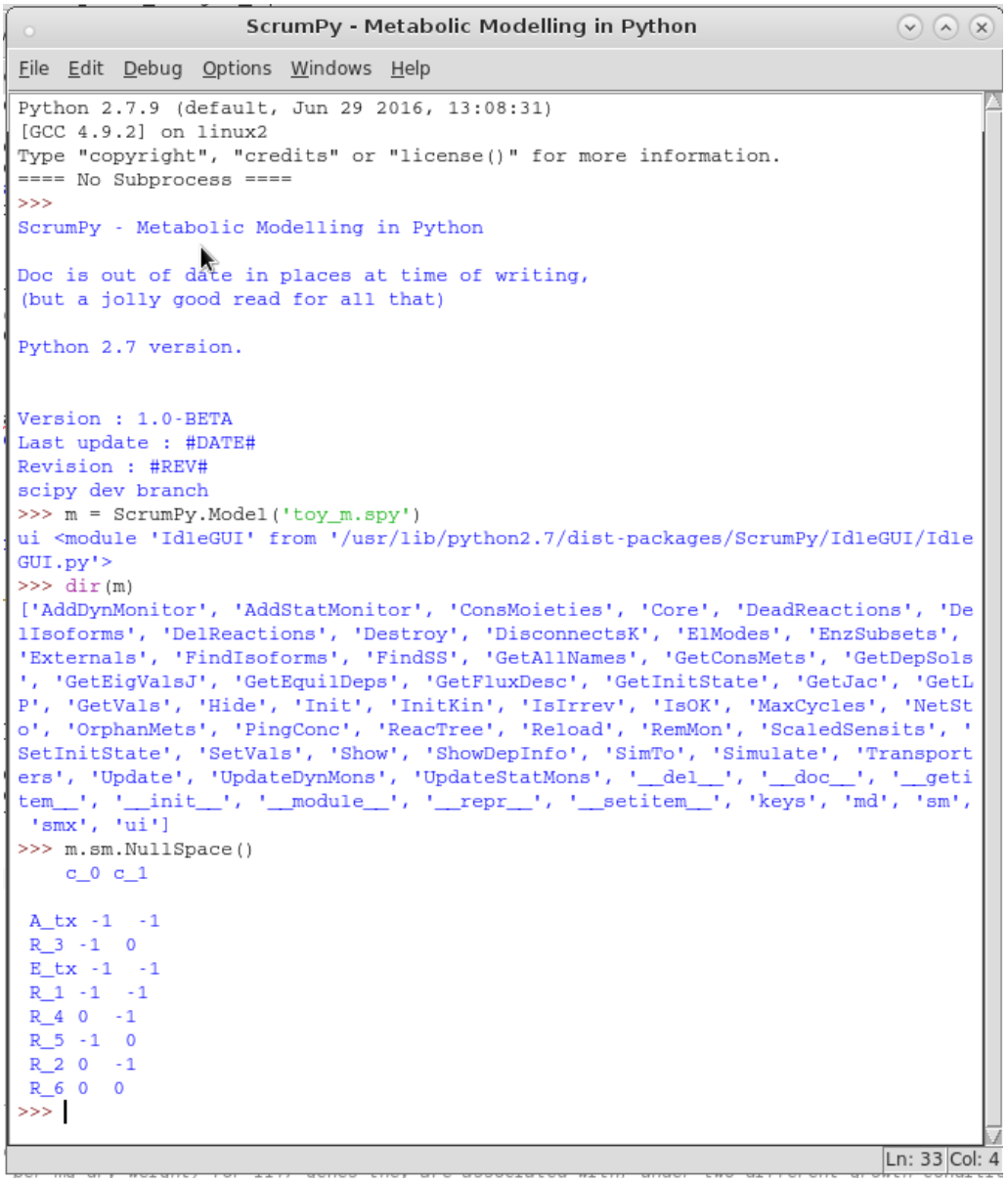



Figure 2.4: The ScrumPy window showing a loaded model assigned to instance *m*. A nullspace of the model is computed by typing m.sm.NullSpace().

# Chapter 3

# Construction of genome scale metabolic models

## 3.1 Introduction

The aim of genome scale metabolic modelling is to represent the entire metabolic capabilities of an organism. A genome scale model (GSM) is typically built from data extracted from annotated genome databases (Fell et al., 2010). GSMs have a wide range of applications such as to analyse high throughput experimental data, investigate metabolic interactions, predicting organism phenotypes, guide metabolic engineering strategies and direct hypothesis-driven discoveries (Feist and Palsson, 2008; Oberhardt et al., 2009; Thiele and Palsson, 2010). Current advances in genomics and genome sequencing have encouraged the reconstruction of GSMs, which play an important role in the emerging field of metabolic modelling. The process of construction typically involves the functional annotation of the genome, identification of associated reactions, determination of correct stoichiometry, assignment of reactions to specific compartments, definition of intercompartmental transporters, definition of biomass composition and estimation of energy requirements. This chapter first presents an overview of the construction and validation steps to create a biologically relevant GSM and then describes the properties of newly constructed GSMs of *A. thaliana* and *C. reinhardtii*.

### 3.1.1 Approaches for construction of models

Over the last decade, several manual, automatic or combined approaches of constructing metabolic models have been used. Manual methods give the maximum control over the structure of a model and result in better accuracy for their intended use. However, the process is extremely time consuming, inconvenient and impractical, especially for large-scale models. Thus, in order to overcome these issues, various automatic processes have been developed (Schmidt et al., 2015). Nevertheless, completely automatic processes are likely to miss important metabolic capabilities of the organism or include wrongly identified functionalities (Orth and Palsson, 2010), and so some level of manual refinement and optimisation remains necessary (Henry et al., 2010; Thiele and Palsson, 2010).

Reconstruction approaches can also be distinguished as `top-down` or `bottom-up` approaches. The top-down approach involves breaking down of high level structures to subsequent smaller or more detailed units. A coarse model is first created, using all available information and is gradually curated to develop a more refined model (Shahzad and Loor, 2012; Çakr and Khatibipour, 2014). In contrast, in the bottom-up approach, individual components are first defined in detail and scaled up by linking them together. Since a high level of detail is provided for smaller components, in the bottom-up process it is easy to test the models from the beginning allowing a better control over their components and coverage. Moreover, since living organisms share the same evolutionary history, many fundamental physiological and biochemical components are common in them and, thus, can potentially be reused between multiple models.

### 3.1.2 Inconsistencies in models

With ready availability of genome databases and semi-automated tools, the number of reconstruction of GSMs for various organisms is increasing. However the accuracy and predictive power of reconstructed models are limited by inherent inconsistencies and gaps (Schmidt et al., 2015). These discrepancies are caused either by the problems in the databases that are used for reconstructions or by the use of untested method or tools used for reconstructions. Some of the problems with databases include incomplete or imprecise genome annotation, presence

of non-functional reactions or non-standard naming of reaction and metabolites. These problems often entail manual re-encoding of data during the model reconstruction process, which is time-consuming and error-prone. On the other hand, lack of detailed documentation on the published models makes it difficult to reproduce reported results. In some cases the old methods and tools are no longer supported and are not available thus the data or tool become unusable and the information is lost.

### Recent development in reconstruction process

The metabolic modelling community is now coming together to design new approach to address many such issues(Fell et al., 2010; Henry et al., 2010; Soh and Hatzimanikatis, 2010). Efforts are being made to unify the format of genome annotations, available from various databases, so that the biochemical information can be created and stored systematically. A method named model SEED is a useful method for detection and corrections of gaps from the models reconstructed from automatic and or semi-automatic processes. It also describes an automatic model reconstruction process, starting from a new genome sequence and finishing on a functional genome-scale metabolic model (Kumar et al., 2007; Henry et al., 2010). Similarly, Thiele and Palsson (2010) published a protocol to generate large scale metabolic models using publicly available genome annotations. More recently, a tool named COMMGEN has been developed to automatically identify inconsistencies between coexisting models and to semi-automatically resolve them (van Heck et al., 2016). The BiGG (Biochemical, Genetic and Genomic) (King et al., 2016) knowledge base is a collection of more then 75 manually curated genome-scale metabolic network reconstructions, which use a set of standarised identifiers called BiGG IDs. Genes in these models are mapped to NCBI genome annotations and metabolites are linked to databases such as KEGG or PubChem. Their web based service also provides an application interface to access and analyse the models. Similarly, MetaNetX (Moretti et al., 2016) is a storehouse of GSMs from different resources imported into a common name-space of chemical compounds, reactions, cellular compartments and proteins. The web based interface of MetaNetX provides a platform to experiment with the existing models or to compare them with new models and analyse them using flux balance analysis. The models constructed in this thesis use the annotations from the BioCyc database (discussed in details

below) for consistency.

### 3.1.3 GSMs of *A. thaliana* and *C. reinhardtii*

#### GSM of *A. thaliana*

Creating metabolic networks to represent plant species is increasingly popular and some metabolic models exists for a few of the plant species. Specially due to the availability of detailed genome annotation of the model plant *A. thaliana* first published in 2000 (Arabidopsis Genome Initiative, 2000), there has been a growing interest in reconstructing its genome scale metabolic model in recent years. There are in total 8 published models for *A. thaliana* to date (Poolman et al., 2009; Radrich et al., 2010; de Oliveira Dal'Molin et al., 2010; Saha et al., 2011; Mintz-Oron et al., 2011; Cheung et al., 2013; Arnold and Nikoloski, 2014; Seaver et al., 2014). Although some of these models uniquely capture specific phenotypes of the organism, all of them vary in their biochemical representation, genome annotations and algorithms they use to build the models. The first GSM of *A. thaliana* was published by Poolman et al. (2009) and represented a heterotrophic plant cell. The model was based on annotations from AraCyc database version 4.5 (Mueller et al., 2003) and covered a network of 1406 reactions with 1253 metabolites. AraGEM (de Oliveira Dal'Molin et al., 2010), which included both photosynthetic and heterotrophic metabolism of *A. thaliana* contains 1526 reactions in 5 different compartments. Both the models, presented by Radrich et al. (2010) and Mintz-Oron et al. (2011), were based on annotations obtained from KEGG. However, a unique feature of model presented by Mintz-Oron et al. (2011), was that it was a tissue-specific model and the information for tissue specificity and compartmentalisation were computed automatically using a network based approach. Among all the models discussed above, three of them, (de Oliveira Dal'Molin et al., 2010; Saha et al., 2011; Cheung et al., 2013) contain the same 5 sub-cellular localisations but the model presented by (Cheung et al., 2013) has a larger coverage of the metabolic network with 2769 reactions and 2618 metabolites.

#### GSM of *C. reinhardtii*

Similarly, more than 6 GSMs for *C. reinhardtii* have been published (Boyle and Morgan, 2009; Manichaikul et al., 2009; Chang et al., 2011; de Oliveira Dal'Molin

et al., 2011; Chapman et al., 2015; Imam et al., 2015) to date. The first ever algal metabolic model by Boyle and Morgan consists of 484 reactions and 458 metabolites in 3 intracellular compartments. Later, another model named AlgaGEM (de Oliveira Dal'Molin et al., 2011) was published which was more inclusive than the first with 1725 unique reactions and 1862 metabolites and included 2249 gene enzyme reaction associations. This model had 4 compartments and was based on a previously published *A. thaliana* model from the same group (de Oliveira Dal'Molin et al., 2010). The model covered the fermentative pathway for production of $H_2$ and fermentative reactions to produce glycolate and acetate. The model iRC1080 (Chang et al., 2011) was extensively compartmentalised into 10 different compartments where the chloroplast alone accounted for more than 30% of the total 2109 reactions in the model. Although it is bigger in size and includes detailed compartmentation it has many unnecessary reactions like sodium mediated transporters and reactions involving flavoprotein (FAD) complex, which are not known to be present in algae. A more comprehensive model, iCre1355, published more recently (Imam et al., 2015), includes a broad range of metabolic functions in different compartments with 2394 reactions and 1133 metabolites.

## 3.2 Methods - Construction and Validation of GSM

### 3.2.1 Resources for genome annotations

High quality and reliable genome annotation is the primary requirement for the construction of GSMs. Fully sequenced genomes are readily available from various databases, however, functional assignment of the sequenced genome is not easy. Specifically from the viewpoint of metabolic reconstruction, annotation of genes encoding metabolic enzymes is of most significance, notably the stoichiometry of the reactions catalyzed by the enzyme and, where available, the reversibility of the reaction. It must be pointed out that genome annotation itself is a bioinformatics problem and does not form a part of the metabolic model (Hartman, 2013). The databases that have been used for reconstruction of GSMs in this study are discussed in more detail below.

### BioCyc

BioCyc is a compendium of Pathway Genome Databases (PGDBs) for more then 7000 sequenced organisms and is a useful resource for the reconstruction of GSMs. It relies on a frame-based representation scheme where data is represented as *frame* and *slots*. A general class and single object such as pathways, gene, enzyme, reactions etc are represented by frames and properties associated with such objects are represented by slots. Frames are identified by a unique identifier, thus allowing the introduction of relationship between the different instances. The PGDBs are generated by a software package called Pathway Tools (Karp et al., 2011), that predicts the metabolic pathways and missing enzymes in a metabolic network. They are then classified into three tiers based on the amount of manual curation and updates they receive. Factors such as literature evidence and experimental elucidations of each annotation are taken into account for this purpose. Tier 1 PGDBs, such as AraCyc, contain detailed entries of metabolites reactions and pathways based on its genome annotations and receive at least a year of curation by scientists. Tier 2 PGDBs, such as ChalmyCyc, are generated by the Pathway Tools, which predicts their metabolic pathways, and receive some degree of literature-based curation. Tier 3 PGDBs are generated solely from the Pathway Tools and do not receive any manual curation.

### AraCyc

AraCyc is the *A. thaliana* specific PGDB (Mueller et al., 2003) and is maintained in collaboration with TAIR (The Arabidopsis Information Resource). The structures for the genes included in AraCyc are based on TAIR genome annotations. The majority of the pathways found in AraCyc have some measure of experimental backup, but additional, computationally-predicted, pathways are also included to maximize the hypothesis-generating power of the database. All manually added pathways are individually reviewed by curators who add comments concerning their potential validity and the sources of the information. Additional pathways are also imported from MetaCyc (Caspi et al., 2006, 2016), based on curator inference. Additional gene ontology and plant ontology functional annotations derived from experimental data and computational predictions are also listed at the TAIR gene pages which are categorized using a different set of evidence codes. The AraCyc PGDB is updated on a regular basis with each new release containing

Table 3.1: Content statistics of different versions of AraCyc database that were used to reconstruct different updates of *A. thaliana* model discussed in this chapter. Data referenced from PMN archives.

| Release | Reactions | Metabolites | Pathways | Enzyme | Citations |
|---|---|---|---|---|---|
| v4.5 | 1723 | 1956 | 288 | 6214 | 2279 |
| v9.0 | 3320 | 3320 | 502 | 7100 | 3210 |
| v11.5 | 3490 | 2613 | 597 | 9041 | 4798 |
| V13.0 | 3635 | 2802 | 620 | 9995 | 5242 |

Table 3.2: Content statistics of different versions of ChlamyCyc. The version v3.5 was used for construction of GSM of *C. reinhardtii* used in this thesis.

| Release | Reactions | Metabolites | Pathways | Enzyme | Citations |
|---|---|---|---|---|---|
| v2.0 | 1688 | 1130 | 283 | 2157 | 2036 |
| v3.5 | 2263 | 1514 | 349 | 3330 | 2708 |

additional annotations and references to experimental evidence. Table 3.1 presents the details of the updates on number of reactions, metabolites, pathway, enzymes and citations included in past four release of the AraCyc database.

### ChlamyCyc

ChlamyCyc is a result of combination of published genome sequence of *C. reinhardtii* bioinformatics predictions and experimental data from omics experiments (May et al., 2009). The database provides a repository to enable and assist systematic studies of cellular processes in *C. reinhardtii*. Table 3.2 shows the content statistics of latest release of *C. reinhardtii*.

### KEGG

KEGG is another well established database used for reconstruction and adopts its own unique identifiers for reactions and metabolites, providing comprehensive descriptions of cellular metabolism and functions (Ogata H et al., 1999). This is a resource for understanding high-level functions and utilities of a biological cell especially for large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. As of February 2017, KEGG has a collection of 10,438 reactions and 17,901 metabolites for 20,946,404 genes of different organisms.

### BRENDA

BRENDA is a collection of enzyme functional and property data, the majority of which is extracted from primary literature. The contents cover information on function, structure, occurrence, preparation and application of enzymes as well as properties of mutants and engineered variants. It includes more then 83,000 enzymes, defined by International Union of Biochemistry and Molecular Biology (IUBMB) naming convention, taking references from 137,000 literature items (Placzek et al., 2016). The latest updates on the web interface of the database give an overview on biochemical processes and helps to visualise enzyme, ligand and organism information from the biochemical viewpoint.

The three databases discussed here, however differ in the annotations of the reactions and the amount of available data. Detailed statistics of the number of reactions present in each of these databases are presented in Figure 3.1

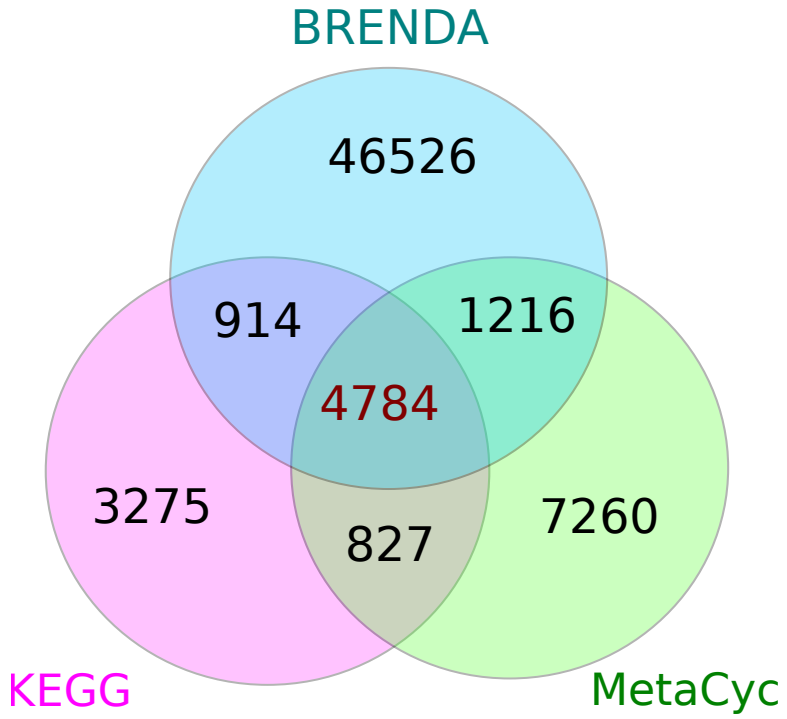



Figure 3.1: Distribution of unique reactions between BRENDA, KEGG and MetaCyc. These data were taken from (Placzek et al., 2016).

### Subcellular localisation databases

Eukaryotic organisms have many cellular compartments, In order to completely understand the metabolic interactions within these compartment each components needs to be represented and studied separately. The information about the compartmentation can be taken from subcellular localisation databases. For example the Arabidopsis Subcellular Database (SUBA) (Heazlewood et al., 2007; Hooper et al., 2017) is a central resource for Arabidopsis protein subcellular location data. Similarly, PredAlgo (Tardif et al., 2012) can be used to predict subcellular location of Chlamydomonas enzymes.

## 3.2.2 Manually added reactions

To reduce the complexity of the model and avoid any inconsistency arising from the reactions involving individual lipid molecules, such reactions were removed from the model to be replaced with a generic lumped reaction that converts acetyl-CoA, NADPH and ATP into palmitate, as a generic fatty acid, as in Poolman et al. (2009, 2013). The reactions are written as:

```
14 NADPH + 8 ACETYL-COA + 7 ATP -> FattyAcid + 7 Pi +
     7 ADP + 14 NADP + 8 CO-A
```

The synthesis of triacyl glyceride was then defined in terms of synthesis of palmitate as :

```
3 FattyAcid + 3 ATP + 1 GLYCEROL-3P -> 1 TAG + 3 AMP + 3PPI + 1 Pi
```

Similarly, to reduce the complexity of representing the chloroplast electron transport chain, two lumped reactions representing the cyclic and non-cyclic photophosphorylation (Poolman et al., 2013; Cheung et al., 2013, 2014) were included in the model as:

```
Light Cyclic Rxn:
7 Photon + 3 ADP + 3 Pi -> 3 WATER + 3 ATP
~


Light NonCyclic Rxn:
14 WATER + 56 Photon + 14 NADP + 18 ADP + 18 Pi ->
```

```
    7 O 2 + 14 PROTON+ 18 ATP + 14 NADPH
~
```

The five complexes used to represent oxidative phosphorylation as represented in Figure 1.6, are shown below. Since protons are considered as external in the model, the inter-membrane translocation proton used to drive the synthesis of ATP was represented as an internal metabolite Pumped-PROTON In total 4 Pumped-PROTONs are required to phosphorylate 1 ADP to ATP. All the reactions are localised in the mitochondria, and, hence the name of reactions and metabolites are suffixed _Mito, in the model.

```
"Complex-I_Mito":
"UBIQUINONE_Mito" + "NADH_Mito" + 5 "PROTON_Mito" ->
"NAD_Mito" + "UBIQUINOL_Mito" + 4 "Pumped-PROTON_Mito"
~

"Complex-II_Mito":
"UBIQUINONE_Mito" + "SUC_Mito" -> "FUM_Mito" + "UBIQUINOL_Mito"
~

"Complex-III":
"UBIQUINOL_Mito" + 2 "Cytochromes-C-Oxidized_Mito" + 2 "PROTON_Mito" ->
"UBIQUINONE_Mito" + 4 "Pumped-PROTON_Mito" + 2 "Cytochromes-C-Reduced_Mito"
~

"Complex-IV_Mito":
4 "Cytochromes-C-Reduced_Mito" + "OXYGEN-MOLECULE_Mito" + 8 "PROTON_Mito" ->
        2 "WATER_Mito" + 4 "Pumped-PROTON_Mito" + 4 "Cytochromes-C-Oxidized_Mito"
~

"Complex-V_Mito":
    4 "Pumped-PROTON_Mito" + "ADP_Mito" + "Pi_Mito" ->
"ATP_Mito" + 4 "PROTON_Mito"
    ~
```

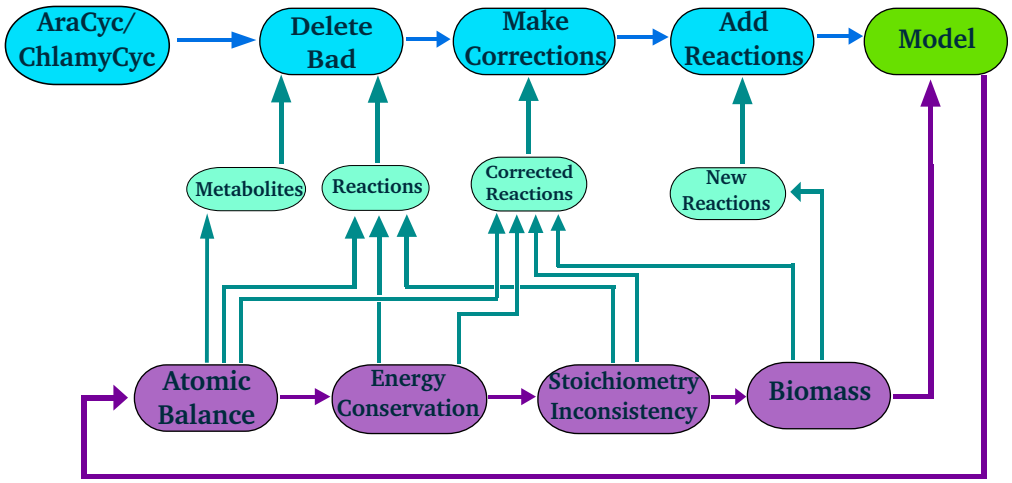



Figure 3.2: Pipeline of construction and validation of GSMs The process shown in blue is the step for construction and that in purple is for the curation which goes through a series of iterations before coming up with a updated model.

### 3.2.3 Construction of GSMs

This section will discuss in details the reconstruction work flow of the GSMs based on the standard practices in the Cell Systems Modelling Group. The approach of reconstruction involves both automatic and manual approaches with an intention of creating more accurate models in a reasonable time. Iterative processes of curation and validation are followed until a biologically suitable model is achieved. Figure 3.2 shows the connection between different components of construction and validation, each of which are discussed in detail below.

#### Extracting information from the database

The construction process begins by creating a draft model using annotations from organism specific BioCyc database. The PyoCyc module from the ScrumPy is used to extract the information from the database. As shown in the example below, the database entry for each reaction contains information about their gene associ-

ation, EC numbers, reaction direction, synonyms or common names, type of the molecule, pathways, enzymatic reaction etc. However, only specific details such as unique id of the reaction, metabolites consumed and produced, their stoichiometries and the directionality of the reaction are necessary to define a reaction. Other specifications are referenced to develop additional understanding of the reaction during the analysis of the model, but do not form a part of the model. Annotations for the SBPase reaction are presented here to illustrate how they are stored in the BioCyc database.

```
UNIQUE-ID - SEDOHEPTULOSE-BISPHOSPHATASE-RXN
EC-NUMBER - EC-3.1.3.37
REACTION-DIRECTION - LEFT-TO-RIGHT
PHYSIOLOGICALLY-RELEVANT? - T
SYNONYMS - Sedoheptulose-1,7-bisphosphatase
IN-PATHWAY - CALVIN-PWY
ENZYMATIC-REACTION - ENZRXNQT-9516
TYPES - Small-Molecule-Reactions
TYPES - Chemical-Reactions
LEFT - WATER
^COEFFICIENT - -1
LEFT - D-SEDOHEPTULOSE-1-7-P2
^COEFFICIENT - -1
RIGHT - |Pi|
^COEFFICIENT - 1
RIGHT - D-SEDOHEPTULOSE-7-P
^COEFFICIENT - 1
```

The unique identifier of each reaction is stored in the slot `UNIQUE-ID`. In a case where specific biological name of a reaction is not defined, a unique reaction number such as *RXN-964* is assigned. The metabolites are taken from the slot *LEFT* and *RIGHT* which represents the the metabolite on left hand side or right hand side of the reaction depending upon their direction, which is stocked from the slot *REACTION-DIRECTION*. For a reaction defined as *LEFT-TO-RIGHT* the metabolites in *LEFT* are the substrate and the metabolites in *RIGHT* are its product. The values stored in the slot *COEFFICIENT* is the stoichiometric coefficient of the metabolite in the reactions. After this information is read from

the database, the reactions are represented in a format readable by the ScrumPy. The SBPase reaction shown in the example above will appear in the model as follows:

```
"SEDOHEPTULOSE-BISPHOSPHATASE-RXN_Plas":
1 "D-SEDOHEPTULOSE-1-7-P2_Plas" + 1 "WATER_Plas" ->
        1 "Pi_Plas" + 1 "D-SEDOHEPTULOSE-7-P_Plas"
~
```

The symbol $\sim$ indicates the end of the information for this reaction and can be substituted with other kinetic information for dynamic modelling.

### Assigning reaction directionality

The directionality of most reactions stored in the PGDBs are based on biological evidence, but for some others it is computed by the Pathway Tool. Thus the direction assigned for a reaction in the database does not necessarily imply the physiological directionality of that reaction. Reactions in the BioCyc database are defined in seven different ways, based on whether or not the reaction occurs in a specified direction in a physiological setting. These options and how they are interpreted in the model is shown in the Table 3.3.

Table 3.3: Direction for each reaction as defined in the BioCyc database annotations and their interpretation in the model respectively. Reactions defined as RIGHT-TO-LEFT were regarded as irreversible to preclude inconsistent enzyme subsets.

| **Database annotations** | **Model** |
|---|---|
| IRREVERSIBLE-LEFT-TO-RIGHT | $\rightarrow$ |
| IRREVERSIBLE-RIGHT-TO-LEFT | $\leftarrow$ |
| PHYSIOL-LEFT-TO-RIGHT | $\rightarrow$ |
| PHYSIOL-RIGHT-TO-LEFT | $\leftarrow$ |
| LEFT-TO-RIGHT | $\leftrightarrow$ |
| RIGHT-TO-LEFT | $\leftarrow$ |
| REVERSIBLE | $\leftrightarrow$ |

### 3.2.4 Handling badly defined metabolites and reactions

The metabolites extracted from the database are not always suitable to be used directly in the model. Multiple changes are incorporated into the draft model generated automatically from the database, including the change in names of metabolites and reactions, correcting the stoichiometry, addition and removal of metabolites and reactions that are not suitable for the model etc. Issues that need to be addressed in order to correctly use them in the model are presented here:

#### Removing bad metabolites

The metabolites used by the reactions, as obtained from the database, are defined either as a *compound* or a *class*. The metabolites defined as class such as *Very-Long-Chain-Alcohols*, *Pimeloyl-ACPs*, *trans-2-cis-5-dienoyl-CoA* do not have a specific empirical formula. The presence of these compounds, could cause potential problems during the analysis of the model by causing stoichiometric imbalances. Similarly, some metabolites defined as compound includes metabolites with undefined functional groups such as *alkyl* group in their formula. Such metabolites cannot be assigned a specific empirical formula. All such metabolites are removed from the model except when they are deemed necessary for specific metabolic function. For example reactions involved with ferredoxin and thioredoxin are required in the chlorophyll synthesis pathway and are included in the model although they do not have a defined atomic structure.

Sometimes, multiple names are assigned to the same metabolite. For example, D-glucose-6-phosphate, GLC-6-P, ALPHA-GLC-6-P all represent glucose 6-phosphate and have the same empirical formula. Such a conflict is resolved by choosing one convenient name to substitute for the rest.

In some cases, the cofactors NADP and NADPH are written as *NAD-P-OR-NOP* and *NADH-P-OR-NOP* in the BioCyc database. This is to indicate that the reaction could use either NADP or NAD. In this case, the reaction is split into two using the alternate pair of cofactors each time and suffixed accordingly. For example the reaction using NAD is suffixed as _NAD and so on.

### Removing Isoforms

Some of the enzymes have multiple forms (isozymes) differing structurally in the amino acid sequence but catalysing the same chemical reaction. They are present with different names hence appear in the model as separate reactions. Such reactions are identified and removed to avoid duplication. Two isozymes for the ATP hydrolysis (ATPase) reaction are shown below for example.

```
 "RXN-11109":
    "ATP" + "WATER" <> "PROTON" + "ADP" + "Pi"
    ~
"3.6.3.6-RXN":
    "ATP" + "WATER" + "PROTON" <> "PROTON" + "ADP" + "Pi"
    ~
```

### Removing redox half reactions

Reactions which involve two or more redox-half-pairs and contain explicit electrons, which cannot occur freely, are called redox-half-reactions. In the database, slots that are similar to pathways refer to sub-reactions. Where possible only the main reaction is included in the model and related sub-reactions are removed.

## 3.2.5 Definition of inputs and outputs

A model created from a database contains only the internal reactions, but for it to operate as a functional unit, external metabolites and their transporters need to be defined. Generally, sources of energy, carbon and nutrients are required as the inputs to a model, while biomass components are defined as the outputs. Depending upon the context of the study, photons, starch or acetate can be used as sources of energy or carbon. Additional sources of carbon can include glucose, sucrose or $CO_2$. Inorganic ions such as $NO_3$, $NH_4$, $SO_4$ and Pi are used as sources of nutrients in models. Diffusion of $O_2$ and $CO_2$ are also represented as separate reactions. Additionally, protons and water are also considered external as they can enter or exit from the metabolic system as necessary (Poolman et al., 2009).

#### Defining biomass composition

In order to describe a growing cell is it necessary to define its molecular composition ie. protein (amino acid), carbohydrates (sucrose, starch), nucleotides (RNA, DNA), lipid (fatty acid), cell wall components (lignin, cellulose, xylan), carotenoids and chlorophyll are used as biomass. The composition of each component is determined either from in-vivo experiments or adopted from published data in literature where relevant and available. Each of the components are assumed to be transported by a single reaction and assigned a flux proportional to their experimentally measured composition. To make them easily identifiable during the analysis process, the names of these reactions are suffixed as _bm_tx in the models.

### 3.2.6 Atomic Balance

Every reaction in the model is required to comply with the the law of mass conservation ie. the metabolites in the left hand side of a reaction should be atomically balanced with the metabolites in the right hand side. For this the elemental composition, in terms of carbon, nitrogen, phosphorous and sulfur of the substrates are compared with those of the products for each reaction. Any discrepancies identified are corrected manually, or if they cannot be balanced are removed from the model. Some of the reactions involving metabolites belonging to the *class* categories, such as starch, xylan, cellulose, which are necessary for various metabolic functions in the model, are checked manually for their atomic balance. Usually starch and cellulose are defined to have one hexose subunit, and xylan one pentose subunit.

### 3.2.7 Identifying energetically inconsistent energy producing cycles

All metabolic models should comply with the law of energy conservation, which implies that energy components such as ATP, NADPH, NADH cannot be produced without some external source of energy in the form of oxidisable carbon or photons. Lack of compliance can arise mainly due to incorrectly defined reaction reversibilities in the model such that a combination of two or more reactions can operate as an internal cycle generating energy components. These cycles have to

be identified and corrected where possible, or removed.

For this, a linear programming problem was defined as shown in Equation 3.2.1 and the ATPase reaction was constrained to carry a positive flux, ie. in the direction of consumption of ATP. All other inputs and outputs are temporarily blocked. A feasible solution to such a problem indicates that the reactions in the solution set together can produce spurious energy. Thus the cycle need to be corrected by changing the directionality of appropriate reactions in the solution set. Each of the reactions is checked and corrected manually with reference to literature evidence or other databases such as KEGG and BRENDA. Likewise, the cycles producing spurious NADH and NADPH are identified by using, a temporary, generic NADH and NADPH oxidase reactions respectively.

$$\begin{aligned} &\text{minimise} && : \ \mathbf{v} \\ &\text{subject to} && \begin{cases} \mathbf{Nv} = 0 \\ v_{tx} = 0 \\ v_{ATPase} = 1 \end{cases} \end{aligned} \tag{3.2.1}$$

The objective is to minimise the sum of the total reaction flux, $\mathbf{v}$, subject to the following constraints:

- $\mathbf{Nv} = 0$ defines steady-state, $\mathbf{v}$ is the vector of all fluxes
- $v_{tx} = 0$ indicates zero flux through external transport reactions
- $v_{ATPase}$ indicates positive flux in ATPase reaction. Similarly, NADPH and NADH are assigned a positive flux, when energy balance in terms of these reductants is to be checked.

### 3.2.8 Identifying stoichiometric inconsistencies

As discussed in the section 2.1.4, the left null space of the stoichiometry matrix can be used to identify the *unconserved* metabolites in the model. These arise where a metabolite has been assigned a composition in one reaction that is consistent in the reaction but is different from the composition in another reaction. This potentially leads to a leak, or appearance of matter violating mass conservation. An algorithm based on mixed-integer linear programming (Gevorgyan et al., 2008) was used in ScrumPy to identify unconserved metabolites in the model.

### 3.2.9 Identifying missing reactions

Identifying the missing reactions is key to obtaining a fully functional model. The process involves both manual and automatic steps. If some obvious reactions are missing from important pathways they can be identified from a manual inspection. Disconnected parts of metabolism can also be identified by computing orphan and dead-end metabolites (see section2.1.4). In other cases when a biomass component cannot be produced, its biosynthetic pathway is backtracked to identify missing reactions. New reactions identified as missing are added in the model with reference to published literature or a reliable database. However, it should be noted that adding a new reaction or reactions in the model can bring about new inconsistencies. Thus all the checks should be repeated every time a change is made to the model.

### 3.2.10 Ability of the model to account for growth

The model is checked for its ability to produce biomass components, first individually, and then in combination, to establish that the model can represent the growth of the organism. Equation 3.2.2 is used to test individual production of biomass. The negative value in the biomass transporter indicates that it is exported from the system.

$$\begin{aligned} &\text{minimise} \quad : \quad \mathbf{v} \\ &\text{subject to} \quad \begin{cases} \mathbf{Nv} = 0 \\ v_{i..j} = -1 \end{cases} \end{aligned} \tag{3.2.2}$$

- $v_{i..j} = -1$ indicates a negative flux in individual biomass transporters indicating its export from the system.

Equation 3.2.2 is modified as follows to test the production of all biomass components in combination.

$$\begin{aligned} &\text{minimise} \quad : \quad \mathbf{v} \\ &\text{subject to} \quad \begin{cases} \mathbf{Nv} = 0 \\ v_{i..j} = \mathbf{t}_{i..j} \end{cases} \end{aligned} \tag{3.2.3}$$

- where $v_{i..j} = \mathbf{t}_{i..j}$ indicates proportional flux through all biomass transporters, consistent with biomass composition.

## 3.3 Results

### 3.3.1 GSM of *A. thaliana*

#### General properties

The GSM of *A. thaliana* was constructed from the AraCyc database. The directionality of all reactions in the model was primarily based on information from AraCyc subject to manual curation with cross-reference to KEGG and BRENDA where necessary. The information about compartmental localisation and intercompartmental transporters were obtained from Cheung et al. (2013). The following corrections were made to the model once the draft model was created.

- 537 reactions have their directionality changed, either to correct inconsistent enzyme subsets (see section 2.1.4) or spurious energy producing cycles.
- 1228 bad metabolites were removed from the model.
- 362 reactions were removed during the process of curation to remove isozymes, energetically inconsistent cycles and to correct inconsistent enzyme subsets. These reactions also include 14 pairs of sub-reactions.
- 34 substitutions were made to unify the name of metabolites with the same empirical formula.
- 26 extra reactions were added for gap filling.

Further, the following checks were made on the model that it :

1. Can produce all biomass components, individually and in combination, under phototrophic and heterotrophic conditions in an experimentally measured proportion.
2. Complies with the law of mass conservation, which means all the reaction in the model are stoichiometrically balanced with respect to C,N, S and P and also is free from unconserved internal metabolites.
3. Is energetically consistent and thus ATP, NADPH or NADH cannot be produced or consumed without some external source of energy such as oxidisable substrates or photon.

4. Is free from inconsistent enzyme subsets.

After completing the curation and validation process, the final model consists of 2588 reactions, of which 234 are transporters (55 biomass transporters and 179 intercompartmental transporters) and 2481 metabolites, distributed in 5 different compartments chloroplast, mitochondria, plastid, peroxisome and vacuole. Each compartment, associated transporters and extra reactions added to the model were written as separate modules and connected to the top level module as represented in Figure 3.3.

All the corrections, list of metabolites and reactions removed, all substitutions made during the curation process are presented as Corrections.spy, Unwanted.spy and Substitutes.spy respectively in the Appendix C. More detailed discussion about the photosynthetic properties of model and its responses to different light conditions is presented in Chapter 4.

### Comparison with previous models

A statistical comparison was made between the current version of the model and its predecessors (Poolman et al., 2009; Cheung et al., 2013). Table 3.5 shows the details of the overall reactions, metabolites, gene association and dead reactions in 4 versions of the model constructed form subsequent updates of AraCyc.

It is clear from Table 3.5 that updates in source database affects the size of the resulting model. Thus to investigate if such changes also affect biochemical properties of the model, all four models were constrained to produce same set of biomass components in exactly same proportions, using Equation 3.2.2. The number of reactions in the resulting solution set for each model is presented in Table 3.6.

The total number of reactions required to produce the biomass components varied in each model. Nevertheless, a consistent feature observed in all the models is that only a small fraction (12-15 %) of reactions are required to produce biomass components in an experimentally measured proportions. One of the reason for the difference in number of reactions in the resulting solution is the use of different set of transporters in all four models. This resulted in rewiring of networks in order to produce the biomass components. Another reason is the updates in irreversibility

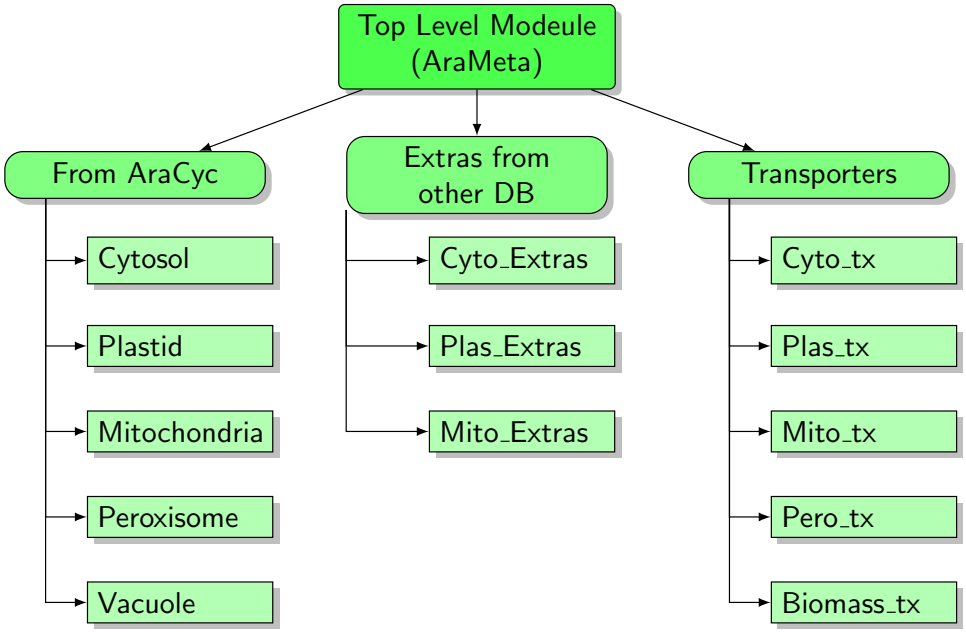



Figure 3.3: Modular structure of components of the GSM of *A. thaliana*. The top level module describes the type of model as Structural and includes the ATPase reaction to represent energy maintenance cost, a list of external metabolites and subsequently imports the rest of the modules included in the model. Each modules has compartment specific reactions suffixed with the first 4 letter of its respective sub-cellular localisation. All associated modules with transporters and extra reactions follow the same pattern.

Table 3.5: General statistics of four models built from different versions of AraCyc database. The number of reactions associated with respective gene is under-determined in v11.5 and v13.0 because (a) manually added lumped reactions has not been accounted for (b) mapping is based solely on the annotations available from AraCyc.

| **Model** | **Reactions** | **Metabolites** | **Reactions with gene** | **Dead Reactions** |
|---|---|---|---|---|
| v4.5(Poolman) | 1408 | 1250 | N/R | 621 |
| v9.0(Cheung) | 2769 | 2739 | 1860 | 1204 |
| v11.5(K1) | 2606 | 2480 | 1807 | 1247 |
| v13.0(K2) | 2613 | 2499 | 1827 | 1269 |

Table 3.6: Statistical comparison of unique reactions required to produce same set of biomass components by four models built from different releases of the AraCyc database.

| Model | Number of reactions to produce biomass |
|---|---|
| v4.5(Poolman) | 232 |
| v9.0(Cheung) | 353 |
| v11.5(K1) | 359 |
| v13.0(K2) | 408 |

of reactions (resulted from both database updates and curation process) in K1 and K2 models, compartmental localisation (Poolman model has only 2 major compartments), updates in transporters (K1 and K2 models excludes seemingly unnecessary transporters that were included in Cheung model). However, it is important to note that despite the difference in total number of reactions, all models showed similar biochemical properties. Reactions from central carbon metabolism including Calvin cycle, glycolysis and TCA cycle were preserved in solution set of all four models.

These results suggest that the regular updates and improvements in the source database has an influence on the structure and hence the prediction capacity of the model. Further the analysis has emphasised that, regardless of the updates in the models, their central carbon metabolism remains unchanged which is important component to support synthesis of biomass and peripheral metabolism.

Some of the major improvements and updates in the current version of the model are listed below.

1. The model is constructed based on a more recent release of AraCyc database, version 13.0, with more genome annotations, covering 597 pathways with 9041 enzymes as compared to 502 pathways and 7100 enzymes in version 9.0. It includes large coverage of metabolic networks, hence gives more possibility of developing newer insight about plant metabolic behaviors.

2. Information about the reaction directions was taken solely from the AraCyc database in the earlier models (Poolman et al., 2009; Cheung et al., 2013) But they have been validated and corrected with reference to other databases like KEGG and BRENDA in the current model.

3. In total 14 sub-reactions were identified and removed. Both Poolman and Cheung models include these such sub-reactions in the model.

4. The updated model can simulate photosynthetic metabolism with greater details, with inclusion of updated light reactions.

### 3.3.2 GSM of *C. reinhardtii*

#### General properties of the GSM of *C. reinhardtii*

The GSM of *C. reinhardtii* was constructed using annotations in the ChlamyCyc v3.5 database. Information about the sub-cellular localisation and intercompartmental transporters were taken from Chang et al. (2011). The following corrections were made after the draft model was created:

1. 324 reactions were corrected for their directionality either to correct enzyme subset or energy conservation inconsistencies.

2. 1090 bad metabolites were removed from the model.

3. 285 reactions were removed during the process of curation to remove isozymes, break the spurious energy consuming cycles and correct the inconsistent enzyme subsets. These reactions also include 9 pairs of sub-reactions.

4. 43 substitutions were made in different names of metabolites.

5. 21 extra reactions were added for gap filling.

Further, the model also satisfies all the validation steps enumerated in Section 3.3.1. The resulting model consists of 1858 reactions and 1931 metabolites in 4 compartments cytosol, chloroplast, mitochondria and peroxisome. There are 73 biomass transporters, representing export of biomass component outside the cell and 269 intercompartmental transporters. Between 339 to 341 reactions were required to produce biomass components, depending upon the experimental condition. In total 1152 reactions from the model can be associated with corresponding genes based on annotations in ChlamyCyc.

## 3.4 Determination of biomass composition of Chlamydomonas

The motivation for the measurement of biomass components under mixotrophic condition was based on the understanding, combined form personal communication with experimental partners in the AccliPhot consortium and published literature (Boyle and Morgan, 2009; Johnson and Alric, 2013; Johnson et al., 2014) that the *C. reinhardtii* grown under mixotrophic condition on acetate media produces more biomass compared to its phototrophic growth. Data collected from this measurement have tried to address this question. Further, the biomass data was integrated in the model and its metabolic properties were compared between phototrophic and mixotrophic growth, and will be presented in Chapter 4.

### 3.4.1 Method and results

Biomass composition for the *C. reinhardtii* model under mixotrophic condition was determined using a torus photobioreactor (PBR) (Takache et al., 2012). These measurements complement the experimental data under phototrophic condition that was already available in the group of Prof. Guillaume Cogne, the AccliPhot collaborating partner at the Process Engineering for Environment and Food (GEPEA) Laboratory, University of Nantes, France. At the date this experiment was performed, no published data was available on biomass composition under mixotrophic conditions for *C. reinhardtii* measured from a photobioreactor setup.

The PBR setup that was used to grow the *C. reinhardtii* under mixotrophic conditions is shown in Figure 3.4. A medium rich in acetate and other nutrients was used to support growth, the detailed composition of which is presented in Table 3.7. A specially designed frame with LEDs was used to provide continuous illumination. Various electrodes were fitted to constantly monitor the level of pH, biomass, $CO_2$ and $O_2$. The pH was maintained at 7.5 by continuous supply of $CO_2$ and the temperature was maintained between 23-24 °C. The amount of acetate consumed and the biomass produced (total dry weight), was measured on a daily basis. Other measurements, such as total cell count, pigment contents, average cell size, were also measured daily while the compositions of starch, glucose, lipid and

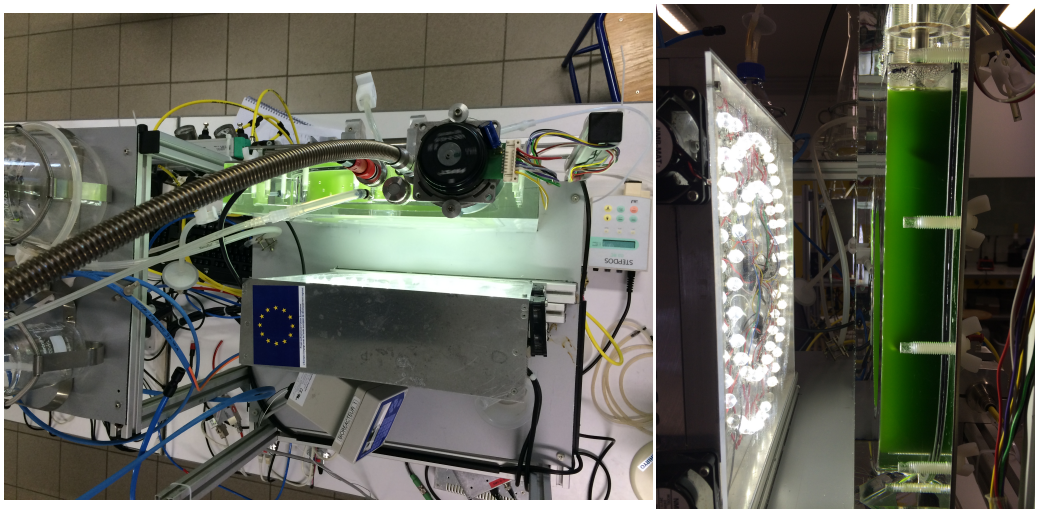


Figure 3.4: Top and side view of the photobioreactor setup. The reactor is fitted with different electrodes to continuously monitor the pH, biomass, level of $CO_2$ and $O_2$. Uniformly fitted led bulbs were used for continuous illumination at 200 uE. A constant flow of liquid medium was maintained by rotating an impeller, thus creating uniform exposure of algae to the light.

overall elemental composition were measured when the final culture was harvested after 6 days of growth.

### Flashing light effect

The movement of fluid between the photic zone, the front surface of the PBR directly exposed to light from LEDs, and the dark zone, the rear surface, was maintained by continuously stirring the medium with a impeller. Such movement of the medium in a PBR is known to create a flashing light effect (Terry, 1986), which has been proven experimentally to be one of the most efficient light regimes in microalgae cultivation, by improving photosynthesis and thus the growth of algae (Grobbelaar, 1991; Liao et al., 2014; Abu-Ghosh et al., 2016). The movement of culture is specially important when the PBR operates in high cell densities such as in the case of mixotrophic growth.

Table 3.7: The composition for 1 liter of the mixotrophic growth medium of *C. reinhardtii*. The compounds were mixed in the order presented above and autoclaved before inoculating with the culture. Final pH of the medium was maintained at 7.5. $NH_4Cl$ serves as the source of nitrogen, $MgSO_4.7H_20$ is the source of magnesium, $KH_2PO_4$ is acidic source of phosphate while $K_2HPO_4$ is also the source of phosphate and is basic. $CaCl_2$ is source of calcium and acetate is obtained from $CH_3COONa$ .

| **Coumpounds** | **Amount** |
|---|---|
| $NH_4Cl$ | 1.2 g |
| $MgSO_4.7H_20$ | 0.3 g |
| $KH_2PO_4$ | 0.092 g |
| $K_2HPO_4$ | 0.4 g |
| $CaCl_2.2H_2O$ | 0.1 g |
| $CH_3COONa$ | 4.2435 g |
| Hutner sol | 1 ml |

### 3.4.2 Comparison of biomass composition

The comparison of percentage fraction of each group of biomass components under phototrophic and mixotrophic conditions is shown in Figure 3.5. The proportion of carbohydrates is about 4% higher under mixotrophic conditions, while the proportion of lipid and protein were about 2-3% higher under phototrophic conditions. Although this difference in composition was minimal, a notable difference was observed in relative growth rate under these two conditions. The growth rate under mixotrophic conditions was 0.0957 $hr^{-1}$ while the growth rate under phototrophic condition was 0.040 $hr^{-1}$.

The composition of each biomass component was obtained in grams % w/w. The proportion for each components were then calculated as described by Cogne et al. (2011). Amount for individual component was calculated using the formula shown below.

$$\mathrm{BMc} = \frac{\mathrm{MF}}{100} * \mathrm{MW} * 1000 * \mathrm{D}$$

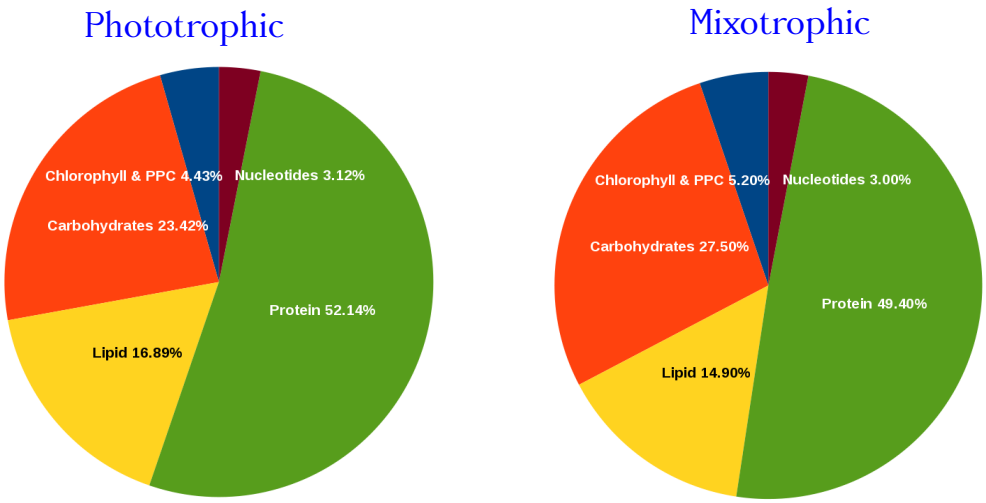



Figure 3.5: Comparison of biomass composition under phototrophic and mixotrophic conditions in *C. reinhardtii*. The measurements were made on culture grown on photobioreactor set up and the normalised molar fraction of each components are considered for comparison.

Where BMc is the composition for individual biomass component, MF is molar fraction, MW is molecular weight and D is the dilution rate or the growth rate. The resulting amount is in the unit mole $kg^{-1}hr^{-1}$

### 3.4.3 Calculating photon stoichiometry

The amount of light exposed to the surface of PBR is known as incident light. Only a fraction of this is used by the organism in photochemistry and is known as stoichiometric photon. Is is dependent upon factors such as cell count per surface area, size and shape of the cells, specific absorption rate, quantum yield of photosynthesis and so on. For the purpose of this thesis, the value was calculated by using a mathematical method developed at the GEPEA (Takache et al., 2012; Soulies et al., 2016). The stoichiometric photon under phototrophic conditions was calculated to be 19.66 mole $kg^{-1}hr^{-1}$ and under mixotrophic conditions was 13.71 mole $kg^{-1}hr^{-1}$.

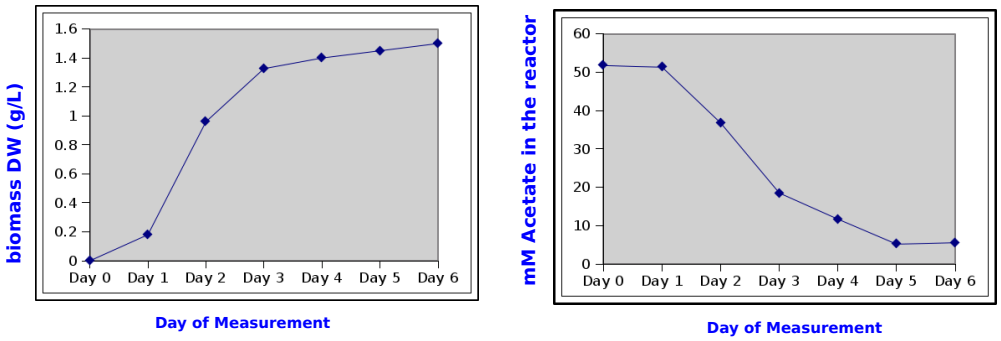



Figure 3.6: Comparison between consumption of acetate per day by the algae in the PBR vs production of total biomass under mixotropic conditions.

### 3.4.4 Calculation of total acetate consumed

The initial culture medium in the PBR contained 52 mM of acetate when the *C. reinhardtii* culture was inoculated. The amount of acetate in the medium was then measured on a daily basis as total organic carbon (ToC), excluding biomass. ToC is calculated by subtracting total carbon from total inorganic carbon in the media. ToC was then used to calculate the concentration of acetate (the formula used is shown below), which was compared against the amount of biomass dry-weight in the culture sample taken at same time. The comparison is presented in Figure 3.6

$$\text{ToC} = \text{TC} - \text{IC}$$

$$\text{concentration of acetate} = \frac{\text{ToC} * \text{Dilutionfactor}}{\text{mol wt of acetate}}$$

## 3.5 Discussion and further developments

Although the updated models cover many details of the metabolic characteristic of the organism, there are some limitations of the model which remain to be ad-

dressed in future developments. Both models constructed here are not balanced in terms of hydrogen and oxygen like their predecessors (Poolman et al., 2009; Cheung et al., 2013; Chang et al., 2011). As such it is extremely difficult to balance protons in GSMs because of the difference in pH between compartments. Some published models (Saha et al., 2011; Mintz-Oron et al., 2011) have dealt with this issue by assuming a constant pH across all compartments but the idea is biologically questionable. Further, both the models are mainly focused on primary metabolism and place less emphasis on secondary metabolic activities. Yet another issue which has not been completely addressed is the sub-cellular localisation and the intercompartmental transporters. Both models have adapted this information from their respective predecessors (Cheung et al., 2013; Chang et al., 2011). However, due to lack of sufficient experimental evidence the information is not detailed and so most of the transporters were added manually to connect the metabolic network. Thus, future developments of the model can be focused on improving these aspects of the models. Nevertheless, the new reconstructions satisfy all theoretical considerations and demonstrate biologically relevant behaviors. The models will be used to study various aspects of photosynthetic metabolism which are discussed in later chapters.

# Chapter 4

# Photosynthetic properties of GSMs of *A. thaliana* and *C. reinhardtii*

## 4.1 Introduction

### 4.1.1 Models to study photosynthetic properties

Knowledge of plant responses to different light conditions provides the opportunity to identify and modify the components involved in photosynthetic metabolism. Genome scale metabolic models offer a platform to predict the behavior of plants and, thus, their responses under different environmental conditions based on the stoichiometry of reactions. Various genome scale metabolic models have been used to study cellular and phenotypic characteristic of plants and algae.

Some of the published models have been used specially to study the change in metabolism under different light conditions. A GSM of rice, representing an expanding leaf, was used to investigate its metabolic responses under different light intensities (Poolman et al., 2013). The model was also used to study the interactions between different organelles and the reconfiguration of the metabolic network to produce biomass when the light intensity increased (Poolman et al., 2014). These studies have proposed an important role of redox shuttling between

compartments under low light conditions and demonstrated that the role of mitochondrial metabolism is likely to change according to the balance between demand and availability of energy. Also, qualitative similarities were reported in interactions between mitochondrial and chloroplastic metabolism when photosynthetic products were exported through phloem and when they were used for growth under increasing light condition. A further development of the model was used to describe the multi-compartmental biochemistry of the chlorophyll synthesis pathway and the effect of light on it (Chatterjee and Kundu, 2015). A separate GSM of rice combined with transcriptomics and metabolomics data characterises the cellular metabolism under a different spectrum of light (Lakshmanan et al., 2015). Similarly, a GSM of Arabidopsis was used to study the change in fluxes through various routes of energy and reductant dissipating reactions as a effect of changing light environment (Cheung et al., 2014, 2015).

Since the full genome sequence of *C. reinhardtii* was published (Merchant et al., 2007), GSMs of the algal species are being constructed. The first GSM of *C. reinhardtii* was used to study metabolic characteristics under autotrophic, mixotrophic and heterotrophic conditions (Boyle and Morgan, 2009). The study concluded that mixotrophic and autotrophic growth were more carbon efficient. Later developments of the model (de Oliveira Dal'Molin et al., 2010) were used to study the pathway producing $H_2$ as a fermentative product, which could be engineered to produce biofuels. A more detailed and compartmentalised model of *C. reinhardtii* (Chang et al., 2011) was used to model the effect of source and quality of light on its growth. The model was also adopted to study the effect of intracellular flux distribution after introducing acetate as source of carbon and energy (Chapman et al., 2015). The latest comprehensive model of *C. reinhardtii* (Imam et al., 2015) was used as a platform to study the effect of varying light intensity in algal metabolism with integration of transcriptomics data.

This chapter will present the study on photosynthetic properties of GSMs of *A. thaliana* and *C. reinhardtii*, whose construction was described in Chapter 3, and their metabolic responses to different light conditions and carbon sources.

## 4.2 Methodology

### 4.2.1 General photosynthetic responses

FBA is used to evaluate the photosynthetic properties of the models. The LP formulation used to investigate dark metabolic fluxes is shown in Equation 4.2.1. The objective was to minimise the sum of total reaction fluxes. In addition to steady state condition (discussed in Section 2.1.6), the necessity to produce all biomass components is set as a constraint for all the analyses.

$$\begin{aligned} \text{minimise} \quad & : \ \mathbf{v} \\ \text{subject to} \quad & \begin{cases} \mathbf{Nv} = 0 \\ v_{\nu} = 0 \\ v_{\text{ATPase}} = \text{ATPase} \\ v_{Starch} = \mathbf{t}_{Starch} \end{cases} \end{aligned} \tag{4.2.1}$$

- Additional constraints $v_{\nu} = 0$ and $v_{Starch} = \mathbf{t}_{Starch}$ were added to the equation to simulate dark metabolism. They represent a flux of 0 to the photon input reaction and a positive flux through the starch transporter respectively.

Equation 4.2.1 was modified as follows to study metabolism under light conditions.

$$\begin{aligned} \text{minimise} \quad & : \ \mathbf{v} \\ \text{subject to} \quad & \begin{cases} \mathbf{Nv} = 0 \\ v_{\nu} = \nu \\ v_{i..j} = \mathbf{t}_{i..j} \\ v_{\text{ATPase}} = \text{ATPase} \end{cases} \end{aligned} \tag{4.2.2}$$

Where,

- $v_{\nu} = \nu$ indicates a positive flux through the photon input reaction.

The cellular maintenance cost, accounted for by the minimal flux carried by ATPase reaction was assumed to be same in all simulation conditions. It is based on the assumption that the changing the biological maintenance cost necessary for biomass synthesis only affects the minimum photon intensity and does not have impact on the organism's responses to higher light intensities.

### 4.2.2 Responses under changing light conditions - light scan analysis

To study the responses of reactions in the model to changing light conditions, equation 4.2.3 was repeatedly solved with increasing input of photon flux at fixed steps, until the the system was saturated and responses of the reactions remain linear. Additional constraints that the cyclic photophosphorylation could not exceed non-cyclic and that the sum of flux through RuBisCO (carboxylase, oxygenase) does not exceed an arbitrary set limit were as described in Poolman et al. (2013). The biomass reactions were constrained to carry flux in an experimentally measured proportion.

$$\begin{aligned} \text{minimise} \quad & : \ \mathbf{v} \\ \text{subject to} \quad & \begin{cases} \mathbf{Nv} = 0 \\ v_{i..j} = \mathbf{t}_{i..j} \\ v_{\text{ATPase}} = \text{ATPase} \\ v_{\nu} = \nu \\ v_{\text{LightNonCyc}} \geq v_{\text{LightCyc}} \\ v_{\text{RubiscoCarboxylase}} + v_{\text{RubiscoOxygenase}} = 0.4 \end{cases} \end{aligned} \tag{4.2.3}$$

Where,

- $v_{\nu} = \nu$ is the photon flux into the system and is gradually increased in fixed steps.
- $v_{\text{LightNonCyc}} \geq v_{\text{LightCyc}}$ constraints that cyclic photophosphorylation cannot exceed noncyclic photophosphorylation
- $v_{\text{RubiscoCarboxylase}} + v_{\text{RubiscoOxygenase}}$ or the sum of the Rubisco carboxylase and oxygenase reactions is set to 0.4 to implement the limit on Calvin cycle flux and reflect the saturation of Rubisco.

All resulting LP solutions were collected as a dataset and used to plot the flux through reactions as a function of the photon flux. Groups of reactions with similar responses were identified.

## 4.3 Results from the Arabidopsis model

Plant metabolism undergoes various changes during day and night. Broadly they can be divided into a heterotrophic phase in the dark and a photosynthetic phase in the light. Metabolic responses under these two conditions are compared on the latest GSM of Arabidopsis.

### 4.3.1 Heterotrophic metabolism

As expected a different set of reactions were active during dark metabolism in the GSM of *A. thaliana* as compared to light metabolism. Central carbon metabolism in the dark is shown in Figure 4.1. During this condition, degradation of starch serves as source of energy by fueling the metabolism. This was evidenced by the flux through the starch degradation reaction. Along with the starch degradation the largest flux was related to the OPPP reactions G6Pdh, 6PGLac and PGluDh. The TCA cycle in the mitochondria was active and was running in cyclic mode using pyruvate produced from the glycolytic reactions in the cytosol. The operation of the complete TCA cycle is attributed to the production of ATP via oxidative phosphorylation, which is evident by the positive flux in all the mitochondrial complex reactions. In the chloroplast, the transaldolase reaction was active while the light specific reactions such as RuBisCO, G3Pdh, FBPase, SBPase, PRK were not active.

### 4.3.2 Phototrophic metabolism

#### General photosynthetic properties

The minimal photon flux at which a solution to the Equation 4.2.3 could be found was 0.32 photon flux unit. Quantum demand, which is calculated as total photon per $CO_2$ fixed at minimum photon flux, is 13.89. The assimilation quotient, calculated as $CO_2$ fixed per $O_2$ released, is 0.97 while photon per $O_2$ released is 13.56. These results are within the range of experimentally observed values and are also consistent with the data reported previously in a GSM of rice (Poolman et al., 2013).

The central carbon metabolism under light conditions is shown in Figure 4.2.

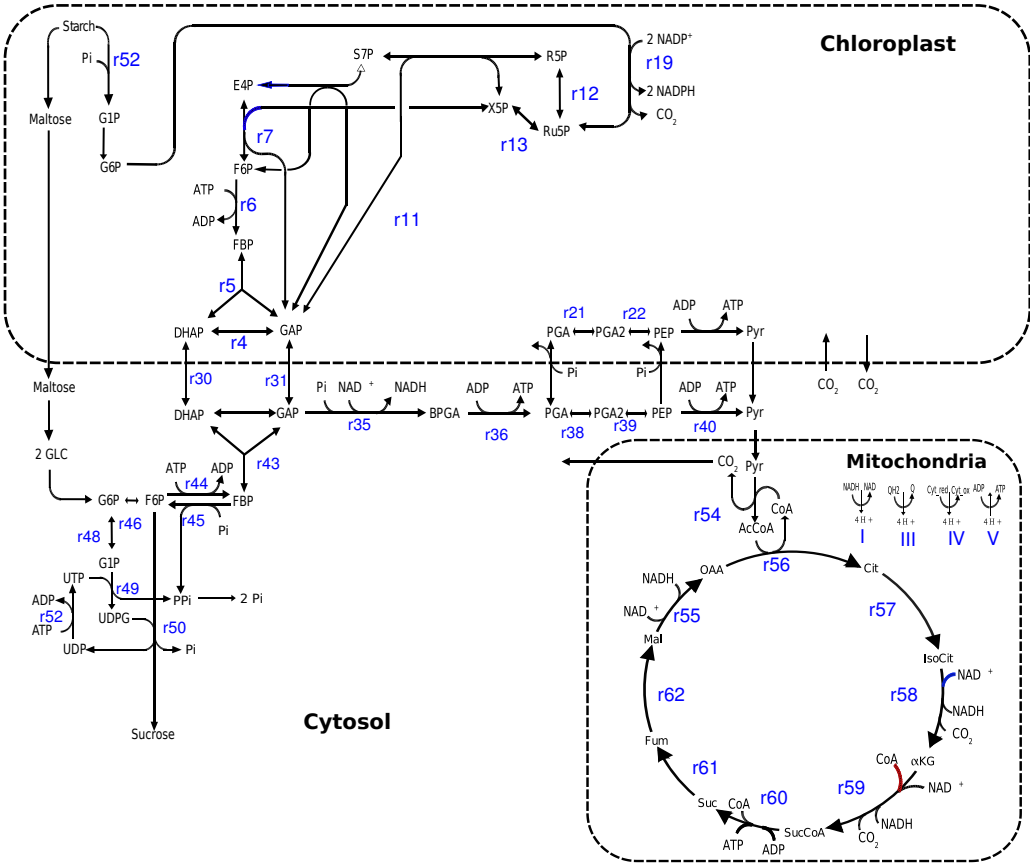



Figure 4.1: Central carbon metabolism of Arabidopsis under the dark conditions as determined from the model. OPPP is active along with the transaldolase in the chloroplast while the complete TCA cycle is active in the mitochondria

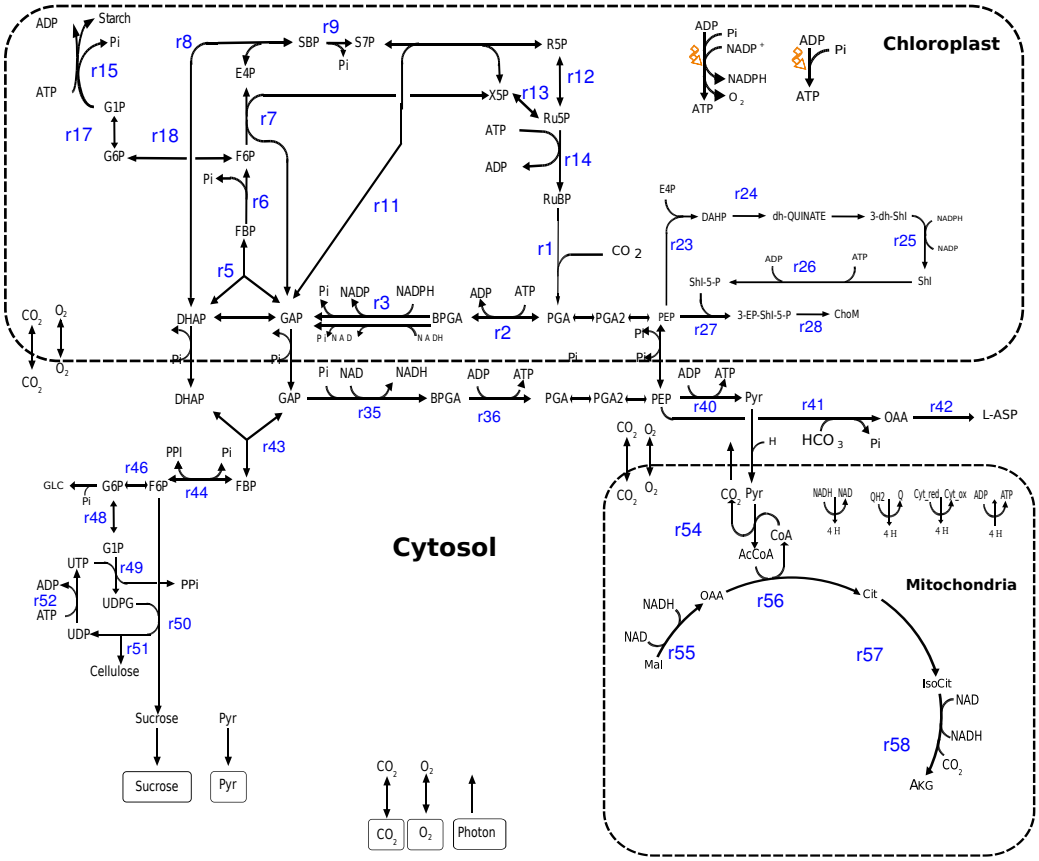



Figure 4.2: Central carbon metabolism of Arabidopsis under the light conditions as predicted by the model. The reductive pentose phosphate pathway was active in the chloroplast and the TCA cycle in the mitochondria was operating in two parts.

In the light phase, there was high flux through the cyclic and non cyclic photophosphorylation reactions which generate the energy and reductants necessary to fuel the metabolism. As shown in Figure 4.2, the reductive pentose phosphate pathway was active and the carbon assimilated by the RuBisCO. The triose and hexose phosphates were exported to the cytosol to support synthesis of other biomass components. The OPPP reactions and the transaldolase reaction which were active in the dark are inactive under the light conditions. The reducing power generated in the chloroplast is also shuttled to the cytosol both by the MAL-OAA shuttle Figure 1.10 and TP-PGA shuttle Figure 1.9.

### Mitochondrial metabolism

In the mitochondria, unlike in the dark, only half of the TCA cycle reactions carry flux Figure 4.2. The pyruvate transported to mitochondria is used to produce the AcCoA which is used by citrate synthase to produce citrate. The $\alpha KG$ produced by the subsequent reactions is then transported to the cytosol. The malate used in the other half of the TCA cycle to produce OAA is shuttled to mitochondria via the MAL-OAA shuttle. The mitochondrial electron transport chain is active but with a minimum flux and the reductants used in the chain come from the activity of pyruvate dehydrogenase, alpha-ketoglutarate dehydrogenase and malate dhydrogenase.

### Lower light conditions

The photosynthetic metabolism starts to change with a small increase in the input of light flux. Most notable is the shift in energy metabolism from mitochondria to chloroplast. The production of mitochondrial ATP starts to decline as the availability of light increases, note the flux of complex I tending near to zero in Figure 4.3. The energy requirement in the chloroplast is now met with the production of ATP and NADPH by cyclic and non-cyclic phosphorylation. Under this condition, there was net evolution of $O_2$ from the chloroplast. The photosynthetic $O_2$ thus released was utilized by mitochondrial respiration producing $CO_2$, which in the model was assumed to diffuse in part back to the chloroplast to be re-assimilated in the Calvin cycle. Demands for energy in the mitochondria is fulfilled by a mitochondrial malate-oxaloacetate shuttle and a shuttle of $\alpha KG$ and GLT (Figure 4.3) which transports the reductants generated at the chloroplast to mitochondria and

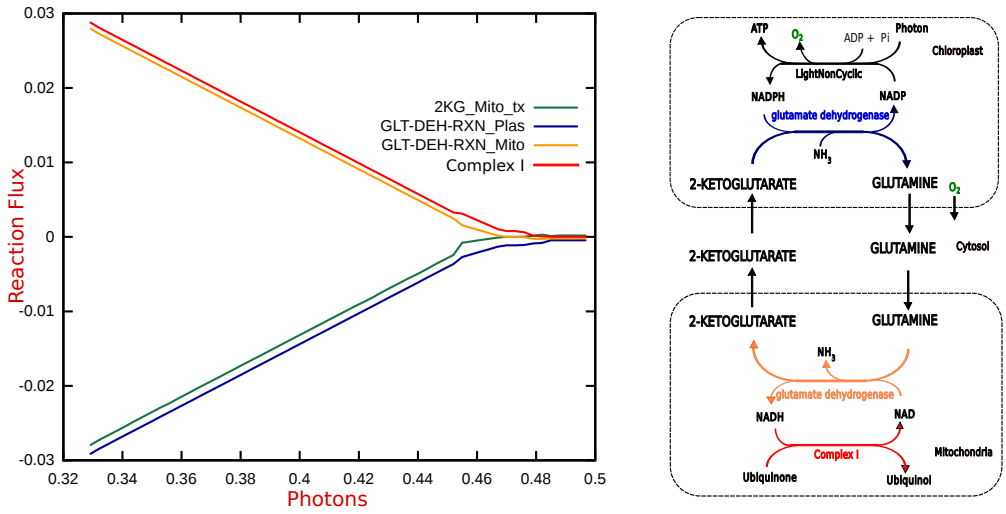



Figure 4.3: Reaction responses under low light conditions in the *A. thaliana* model are shown in left, where, photon flux is represented in the x-axis and reaction flux in the y-axis. The network diagram derived by studying the responses of reactions is shown on the right. The reactions in the diagram are color coded as in the response curve.

provide NADH to power the mitochondrial electron transport chain and oxidative phosphorylation.

### Responses under high light conditions

Using the light scan analysis technique, the reactions that particularly responded to a high input of photon flux were identified. This set of reactions could be categorized into two broad groups, one involved with the central carbon metabolism that contribute to maintaining normal photosynthetic function and the second involved with photo-protection mechanisms.

The response curves of the reactions that show significant change with increasing input of photon flux are shown in Figure 4.4. Only some specific reactions, representing a group with similar responses, are shown for clarity. Based on the activity of reactions, the plot can be divided into 3 distinct regions. Region A is the lower light range with normal photosynthtic metabolism as discussed above. In the region B, as the photon flux increases the reactions involved with the xantho-

phyll cycle start to turn on. The coordinated activity of antheraxanthin epoxidase (represented by light yellow line in Figure 4.4) and violaxanthin deepoxidase form a cycle converting antheraxanthin to violaxanthin while concomitantly converting ascorbate to dehydro-ascorbate. Another set of reactions, glutathione dehydrogenase (represented by green line in Figure 4.4) and glutathione-disulfide reductase (represented by pink line in Figure 4.4) are also active in the same region. This set of reactions complete a cycle of ascorbate-glutathione combined with xanthophyll cycle (see left hand side diagram in Figure 4.5).

In the region C, glutamine synthase (represented by a blue line in Figure 4.4), and carbamoyl phosphate synthase start to turn on with similar flux response throughout the region. These also form a cycle of glutamate-glutamine with the help of carbamate kinase (represented by dark yellow color) and carbonate dehydratase reactions which operate with similar flux values in the region C. The carbamate kinase reaction helps the cycle by producing ammonia by using carbamoyl phosphate produced by carbamoylphosphate synthase reaction. The $CO_2$ produced in the process is recycled back to bicarbonate by carbonate dehydratase reaction (see right hand side diagram in Figure 4.5).

It is clear that the reactions which start to show responses in the high light regions can be associated with their function in photo-protection mechanisms. The function is achieved by a group of reactions operating in a cyclic mode dissipating energy components. However, there is evidence of many other energy dissipating cycles reported to be active under high light conditions which were not seen to show a response in this analysis. Hence, to examine if any other groups of reactions operate in similar fashion under high light conditions, the simulation was repeated after breaking the cycle shown in the region C of Figure 4.4, by setting a flux of zero to the carbamoyl phosphate synthase reaction. Surprisingly, many such cycles, in a group of between 2-4 reactions, were found from subsequent simulations. Two of the subsequent scan responses are shown in Figure 4.6. In total 19 such cycles were identified and are presented in Figure 4.7. After this point the flux at the high light conditions was distributed among a large group of reactions. Thus the analysis was stopped with the assumption that the large group of reactions are less likely to have exactly the same activity *in vivo* to operate like a cycle described here.

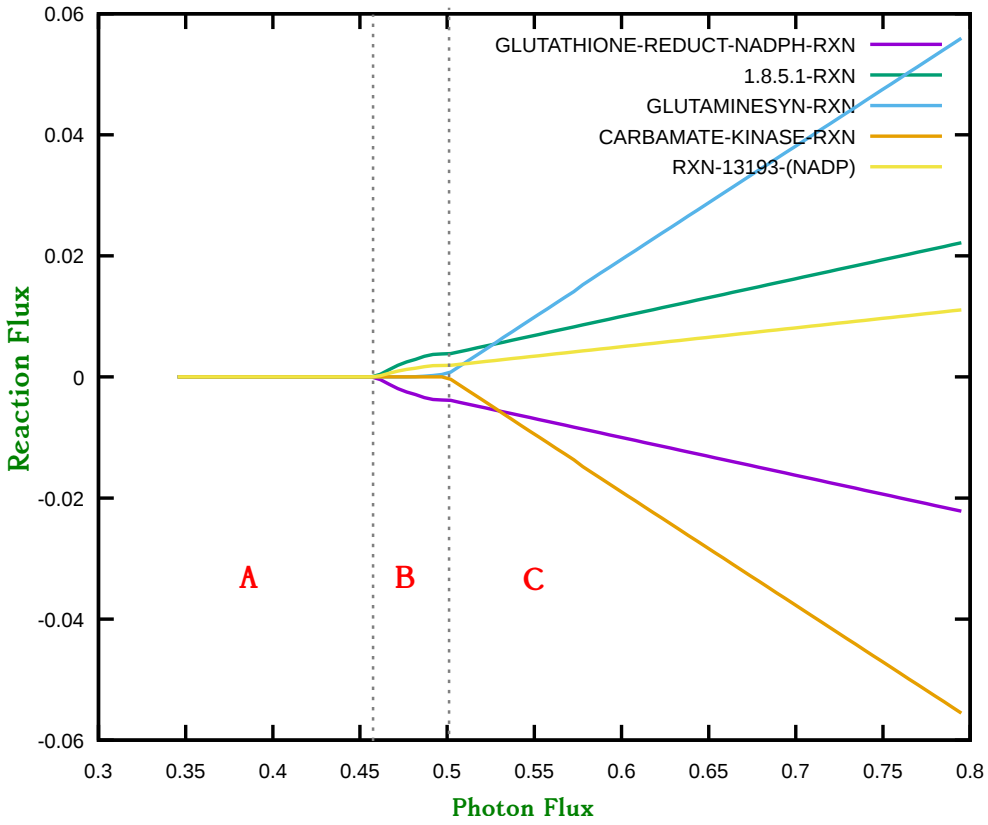



Figure 4.4: Reactions responding to high light conditions. The photon flux is represented on the x-axis and reaction flux on the y-axis. Based on the pattern of activity of reactions the plot can be divided into 3 regions A, B and C. Region A represents normal photosynthetic condition as discussed in Section 4.3.2. In the region B reactions from the xanthophyll cycle show similar flux responses as the reactions involved in the ascorbate-glutathione cycle. The region C represents the responses of reactions belonging to the glutamate - glutamine cycle. The network diagrams of the cycle formed by reactions operating in region B and C are represented in Figure 4.5.

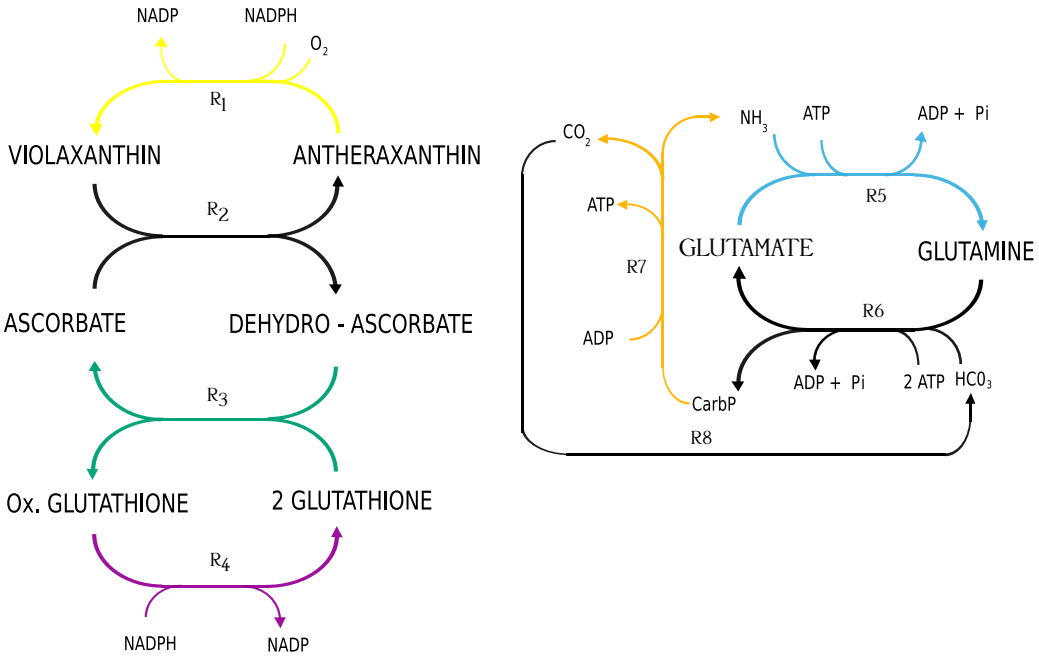
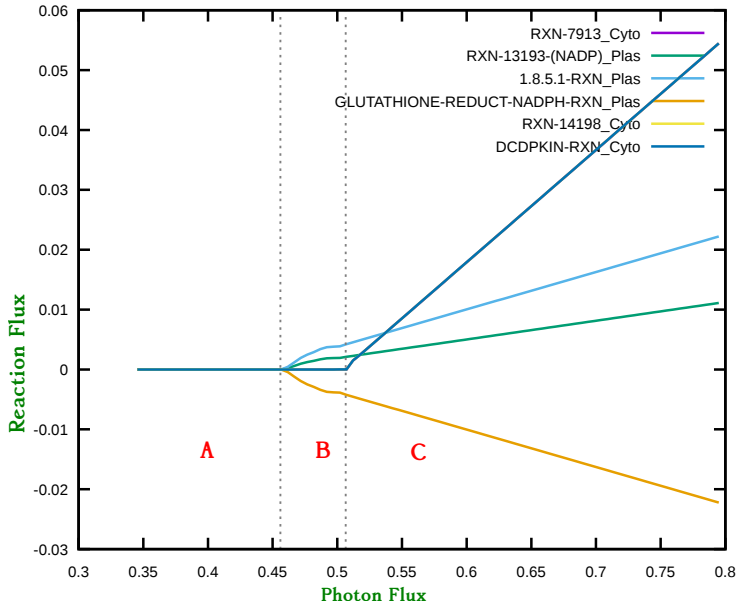
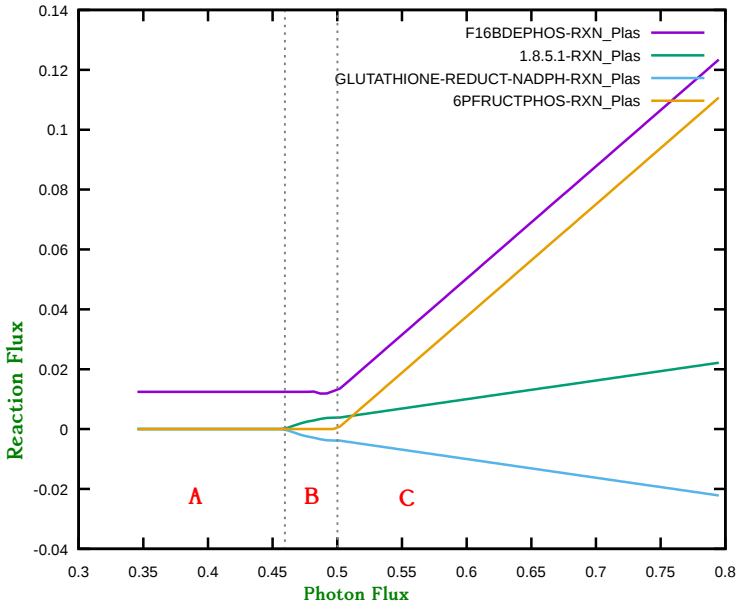



Figure 4.5: Network diagrams of the reactions that respond to high light conditions as shown in Figure 4.4 The network diagram on the left represents the reactions that are active in the region B and the diagram on the right represents the reactions that are active in region C. The reactions in the response curve are color coded to reactions in the network diagram. R1, GlutmSyn; R2, VDE; R3, GluthDh; R4, GluthRd; R5, GlutmSyn; R6, CarbPSyn; R7, CarbK; R8, CarboDht.




(a)




(b)

Figure 4.6: Metabolic responses under high light conditions after subsequent removal of reactions that dominated the response curve and when the photon flux input increased. (a) DCDPKIN and RXN-14199 dominates the response curve after carbamate kinase (see Figure 4.4) is removed (b) FBPase and 6FP dominates the region C after the dominating reactions from (a) are removed.

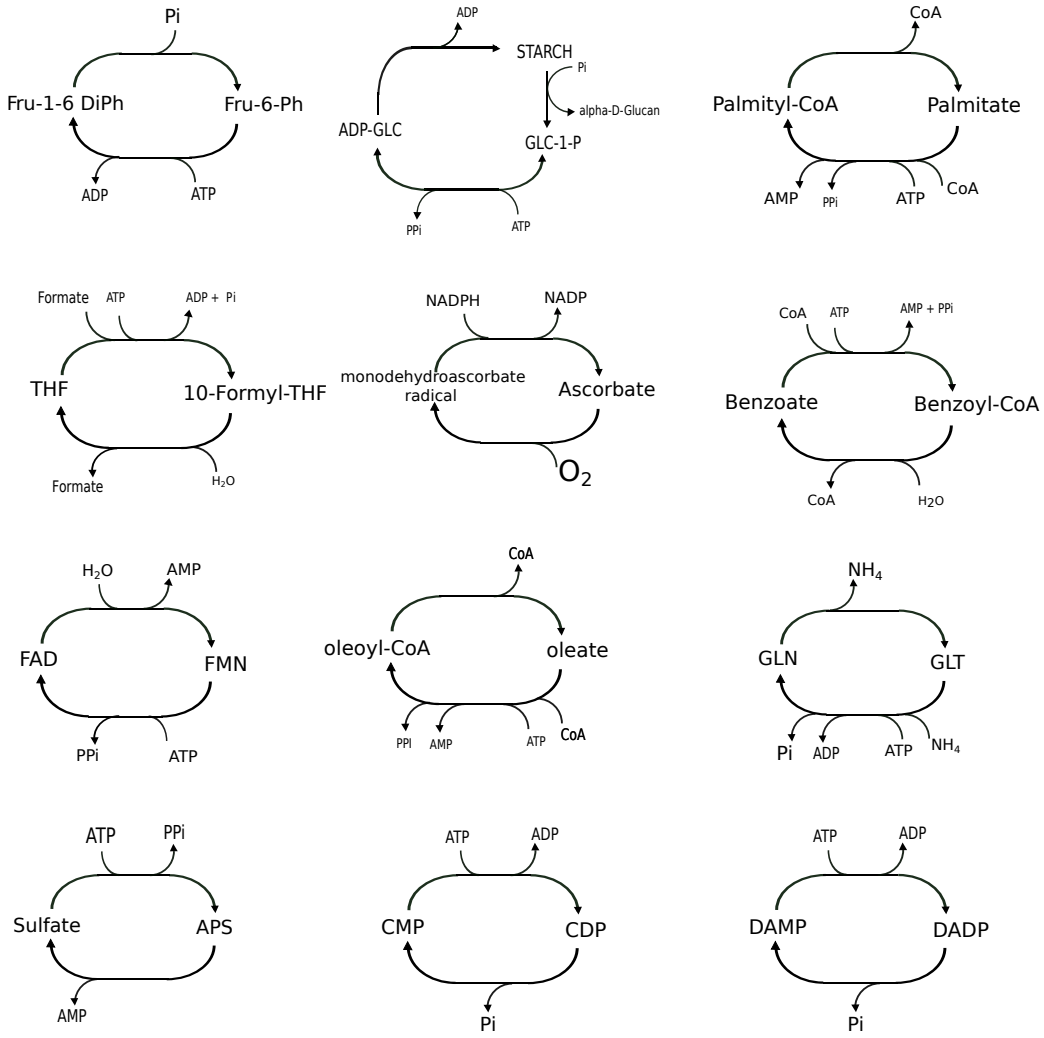



Figure 4.7: The metabolic cycles identified from the light scan analysis. Reactions involved in these modes respond with bigger change in flux under high light input and appear subsequently when an existing cycle is removed.

### 4.3.3 Discussion

#### Dark metabolism

There has been relatively little research on plants under dark conditions and thus the dark metabolism is poorly understood compared to metabolism under light (Wang et al., 2016). The modelling results obtained here could be a guide to further develop our understanding of dark metabolism in plants. The results presented here are consistent with previous experimental observations. For example, measurement of metabolite levels measured from the leaves of spinach, pea, wheat and barley during the night have shown high levels of hexose phosphates in the chloroplast which indicates that the hydrolysis of starch supports the rate of sucrose synthesis in the dark (Stitt et al., 1985). More recently the promotion of starch degradation during the dark is also evidenced by measurement of darkness-responsive proteins in Arabidopsis chloroplasts (Wang et al., 2016). The analysis has demonstrated that the model is able to successfully predict the breakdown of starch to support metabolism.

#### Normal and lower light metabolism

The photosynthetic properties presented here under the light conditions are consistent with various experimental observations and computational analysis. For example the cyclic and noncyclic photophosphorylation are reported to be the major contributors in the steady state photosynthesis of rice leaves (Makino et al., 2002). The cyclic and linear electron transfer in plant leaf is experimentally verified to induce the synthesis of ATP and reductants needed to activate the Calvin cycle (Joliot and Joliot, 2002, 2005). Similarly, the non-cyclic activity of the TCA cycle was studied with stable isotope labelling measurements in photosynthetic leaves, and that showed much smaller flux through the part of the TCA cycle from 2-KG to fumarate (Tcherkez et al., 2009), which is consistent with the model prediction. The observation of recycling of $O_2$ and $CO_2$ and shuffling of reductants between chloroplast and mitochondria, under low illumination conditions was also reported from a similar analysis on a GSM of rice (Poolman et al., 2013).

The oxidation of MAL and PYR to support the ETC, which generate ATP and support cytosolic metabolism, is a more effective way to rebalance energy requirement

in the mitochondria than in the chloroplasts. The mitochondria can generate more ATP for an electron pair between $H_2O$ and NADPH compared to the chloroplastic photophosphorylation. As the availability of light increased the ATP is provided by the activity of photophosphorylation thus showing a clear shift in metabolism to the chloroplast with the use of the alternative oxidase. This observation supports the idea that the mitochondrial metabolism has a role in optimizing the photosynthetic performance (Padmasree et al., 2002). Moreover, the results also show that the metabolic interactions between different compartments in the light are not fixed but change according to the conditions.

### Metabolism under high light conditions

The cycle of violaxanthin-zeaxanthin also called as the xanthophyll cycle is known to participate in non-photochemical quenching. It has been reported previously to be involved in protecting plants under stress conditions (Goss and Jakob, 2010; Maxwell et al., 1999). From the modelling results it is clear that the set of 4 reactions have similar responses to the increasing input of photon flux, indicating that these set of reactions have similar biological properties. As shown in the Figure 4.5, the combination of 4 reactions operate with a net oxidation of 2 NADPH, which can be attributed to its function to dissipate reducing energy which is in excess under the high light conditions. However, it is important to note that, biologically, the interconversion of violaxanthin to antheraxanthin is a very slow process. Moreover, the non-photochemical property is due to the activity of violaxanthin itself and not the interconversion between violaxanthin and antheraxanthin. Thus in the modelling context, since the set of 4 reactions operate together and their net stoichometry oxidises 2 molecules of reducing equivalent, the unique combination of these reactions forming the cycle is reported as a energy dissipating cycle.

Similarly, the concomitant response of glutamine synthase and carbamoyl phosphate synthase along with carbamate kinase and carbonate dehydratase also operate as a cycle with net hydrolysis of 2 ATP. The fact that the cycle is active in the high light range can be attributed to its function in photo-protection. There is experimental evidence of the function of glutamate-glutamine cycle in assimilation of ammonia in higher plants (Masclaux-Daubresse et al., 2006) but the combination of these reactions operating together for a photoprotective function has not been

reported previously. Thus the analysis has identified a new function of well known reactions in a photo protective mechanism and also demonstrated the capacity of model to make new predictions.

The observation that a set of reactions show similar responses under high light conditions and operate in a cyclic manner was a motivation to further study energy dissipating metabolic cycles in the model. The detailed discussion of the significance of such cycles is presented in Chapter 6.

## 4.4 Results from the GSM of *C. reinhardtii*

### 4.4.1 Phototrophic metabolism

#### General photosynthetic response

The minimal photon flux that is required by the model to produce biomass components is 16.59 mole $kg^{-1}hr^{-1}$ which is very close to the experimentally measured stoichiometric photon flux of 19.66 mole $kg^{-1}hr^{-1}$. Photosynthetic quotient is 1.14, which is very close to an experimentally obtained value of 1.134 (see Section 3.4) and within the range of values reported for algae (Burris, 1981). Similarly, the quantum demand is 9.94, corresponding to 8.65 photons per $O_2$ released. In total 351 reactions were required for the model to produce all biomass components.

#### Normal light conditions

The central carbon metabolism of *C. reinhardtii* predicted under phototrophic condition is presented in Figure 4.8. As expected, the majority of flux is directed towards the Calvin cycle. Some of the energy generated from the light cyclic and non-cyclic photophosphorylation is utilised in the regenerative limb of the Calvin cycle. Some of the triose phosphate from the chloroplast is tranported to the cytosol, converted to PEP and antiported back to the chloroplast to enter the shikimate pathway and further used in the production of the chlorophylls. Pyruvate in the chloroplast is used in various amino acid synthesis pathways including lysine, cystine, methionine, valine. In the cytosol, pyruvate dehydrogenase converts pyruvate to acetyl CoA, which is then used to produce linoleic acid.

Table 4.1: Summary of energy components being produced and consumed in different compartments in the *C. reinhardtii* model. Production of energy components in the mitochondria is less as compared to that in the Chloroplast.

| Compartment | Cyto | Mito | Plas |
|---|---|---|---|
| ATP | 0.8386 | 0.0022 | 9.4905 |
| NADPH | 1.1855 | 0.0 | 6.4911 |
| NADH | 0.1194 | 0.0059 | 0.0399 |

In the mitochondria, only half of the TCA cycle is operational. The electron transport chain maintains a minimal flux and is fueled by reductants shuttled from the chloroplast. As shown in Table 4.1, ATP produced in the mitochondria is much less compared to ATP produced in the chloroplast. This also support the fact that, under full illumination conditions, the reductants necessary for the mitochondrial ETC comes from the photo-phosphorylation in the chloroplast.

### Responses under high light conditions

The light scan analysis technique was used to study the response of the *C. reinhardtii* model to an increasing input of photon flux, under phototrophic conditions. The analysis shows that reactions belonging mainly to the central carbon metabolism involving the Calvin cycle reactions, nucletide metabolism, glycolysis and amino acid synthesis pathways showed significant changes in activity in response to increasing photons. The other group of reaction that responded were related to the protection mechanism in a similar pattern observed in the *A. thaliana* model.

The response curve for some of the representative reactions is shown in Figure 4.9. Based on the pattern of responses and the point at which a set of reactions turn on or saturate, the response curve is divided into 4 regions. Region A shows the normal photosynthetic activity. The photorespiratory reactions start to turn on in region B and start to saturate at the region C, where the xanthophyll cycle, as discussed above, starts to turn on. The cyclic activity of starch synthesis and degradation reactions, acting as an energy dissipation mechanism, start to turn on in the region D. The recurring pattern of two or more reactions involved in energy

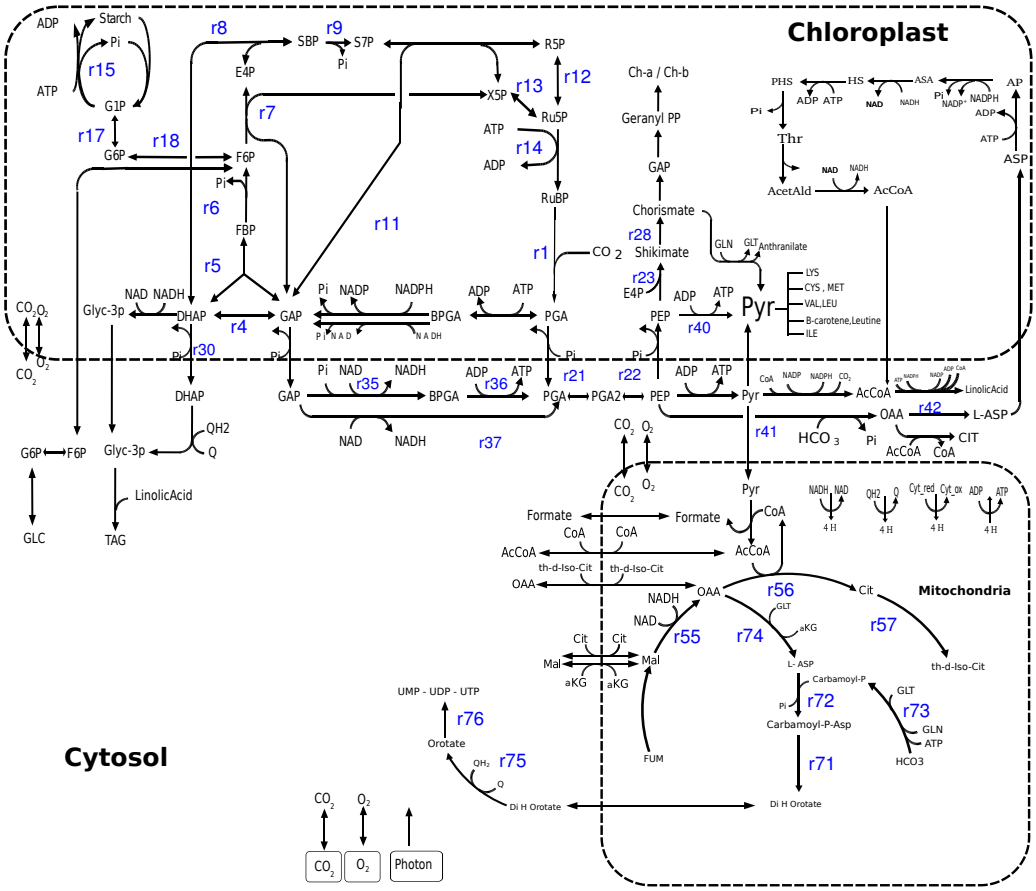



Figure 4.8: Central carbon metabolism of *C. reinhardtii* under phototrophic condition as predicted by the model. The predictions have shown that the $CO_2$ fixed in the Calvin cycle and the triose phosphates are transported to the cytosol to be used to make pyruvate that is ultimately tranported to both chloroplast to be used for synthesis of various amino acids and to the mitochondria to be used in the TCA cycle. The energy required for the metabolism was fueled by the activity of cyclic and non-cyclic photophosphorylation.

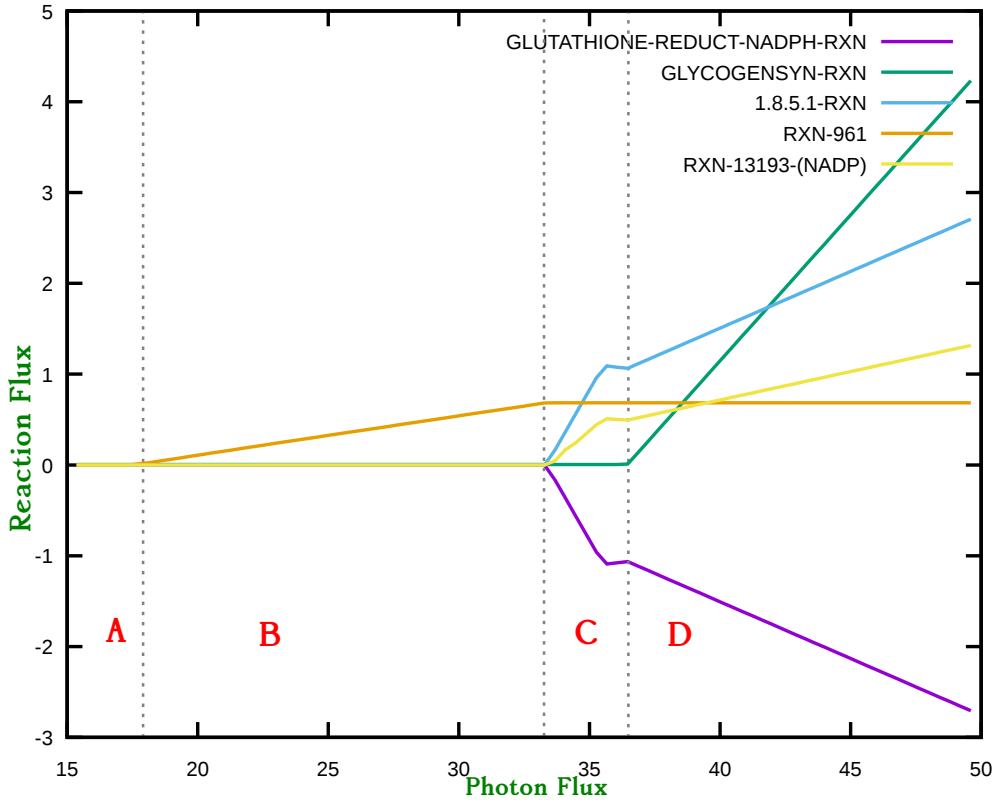



Figure 4.9: Responses of reactions under high light conditions in Chlamydomonas. Representative reactions from sets of reactions with similar responses are shown. All the reactions shown here are localised in the chloroplast

metabolism carrying higher flux with increasing light input was also observed as in the case of *A. thaliana* model.

### 4.4.2 Mixotrophic metabolism

#### General photosynthetic properties

The metabolism under mixotrophic conditions was predicted by using the LP formulation described in Equation 4.2.1. The experimentally obtained biomass composition (see section 3.4) was constrained as outputs while both acetate and $CO_2$ were allowed to be used as source of carbon. The amount of $C_2H_4O_2$ used in the metabolism, as predicted by the model is 23.676 mole $kg^{-1}hr^{-1}$which is com-

parable to the experimental value of acetate consumed on day 3 of the experiment when the growth rate was near maximum (see Figure 3.6). The minimal photon flux that is required by the model to produce biomass components in the proportion defined above is 9.98 mole $kg^{-1}hr^{-1}$ which is very close to the stoichiometric photon flux calculated from the experimental measurements 13.71 mole $kg^{-1}hr^{-1}$. The photosynthetic quotient, ie. $O_2$ produced per $CO_2$ fixed is 0.465 which is relatively low compared to the value under phototrophic conditions but this is the result of both acetate and $CO_2$ being available as carbon sources. However, it should be noted that the elemental balance equation considers only acetate as the source of carbon. In total 363 reactions were required to produce biomass components.

#### Responses under normal light conditions

The network diagram to represent assimilation metabolism under mixotrophic conditions is shown in Figure 4.10. The prediction shows decrease in activity through the Calvin cycle reactions and increased activity in the TCA cycle which is operating in a non-cyclic manner. The change in flux pattern can be attributed to the acetate assimilation pathways which carry most of the flux in the network. In the cytosol, the acetate was converted to acetyl CoA which was then transported to the mitochondria to be fed into the TCA cycle where it was converted to citrate. Some of the citrate was converted to succinate and the remainder converted to malate in the glyoxylate cycle. By doing so the network bypasses two of the $CO_2$ producing reactions in the TCA cycle. The remaining half of the cycle is continued by the succinate dehydrogenase reaction producing fumarate while concomitantly producing reduced ubiquinol which is used in oxidative phosphorylation.

### 4.4.3 Discussion

The results from this study have emphasised the role of central carbon metabolism as a multi-functional network wherein the intermediates generated from the carbon assimilation are used in maintaining the photosynthetic metabolism and supporting biosynthetic pathways. Moreover, published results from various experimental work support the predictions made by the model. For example the increase in flux through the reactions involved in central carbon metabolism and other biosynthetic pathways associated with amino acid metabolism is consistent

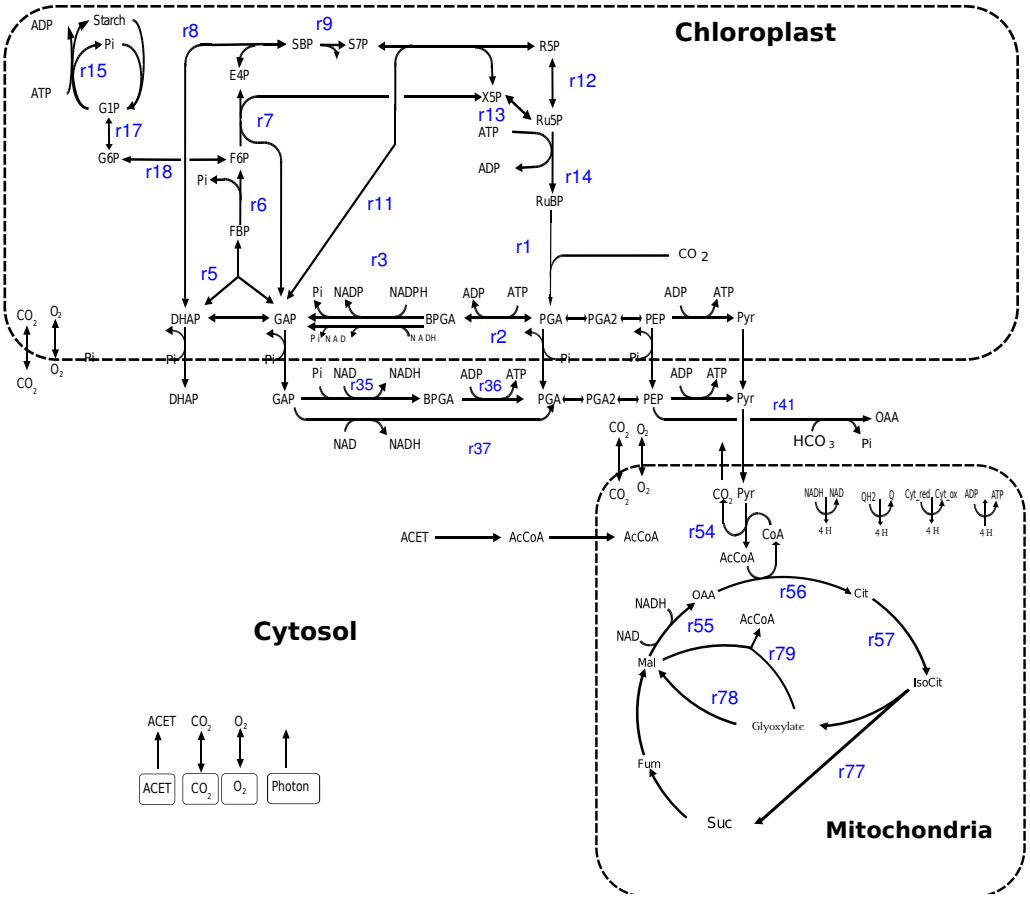



Figure 4.10: Central carbon metabolism under mixotrophic conditions. The acetate coming into the system is transported to the Mitochondria where it is used by the TCA cycle.

with experimental work in *C. reinhardtii* under high-light stress (Davis et al., 2013). The measurements from the nuclear magnetic resonance and mass spectrometry analysis showed a significant increase in the level of amino acid pools in high-light acclimated *C. reinhardtii* cultures. Similarly, a separate experiment in which the metabolite levels were measured from *C. reinhardtii* treated with a fluctuating light environment also showed a rapid increase in the metabolites involved in the central carbon metabolism and the Calvin cycle reactions (Mettler et al., 2014). Likewise, the activity of photorespiratory reactions, as observed from the light scan analysis and the subsequent metabolism of glycolate is also reported to be involved in photo-protection mechanism by promoting non-assimilatory electron transport in *C. reinhardtii* (Renberg et al., 2010). All these observations highlight the predictive accuracy of the model.

The mixotrophic metabolism predicted by the model is consistent with the experimental observations and also comparable to some of the published results from modelling analysis (Chapman et al., 2015; Imam et al., 2015). The decrease in activity of photosynthetic activity under mixotrophic condition in *C. reinhardtii* is attributed to the presence of alternative pathways to compensate the electron flow (Heifetz et al., 2000; Johnson and Alric, 2013; Johnson et al., 2014). Another interesting observation was the decrease in activity of the oxidative pentose phosphate pathway reactions and increase in flux through the acetate assimilation pathways in mitochondria. The activity of the isocitrate lyase reaction under mixotrophic conditions, which helps to bypass the $CO_2$ producing reactions by directly producing succinate and glyoxylate from isocitrate, have also been reported from a experimental work (Plancke et al., 2014). These results show that the model developed here is able to accurately represent both photorophic and mixotrophic metabolism and explain the metabolic interactions under both conditions. They also emphasise the role of central carbon metabolism in supporting photosynthesis.

## 4.5 Conclusion

Overall, the analysis presented here have highlighted the importance of primary photosynthetic metabolism and its flexibility, both in plant and algae, in maintaining normal cellular functions, synthesis of biomass components, and photoprotection mechanism under high light conditions. A unique pattern of set of

reactions responding to high light condition, forming metabolic cycles, was identified which showed the scope of further investigation of such cycles in the network. This aspect will be discussed in more details in Chapter 6. Similarly, the role of mitochondrial metabolism under light condition, specially in balancing the energy requirement have been emphasised. The analysis also show that the reactions mediated by the thioredoxin responded as expected in model flux distribution, despite the fact that it is not specifically included. Moreover, the analysis have demonstrated that both the models are able to correctly represent well known photosynthetic properties and thus are ready to be used for exploring newer features in the metabolic networks.

# Chapter 5

# Effects of enzyme knockouts on GSM of *A. thaliana*

## 5.1 Introduction

With the increasing demand of food supplies, there is a compelling need to improve productivity of crop plants, which has been one of the main focus of the photosynthetic research (Raines, 2003). There is overwhelming evidence from *in vivo* and *in silico* analysis to suggest that improving photosynthetic capacity of plants can help improve crop productivity (Poolman et al., 2000, 2003; Raines, 2003; Lefebvre et al., 2005; Simkin et al., 2015, 2016). Creating transgenic plants is a popular choice to study the regulation of photosynthesis and understand properties of the enzymes involved. In this regard, plants with increased or reduced activity of individual enzymes from the Calvin cycle have been used to study the regulation of photosynthetic carbon flow in higher plants (Raines, 2006; Simkin et al., 2015). This chapter is centered on the knock-out and deregulation of Calvin cycle enzymes in conjunction with the experimental work done by the collaborating partners in the AccliPhot consortium at ETH Zurich. Discussion of the physiological impact of knocking out Calvin cycle reactions from the model will be presented followed by the results from the model after removing these enzymes singly and in combination.

The Calvin cycle is an integral component of the central carbon metabolism in

plants (see section 1.4.1). This study will focus mainly on the four enzymes, SBPase, FBPase, GAPdh and Ru5Pk whose activities are regulated by the ferredoxin thioredoxin system (see section 1.4.1). Briefly, these enzymes are active in the light and inactive in the dark due to thioredoxin regulation. More details about the metabolic properties of these enzymes is presented below.

### SBPase

SBPase (EC-3.1.3.37) catalyzes the conversion of SBP to S7P and phosphate and does not have a cytosolic isoform. Besides thiredoxin-mediated regulation, the activity of SBPase is also regulated by stromal pH and $Mg^{2+}$ levels, which vary in response to changing light conditions (Buchanan, 1980, 1991). It functions to regenerate RuBP in the Calvin cycle (Raines et al., 2000). In addition to its role in the regenerative limb, SBPase is known to influence the control of carbon assimilation along with RuBisCO thus also acting as a limiting factor for $CO_2$ fixation in photosynthetic leaves (Poolman et al., 2000; Raines et al., 2000). Many transgenic species of plants with either reduced or increased activity of SBPase have been used to study the control it imparts in photosynthetic metabolism. These studies have unanimously demonstrated that the reduction in SBPase activity severely affects $CO_2$ assimilation and growth of the plants (Raines et al., 2000; Olcer et al., 2001; Harrison et al., 2001; Raines, 2003; Lefebvre et al., 2005; Tamoi et al., 2006; Liu et al., 2012; Ding et al., 2016). Thus, SBPase may serve an useful target for various metabolic engineering studies to improve photosynthesis.

### FBPase

FBPase (EC-3.1.3.11) is another important enzyme of the regenerative limb of the Calvin cycle and catalyses the conversion of FBP to F6P. It has a cytosolic isoform that is not redox regulated (Serrato, Barajas-Lopez, Chueca and Sahrawy, 2009; Serrato, Yubero-Serrano, Sandalio, Munoz-Blanco, Chueca, Cabllero and Sahrawy, 2009). The chloroplastic isoform has 3 cysteine residues in the structure, 2 of which form disulphide bond and are reduced by thoredoxin during light activation thus controlling its regulation (Chiadmi et al., 1999). Out of 3 FBPase genes identified in *A. thaliana*, two are localised in the chloroplast and function in regeneration of RuBP in the Calvin cycle, and one of them is localised in the cytosol and is known to play an important role in sucrose metabolism (Lee et al., 2008). Various

experimental transgenic experiments have established the role of FBPase in $CO_2$ assimilation as well as in synthesis of various biomass components (Lloyd et al., 1991; Sahrawy et al., 2004; Tamoi et al., 2006; Rojas-González et al., 2015).

### GAPdh

GAPdh (EC-1.2.1.13) catalyses the formation of 1,3-bisphosphoglycerate by oxidising glyceraldehyde-3-phosphate whilst reducing NAD+ to NADH. It has 2 isoforms one localised in the chloroplast and the other in the cytosol. It is encoded by a total of 7 genes in Arabidopsis out of which two of them, GAPCp1 and GAPCp2, are localised to the chloroplast and are involved in photosynthesis (Munoz-Bertomeu et al., 2009). In addition to its chloroplast isoform, plants contain a cytosolic non-phosphorylating NADP-dependent GAPdh (EC 1.2.1.9) which generates reducing power in the form of NADPH for biosynthetic processes (Habenicht et al., 1994). The activity of the cytosolic isoform was found to increase 2-fold under oxidative stress, which underlines its potential role to maintain reductant levels in the cytoplasm of plant cells under such stress conditions (Bustos et al., 2008).

### Ru5Pk

Phosphoribulokinase (EC-2.7.1.19) is part of the regenerative limb of the Calvin cycle and is involved in $CO_2$ fixation in photosynthetic organisms. It catalyses the formation of ribulose 5-phosphate and ATP to ribulose 1,5 bisphosphatase, ADP and phosphate. In addition to the thioredoxin system, activity of both enzymes, GAPdh and Ru5PK, are also regulated by a nuclear-encoded protein called CP12, which forms a Ru5Pk/GAPDH/CP12 multi-protein complex in the chloroplast (Pohlmeyer et al., 1996; Marri et al., 2009; López-Calcagno et al., 2017) The deactivation and activation of GAPdh and Ru5Pk activity is mediated by the the formation and dissociation of the protein complex, which occurs in response to changes in availability of light (Howard et al., 2008).

### Modelling studies of the Calvin cycle enzymes

Smaller models with just the Calvin cycle reactions have been used to study the effect these enzymes have in photosynthetic metabolism. A detailed metabolic model of the Calvin cycle was used to demonstrate that SBPase potentially ex-

ert considerable control over carbon assimilation and starch deposition (Poolman et al., 2000). Similarly, a model of chloroplast metabolism was used to investigate viable pathways under light and dark conditions and under dark with addition of SBPase (Poolman et al., 2003). The investigation is further extended here to study deregulation of both SBPase and FBPase. Moreover, a large-scale model of a plant has never been used before to study the effect of knockouts of these Calvin cycle enzymes. In this regard, the study of metabolic fluxes under knockout conditions presented here can be a guide to further explore the photosynthetic properties of these enzymes and their importance in plant metabolism.

## 5.2 Method

### 5.2.1 Investigating knockout properties with the GSM

The GSM of *A. thaliana* constructed in Chapter 3 was used in this study. Reactions catalysed by each of the enzymes SBPase, FBPase, G3Pdh, Ru5Pk were identified and corresponding LP constraints set to zero individually and in combination, using the formulation as described in Equation 5.2.1. If a feasible solution was possible to the LP problem, indicating viability of the mutant, the rerouting of the fluxes in the network was noted. The fluxes were then compared with the solution under normal conditions, hereafter referred to as the wild-type solution. The information is used to study the role of other enzymes in the network to make such changes possible and to develop an understanding of the physiological impact that is likely to happen *in vivo*. Mainly, the following changes were noted:

- reactions whose fluxes increase in the mutant compared to the wild-type
- reactions whose fluxes decrease in the mutant compared to the wild-type
- reactions whose fluxes were reduced to zero in the mutant
- reactions that carry flux in the mutant but not in the wild-type
- reactions whose flux does not change as a result of the mutation

Unless specified otherwise, the terms wild-type and mutant are used to state the model solution under normal and the knockout conditions respectively.

$$
\begin{aligned}
\text{minimise} \quad & : \ \mathbf{v} \\
\text{subject to} \quad & \begin{cases}
\mathbf{Nv} = 0 \\
v_{i..j} = \mathbf{t}_{i..j} \\
v_{\text{ATPase}} = \text{ATPase} \\
v_{\nu} = \nu \\
v_{\text{LightNonCyc}} \geq v_{\text{LightCyc}} \\
v_{\text{RubiscoCarboxylase}} + v_{\text{RubiscoOxygenase}} = 0.4 \\
v_{\text{KO-rxn}} = 0
\end{cases}
\end{aligned}
\tag{5.2.1}
$$

Where,

- $v_{\text{KO-rxn}} = 0$ indicates that the flux through the reaction under study (either SBPase, FBPase, GAPdh, Ru5Pk or combination of them under dual knockout conditions) is zero.
- all the remaining constraints have similar meaning as described in Equation 4.2.3

### 5.2.2 Elementary modes analysis

To study the effect of deregulation of SBPase and FBPase, a smaller model of Calvin cycle reactions was adapted from a previous study (Poolman et al., 2003). This is a self-sufficient smaller model which uses input of photons as source of energy and $CO_2$ as source of carbon and output of triose phosphates. The reactions that are known to be active only under the light conditions, such as SBPase, FBPase, Rubisco, Ru5PK and GAPdh, were blocked to represent dark metabolism and elementary modes (see Section 2.1.5) were computed. Subsequently, SBPase and FBPase were then allowed to be active under dark conditions, thus representing their deregulation, and elementary modes were computed again.

## 5.3 Results

The overall summary of results from the knockout investigation is presented in Table 5.1. The reactions that increased, decreased, turned on, turned off or carried the same flux as a result of the knockout are summarised in the Table 5.2.

Additional experimental results obtained from the collaborating partner at the ETH Zurich are presented in Appendix A. Experimental knockouts of SBPase and FBPase were viable although with severely retarded phenotype. The mutant lines sbp-1 and sbp-2, of SBPase and fbp-2 of FBPase were compromised in growth with pale leaves after 21 days of growth as compared to the wild-type plant. In contrast, the mutant line gapb-1 for GAPdh grew healthy with little observable phenotype after 21 days of growth. The knockout mutant for Ru5Pk however was not viable. These observations are consistent with the results from the model prediction. The central carbon metabolism of the wild-type solution is represented in Figure 4.2. The following sections will present in details the rearrangement of fluxes under each knockout scenerio.

Table 5.1: Comparison of experimental and model enzyme knockouts. Model predictions are based on feasible LP solution after removal of reaction catalysed by respective enzymes. Experimental results are based on growth of mutant plant. I/P, experiment in progress.

| **Knockout** | **Model** | **Experimental** |
|---|---|---|
| SBPase | Viable | Viable |
| FBPase | Viable | Viable |
| G3Pdh | Viable | Viable |
| Ru5Pk | Non-viable | Non-viable |
| SBPase + FBPase | Viable | I/P |

### Single knockout of SBPase

The active reactions in the central carbon metabolism after the removal of the SBPase reaction are represented in Figure 5.1. Significant changes were observed in the chloroplast specially in the regenerative limb of the Calvin cycle. The transaldolase reaction, which is normally inactive under light condition, was active as an effect of the knockout and carried a flux equal to the flux through the SBPase reaction in the wild-type solution. The flux through the FBPase reaction increased and the increase in flux was also equal to the flux through SBPase reaction in the wild-type solution. Since the SBPAld reaction that uses E4P and DHAP to produce SBP was turned off, the E4P produced by the 2-transketolase reaction was

Table 5.2: Summary of model flux responses to enzyme knockouts compared to the wild-type. FB-G6P, knockout of both isoform of FBPase and the G6P transporter; FB-SB, dual knockout of SBPase and FBPase. Symbols: $\times$, knocked out; $\uparrow$, increased; $\Uparrow$, increased from zero; $\downarrow$, decreased; $\Downarrow$, decreased to zero; $\Leftrightarrow$, changed direction; and -, remained unchanged.

| KO reaction: | SBPase | FBPase | FB-G6P | FB-SB | GAPdh |
|---|---|---|---|---|---|
| Chloroplast | | | | | |
| SBPase | $\times$ | $\uparrow$ | $\uparrow$ | $\times$ | - |
| Transald | $\Uparrow$ | $\Uparrow$ | $\Uparrow$ | $\Uparrow$ | - |
| FBPAld | $\uparrow$ | $\Downarrow$ | $\Downarrow$ | $\Downarrow$ | - |
| SBPAld | $\Downarrow$ | $\uparrow$ | $\uparrow$ | $\Downarrow$ | - |
| FBPase | $\uparrow$ | $\times$ | $\times$ | $\times$ | - |
| PGI | - | $\downarrow$ | - | $\uparrow(\Leftrightarrow)$ | - |
| TriPIsom | - | $\downarrow$ | - | $\Downarrow$ | $\Downarrow$ |
| PGM | - | - | $\uparrow$ | - | - |
| StSyn | - | - | $\uparrow$ | - | - |
| PPI | - | - | $\uparrow$ | - | - |
| G1PAt | - | - | $\uparrow$ | - | - |
| PGKin | - | - | - | - | $\Downarrow$ |
| GAPdh | - | - | - | - | $\times$ |
| Cytosol | | | | | |
| PGM | - | - | $\Downarrow$ | $\Downarrow$ | - |
| FBPAld | - | $\uparrow$ | $\Downarrow$ | $\uparrow$ | - |
| FBPase | - | $\uparrow$ | $\times$ | $\uparrow$ | - |
| TriPIsom | - | $\uparrow$ | - | $\uparrow$ | $\uparrow$ |
| PGI | - | - | - | $\Uparrow$ | - |
| LacDh | - | - | - | - | - |
| ASPTran | - | - | - | - | $\uparrow$ |
| GAPdhC | - | - | - | - | $\Uparrow$ |
| PGKin | - | - | - | - | $\Uparrow$ |
| Mitochondria | | | | | |
| Cpx-II | - | - | - | - | $\uparrow$ |
| Cpx-IV | - | - | - | - | $\uparrow$ |
| Transporters | | | | | |
| G6P-Plas | - | - | $\times$ | $\uparrow(\Leftrightarrow)$ | - |
| GAP-Plas | - | $\uparrow$ | $\Downarrow$ | $\uparrow$ | - |
| DHAP-Plas | - | - | - | - | $\uparrow$ |
| ATP-ADP-Plas | - | - | - | - | $\uparrow$ |
| Photon-tx | - | - | $\uparrow$ | - | $\uparrow$ |

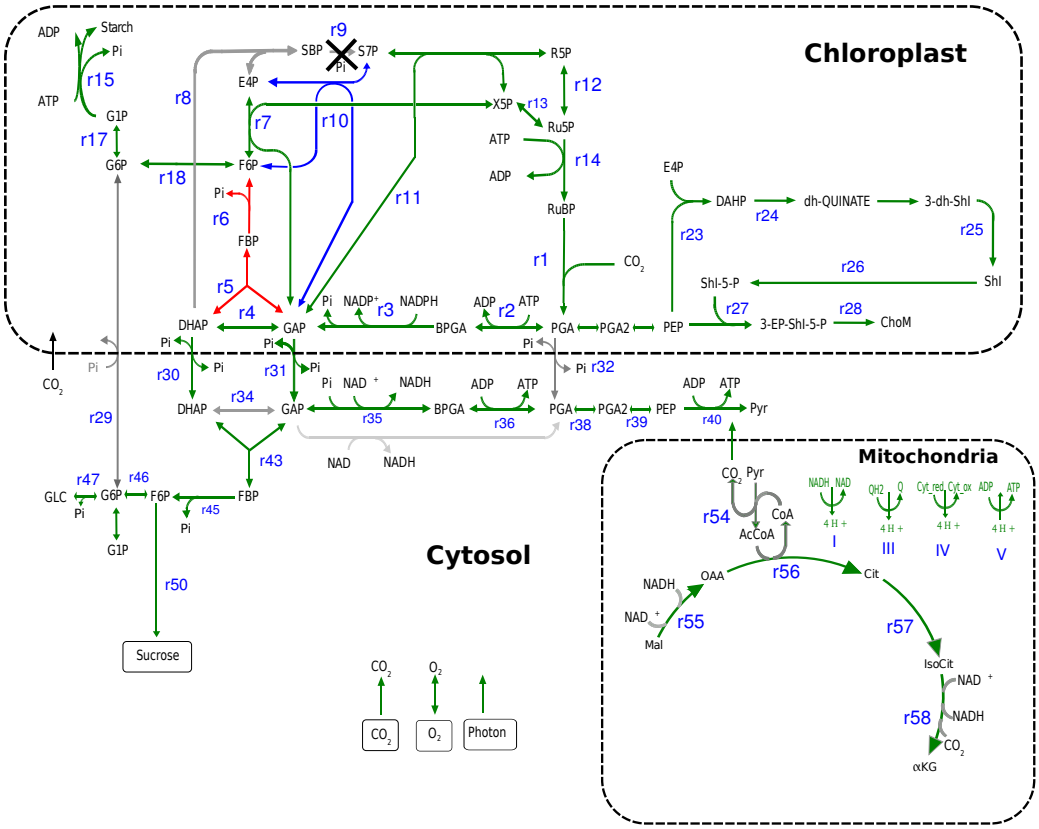



Figure 5.1: Central carbon metabolism of the SBPase knockout mutant. The reactions marked with a X is removed from the model. The reactions in green carry the same flux as the wild-type solution, reactions in blue are inactive in wild-type but turn on in the knockout, and reactions in red have increased flux values compared to the wild-type solution. Reactions in grey carry no flux in the knockout mutant. The flux through the SBPase reaction was substituted by the increase in flux through the FBPase reaction and turning on of the transaldolase reaction.

used by the transaldolase reaction. The S7P produced by the transaldolase reaction was than used by 2-transketolase reaction completing the regenerative limb of the Calvin cycle. The rest of metabolism remained unaffected. Assimilation quotient, quantum efficiency and quantum demand remained unchanged. The objective value however was increased in the mutant and the difference was equal to the flux of SBPase in wild-type solution.

### Single knockout of FBPase

The active reactions in the central carbon metabolism after removal of the chloroplastic FBPase reaction is represented in Figure 5.2. The most notable change under this knockout condition was the flux carried by the transaldolase reaction, which carried no flux in wild-type solution. The flux through the SBPase, cytosolic FBPase and GAP transporter reaction increased. The sum of the increased flux through SBPase, transaldolase and cytosolic FBPase reactions was equal to the flux of FBPase in the wild-type solution. The flux through the PGI reaction decreased while the flux through G6P transport reaction fell to zero.

### GAPdh

The active reactions in the central carbon metabolism after removing the reaction catalised by GAPdh in the chloroplast are represented in Figure 5.3. The major changes were observed in the reductive limb of the Calvin cycle. As the chloroplastic GAPdh reaction was knocked out, the flux through the PGKin reaction was also zero. Flux through the triose phosphate transporters increased thus supporting the flux through cytosolic isoforms of the GAPdh and PGKin. The PGI reaction in the chloroplast was not carrying any flux, thus the carbon required for the synthesis of starch was transported from the cytosol by the G6P transporter. In the mitochondria, the TCA cycle reactions did not carry any flux. However, a smaller increase was observed in the flux through Complex III and Complex V in the mitochondria. Smaller changes were observed in rest of the metabolism across all compartments.

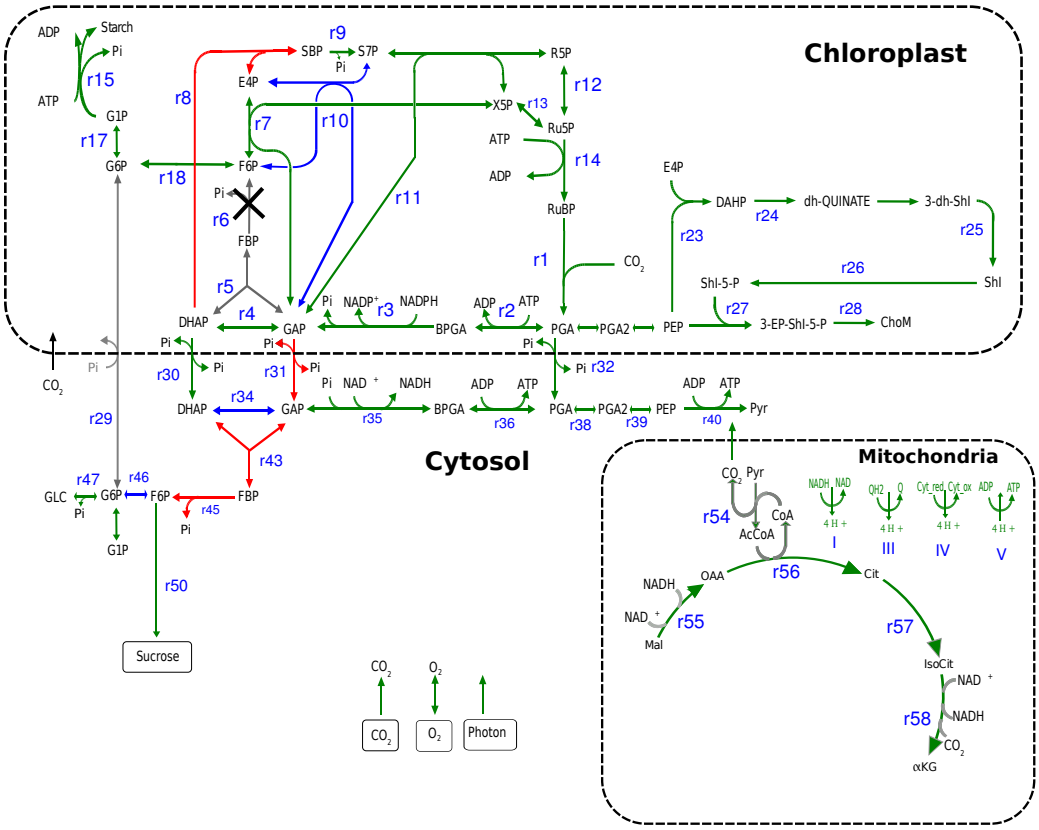



Figure 5.2: Central carbon metabolism of the FBPase KO mutant. The flux through the SBPase reaction has increased while transaldolase and the cytosolic isoform of FBPase is activated. Colour scheme is same as in Figure 5.1.

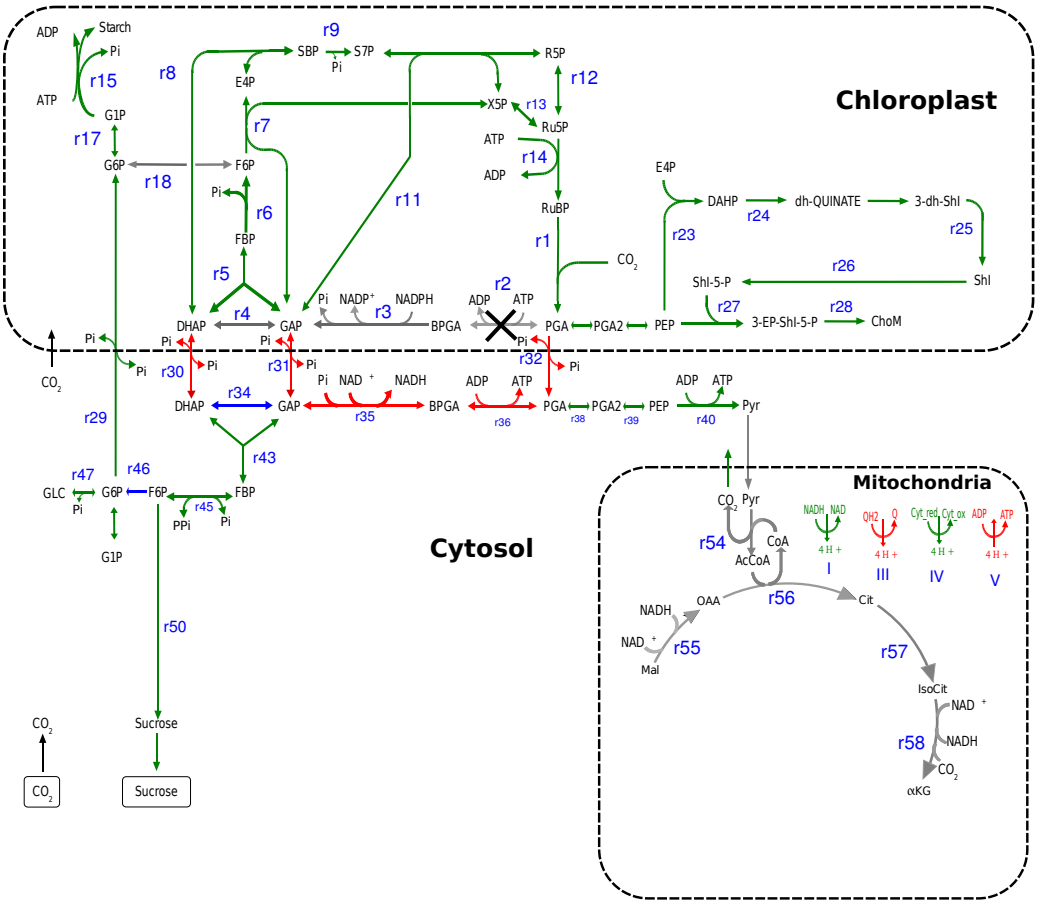



Figure 5.3: Central carbon metabolism after removing reaction catalised by GAPdh in the chloroplast. A major change in this mutant was the shift in activity of regenerative limb of the Calvin cycle from the chloroplast to the cytosol.

### Ru5Pk

A feasible LP solution was not possible in case of the Ru5Pk KO indicating that the presence of Ru5Pk is vital for continual operation of the regenerative limb in the Calvin cycle and hence photosynthesis in plants. Homozygous lines of the Ru5Pk KO mutant were not viable during *in vivo* experiments as well.

### Dual knockout of FBPase (c+ ch) and G6P transport

Experimental results (Rojas-González et al., 2015) suggest that when the chloroplastic isoform of FBPase is absent the cytosolic isoform could not compensate for its activity. Similarly, no activity of the G6P transporter is found in the mature photosynthetic leaves (Kammerer et al., 1998; Niewiadomski et al., 2005). Hence, the FBPase knockout investigation was repeated after removing these reactions from the model. The central carbon metabolism of this knockout mutant is represented in Figure 5.4. Although the qualitative results were similar to the single knockout of FBPase, the flux through the starch synthase reaction increased. The starch was then converted to maltose and transported to the cytosol. The maltose in the cytosol was converted to glucose, by the amylomaltase reaction, which was further used to support carbohydrate metabolism.

### SBPase, FBPase dual knockout

Although a clear complementary role between SBPase and FBPase is seen when either one of them were knocked out from the model, the dual knockout of both of these reaction still produced a feasible solution. The central carbon metabolic network of this double knockout is shown in Figure 5.5. As with the SBPase and FBPase single knockouts, the transaldolase reaction helped support flux through the regenerative limb of the Calvin cycle. G6P was imported into the chloroplast as opposed to its export in the wild-type. This import supplied the carbon necessary for the synthesis of starch. Further, the PGI reaction was operating in the reverse direction, converting G6P to F6P, thus making F6P available for the Calvin cycle.

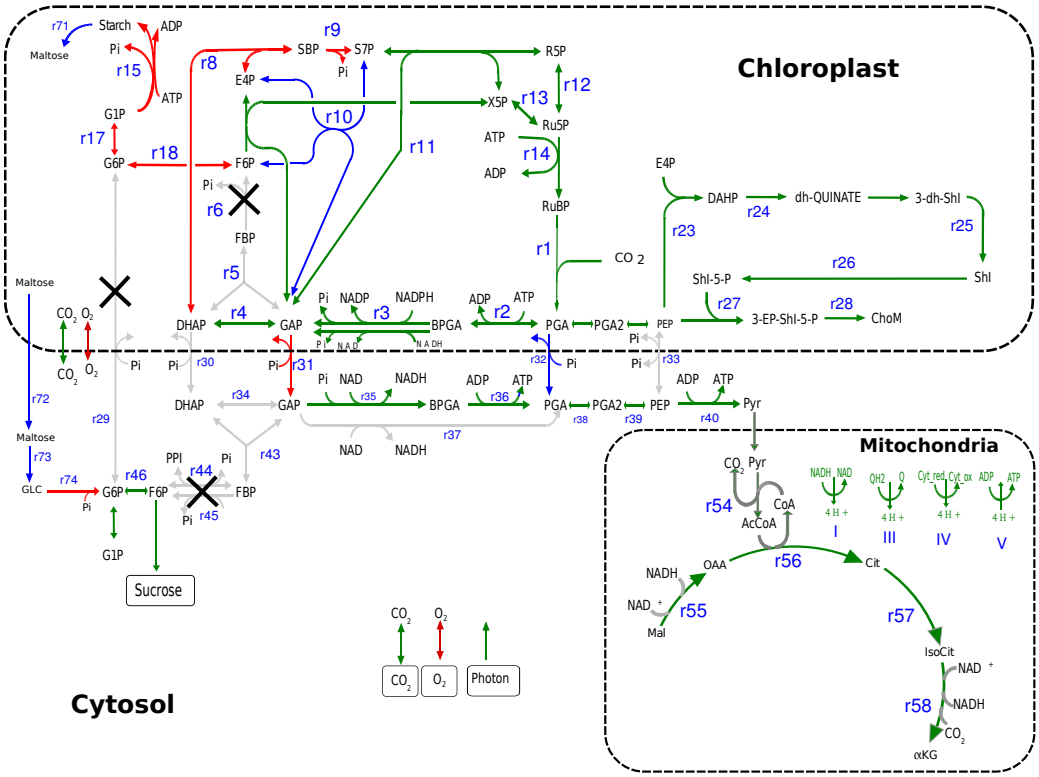



Figure 5.4: Central carbon metabolism of the dual knockout of both isoforms of FBPase and G6P transporter. The flux through the SBPase reaction has increased and the transaldolase reaction start to carry positive flux. The turnover of starch in the chloroplast increase. The starch was converted to maltose and transported to the cytosol and used to support glucose metabolism.

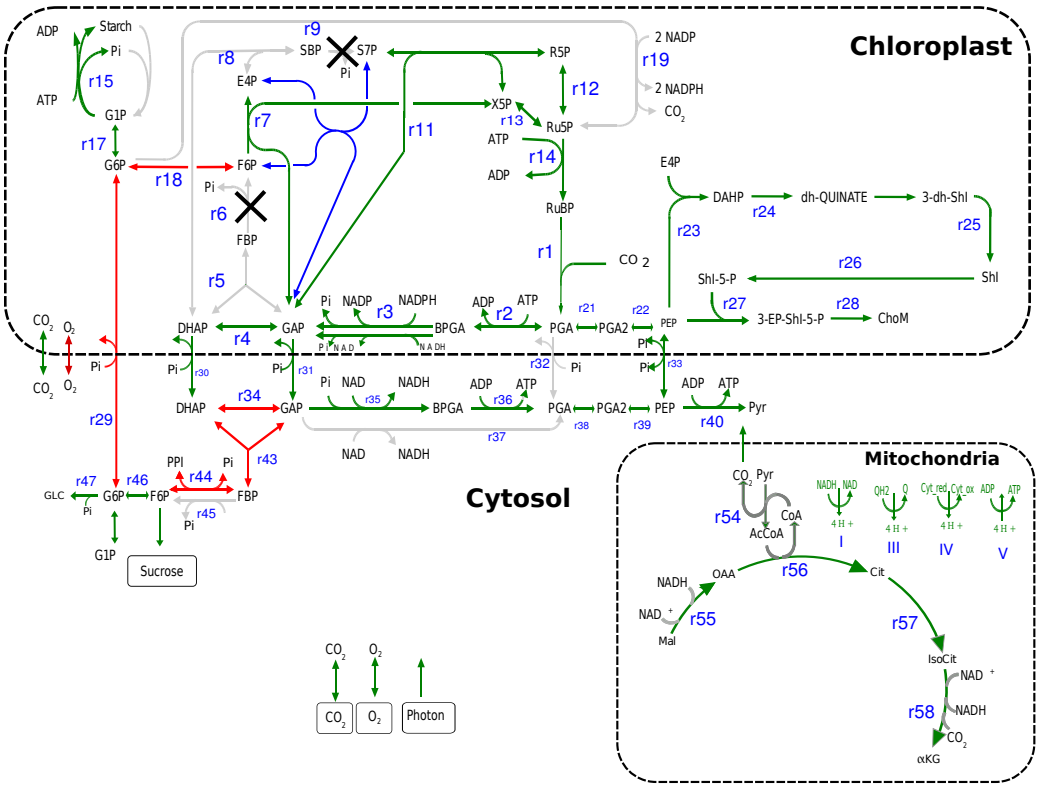



Figure 5.5: Central carbon metabolism of the SBPase, FBPase dual KO mutant. Transaldolase reaction helps to maintain flux through the regenerative limb, the flux through cytosolic isoform of FBPase is increased and G6P is being imported to chloroplast as opposed to its export in the wild-type.

### 5.3.1 Elementary modes analysis of Calvin cycle to study effect of deregulation

In total, 5 elementary modes were computed under dark conditions one mode each producing GAP,PGA, DHAP, E4P, R5P, G6P (first 5 modes shown below). When SBPase was included one additional mode was seen, completely oxidising G6P from starch with reduction of NADP (6th one below). These modes are consistent with those presented by Poolman et al. (2003). Further, when FBPase was included, one more additional mode, 7th in the list below, was computed, with its net stoichiometry exactly same as the 6th mode. The reactions involved in the elementary mode 6 and 7 are represented in Figure 5.6. The net stoichiometry of all elementary modes are listed below.

1. 6 NADP + Starch -> 3 $CO_2$ + 6 NADPH + DHAP
2. 6 NADP + Starch -> 3 $CO_2$ + 6 NADPH + GAP
3. 4 NADP + Starch -> E4P + 2 $CO_2$ + 4 NADPH
4. 2 NADP + Starch -> R5P + $CO_2$ + 2 NADPH
5. Starch -> G6P
6. 12 NADP + Starch -> 6 $CO_2$ + 12 NADPH
7. 12 NADP + Starch -> 6 $CO_2$ + 12 NADPH

## 5.4 Discussion

One of the original goals of this study was to investigate if the GSM of *A. thaliana* is able to represent a growth of the plant after removal of some of the important Calvin cycle reactions. Further, the aim was to understand the metabolic rearrangement of the fluxes at a cellular level under each knockout conditions, that make the mutation viable. The results show that the model is able to cope with the removal of single reactions catalysed by SBPase, FBPase and GAPdh and give a feasible solution to the LP problem, indicating that the *in vivo* knockout of these enzymes will be viable. The single knockout of Ru5Pk was correctly predicted to

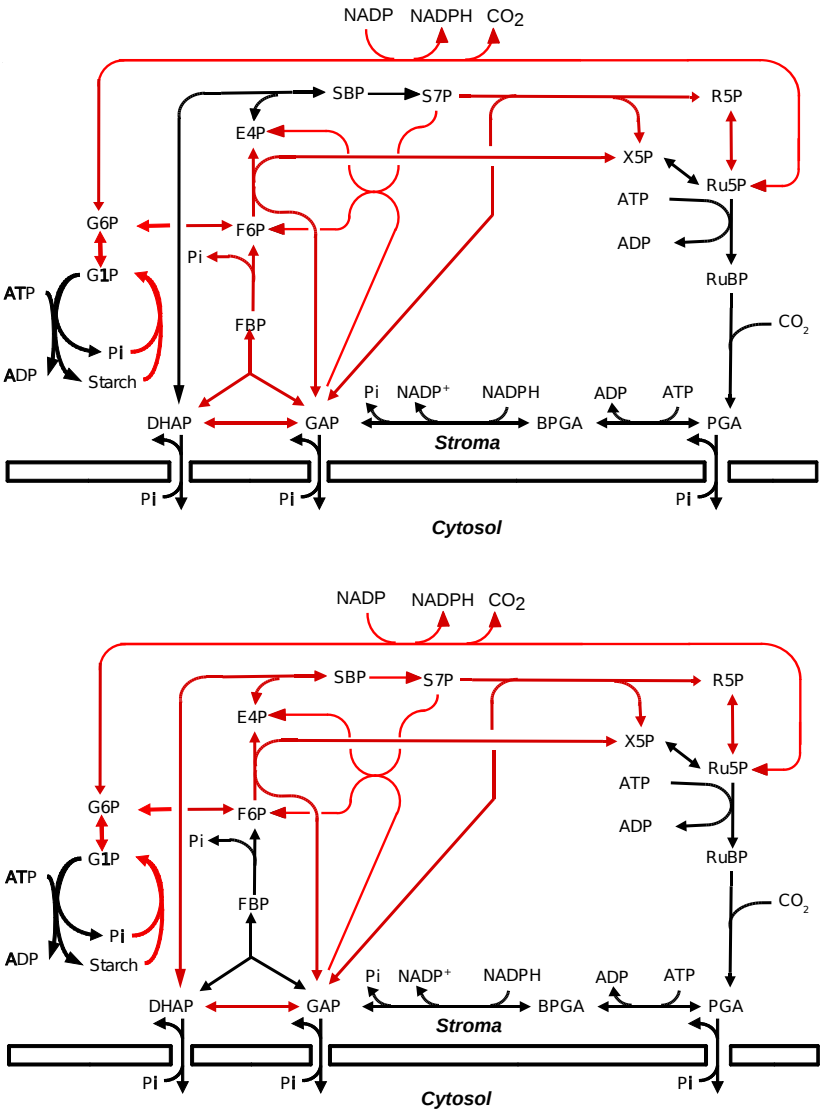



Figure 5.6: Reactions involved in the elementary modes of chloroplast metabolism in the dark after deregulation of FBPase (top) and SBPase (bottom). Reaction participating in the modes under each scenerio are colored in red. Both the modes are involved in oxidation of glucose-6-phosphate generated from degradation of starch and are used to reduce NADP thus producing excess reductants in the system. The diagram is adapted from Poolman *et al* (2003)

be non-viable. The predictions are qualitatively consistent with the experimental observation presented here and those presented by Liu et al. (2012). Moreover, the results show that the rearrangement of fluxes within the chloroplast, more specifically within the Calvin cycle, is capable of sustaining the knockout of any of the reactions. Although the model can discover compensating routes in the metabolic network, it cannot predict whether the plant's regulatory mechanisms can readjust to activate the alternative route, where growth is poorer than wild type. This suggests that the alternative route lack the capacity to carry the wild-type flux. Nevertheless, the results demonstrate the flexibility of the chloroplastic metabolism, since the alternate routes in the network under knockout conditions involves reactions that are not normally active simultaneously.

### SBPase

Many *in vivo* experiments have been performed to study the effect on photosynthesis caused by SBPase mutations. In an experiment using antisense tobacco plants, Harrison et al. (2001) demonstrated that a small decrease in SBPase activity affected the RuBP regeneration capacity but not the activity of RuBisCO, indicating some degree of assimilation could still be active. Similarly, Raines et al. (2000) have reported a significant reduction in the rate of light and $CO_2$ saturated photosynthesis with a small decrease in the activity but the plants were still viable. Further, the investigation demonstrated that the reduction in SBPase activity made no significant change in the level of sucrose but the level of starch decreased linearly with reducing activity of SBPase. That means that although the plant is not able to deposit starch for night time metabolism, it is enough to support its growth. Moreover, other factors, such as the stages of leaf development, are also known to influence the control SBPase exerts on photosynthesis (Olcer et al., 2001). The photosynthetic rate of the young leaf with reduced levels of SBPase was found to be higher in young leaves compared to mature fully grown leaves under the same mutation condition.

### FBPase

A controlled reduction in the activity of the chloroplastic FBPase in an antisense pea plant was reported to cause increases in leaf fresh weight, carbon assimilation rates and total leaf carbohydrates, specially sucrose, as compared to the wild-type

plant (Sahrawy et al., 2004). However, a more recent study presented by (Rojas-González et al., 2015) contradicts the results. They performed experiments with reduced levels of both chloroplastic and cytosolic FBPase, individually and in combination. Not much difference was observed in the growth of the plant with decreased levels of cytosolic FBPase. However, the mutants with reduced activity of chloroplastic FBPase and combined reduction in activity of both isoforms of FBPase resulted in plants with reduced shape and size of the leaves, reduced photosynthetic capacity, lower content of soluble sugars and less starch accumulation as compared to the wild-type. Their results also show that these mutations did not affect the seed viability and germination capacity of the plant. On the other hand, transgenic plants with a 1.7 fold increase in level of FBPase activity showed the same physical phenotype as the wild-type plant moreover with an increase in the amount of starch in the source leaves (Tamoi et al., 2006).

### GAPdh

Experimental results from the knockout mutation of GAPdh isoforms (GAPCp1 and GAPCp2, both plastid localised) have shown that reduced activity of the enzyme severely affects the phenotype with slower root development, dwarfism and sterility caused by drastic changes in sugar and amino acid balance in Arabidopsis (Munoz-Bertomeu et al., 2009). Further experimental evidence also suggests that these genes have an important role in supplying serine to non-photosynthetic organs such as roots via the phosphorylated serine biosynthesis pathway (Munoz-Bertomeu et al., 2010). The transcriptomics and metabolomics analysis by Anoman et al. (2015) suggested that the lack of GAPCp genes affects nitrogen and carbon metabolism brought about by changes in levels of glycerate and glutamine. Their analysis concluded that GAPdh acts as a metabolic connector of glycolysis with other pathways such as serine biosynthesis and ammonium assimilation pathways. However, a recent study based on knockout of two genes of GAPdh (gapcp1, gapcp2) argued that all these activities take place in heterotrophic conditions and the activity of GAPCp is less important in photosynthetic cells (Anoman et al., 2016). The experimental result used in this study (Appendix A) also shows that the growth of the knockout mutant of GAPdh (gapb1) is comparable to its wild-type growth, indicating that the mutation of single GAPdh gene has no phenotypic effect.

### 5.4.1 Newer properties of the enzymes

Based on these previous observations, an impaired phenotype of the complete knockout mutation, for both SBPase and FBPase, as shown in the experimental results are not surprising. Such knockouts are known to reduce the metabolic capability of the organism. Nevertheless, the results presented here, of the rearrangement of the fluxes under each knockout condition, have demonstrated possible metabolic routes under each knockout condition that could make the mutation viable demonstrating scope for further research. For example, SBPase and FBPase were shown to complement each other's activity when either of them is active. The complementary role of the SBPase and FBPase can also be attributed to their common evolutionary origin and sequence similarity at the protein level. In total, 12 out of 19 amino acids sequences involved in the substrate binding site of FBPase are common to SBPase. Such highly conserved residues in their active sites indicate their similar catalytic mechanism (Raines et al., 1992). Most of the experimental work published so far are limited to mutation, either knock down or over expression, of a single enzyme. Combined mutation of these enzyme such as over expression of FBPase when SBPase is knocked out and vice versa could be performed.

Further, the results presented here also highlight a novel role of the enzyme transaldolase. Although the reaction catalysed by the enzyme is known to be active only in the dark, and has different kinetic properties compared to SBPase and FBPase, it helps to support the flux through the Calvin cycle under all knockout conditions. Although there is some evidence for the light inhibition of transaldolase (Anderson, 1981), it has been contradicted by a more recent observation (Caillau and Quick, 2005), which suggest that the enzyme is differently regulated at the level of gene expression and in response to environmental factors in plant tissue. Previously it has been discussed (Poolman et al., 2000) that if the plants were to be viable without the activity of SBPase there must be one or more secondary route that the Calvin cycle should operate through. Similarly, elementary mode analysis (Poolman et al., 2003) showed that one such possible alternate route is via transaldolase. This conclusion complements the current results from the GSM evidenced by the flux of transaldolase in the mutant being equal to that of SBPase in the wild-type. Thus, the physiological significance of the alternative flux

modes involving transaldolase in the light holds an interesting scope for further investigation.

### 5.4.2 Prediction of phenotype from FBA studies

Analysis of the model allows us to identify alternative pathways that effectively bypass the missing reaction, allowing continued carbon assimilation. Although this technique has in this case proved to be a good predictor of (non)viability of the clones, prediction of the phenotypic impact based on the analysis of the network could be further developed. On the other hand, the structural analysis do not consider any kinetic properties or the change in activity level of enzymes under different environment conditions. Similarly, some of the reactions participating in the solutions set for the mutants are known to have very low activity under photosynthetic conditions. A notable example is that the G6P transport has been reported to have no activity in fully grown leaves (Niewiadomski et al., 2005). This evidence also emphasises the need for detailed experimental data to accurately relate the metabolic activity with the results predicted from FBA studies.

### 5.4.3 Deregulation of SBPase and FBPase could be lethal to plants

The aim of the elementary mode analysis on the smaller model of the Calvin cycle was to investigate possible viable pathways when an naturally occurring regulatory mechanism of two of the light active reactions, SBPase and FBPase, is deactivated at the same time. The analysis has been able to identify a viable pathway through the Calvin cycle under such conditions. From a more complete analysis of the model of Poolman et al. (2003), it has been reported that the deregulation of SBPase activity could have adverse effects on the plant phenotype. When both of these enzymes are deregulated, the system has two alternate routes through which starch is oxidised to $CO_2$ with no production of carbohydrate but reducing NADP to NADPH. This will accumulate a lot of reductant in the system, potentially more than could be used in the network. Thus based on these observations, it can be predicted that such deregulation will exhibit increasingly impaired phenotypes in response to dark periods.

### 5.4.4 Conclusion

Overall, the GSM of Arabidopsis was able to correctly predict the viability or non-viability of all individual KO mutants and indicate the possibility of viable dual knockouts based on analysis of metabolic fluxes in the network. The analysis also highlighted a novel role of the enzyme transaldolase in photosynthetic metabolism and further proposed newer features of the metabolic network such as the complementary roles of SBPase and FBPase. The results have also illustrated a supportive role of the cytosolic isoforms of the Calvin cycle enzymes in maintaining continuous metabolic flux in the network. Correspondingly, the need for further experimental data and development of more theoretical methods to accurately predict the phenotypic impacts of plant mutation is also portrayed. Moreover, the study has also identified a scope for further experimental work that will help further develop our understanding of the photosynthetic machinery of plants.

# Chapter 6

# Energy dissipation mechanisms of Arabidopsis and Chalmydomonas

## 6.1 Introduction

Light is the source of energy for plants and algae and is indispensable for photosynthesis which drives their growth and development. It is absorbed by the antenna pigments in the thylakoid membrane inside the chloroplast (see Chapter 1). As shown in Figure 6.1, the absorbed light excites the chlorophyll molecule and the energy thus generated can either be used in the photochemistry, i.e. is used in the $CO_2$ assimilation, is dissipated as heat or is radiated as florescence. A common strategy that has evolved in photosynthetic organisms for the dissipation of excess energy as heat is collectively referred to as non-photochemical quenching (Demmig-Adams and Adams, 1996; Müller et al., 2001). There are other mechanism, such as metabolic cycles operating in different part of metabolism, that are involved in dissipation of energy as heat. However these mechanisms are less well studied. This chapter will present methods to identify energy dissipating routes in metabolism and discuss energy dissipating nodes and modes identified in the GSMs of *A. thaliana* and *C. reinhardtii*.

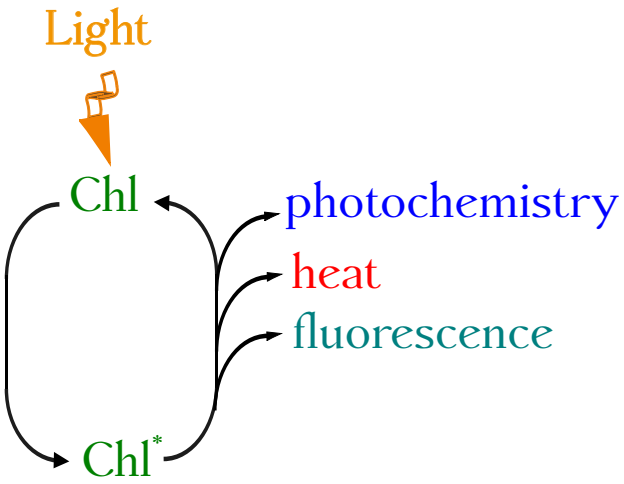



Figure 6.1: Representation of the fate of light absorbed by chlorophyll molecule in light harvesting complex. Absorbed light energy excites the chlorophyll molecules; the energy thus generated can either be used in the photochemistry, dissipated as heat or radiated as fluorescence. Chl*; excited state of chlorophyll.

### 6.1.1 Biochemical basis of energy dissipation mechanisms

As plants and algae are exposed to different light intensities over the course of a day, they have to continuously adapt to cope with such changes. Both low and high light exposure limits photosynthetic performance in plants and algae (Martins et al., 2014; Beckers et al., 2016). The production of energy components, ATP and NADPH, by cyclic and noncyclic light reactions is potentially much higher, under supra-optimal light conditions, than that needed by the bio-synthetic pathways (Beckers et al., 2016). The rate of assimilation of inorganic nutrients like nitrate and sulfate is also increased under high light conditions (Mettler et al., 2014). Availability of NADPH and ATP, in excess, beyond that required for the actual demand, yields a metabolic situation in which no reducible $NADP^+$ would be available. As a result of this, excited chlorophyll molecules, can activate oxygen to its singlet state producing reactive oxygen species (ROS) such as superoxide anions, hydroxyl radicals and hydrogen peroxide (Wilhelm and Selmar, 2011). Light induced production of ROS is increased when the absorption of light energy becomes excessive compared to the photosynthetic activity potentially leading to cell death

in leaves (Ksas et al., 2015). The time scale of the dynamics of these responses can be of the order of seconds owing to rapid movement of leaves, transient shading, rapidly moving clouds etc (Pearcy, 1990; Geiger and Servaites, 1994) which ultimately affect the long term acclimation and growth of the organisms.

Plants and algae have evolved many biochemical, physiological and structural changes thus enabling them to protect themselves by adapting to changing environmental conditions (Walters, 2005). High light acclimation mechanisms work on different timescales and are complex multicomponent processes (Polukhina et al., 2016). The xanthophyll cycle is one of the well-known non-photochemical quenching mechanisms to dissipate excess energy as heat under high light conditions. However, the interconversion of violaxanthin and zeaxanthin in the cycle is a slow process, leading to delayed adaptation to changing light thus causing suboptimal photosynthetic efficiency (Kromdijk et al., 2016). Another metabolic process that is known to be active under high light conditions is the photorespiratory pathway and has been experimentally recognised in plants as a potentially important light stress response to dissipate excess reducing equivalents and energy (Voss et al., 2013). Although the process of photorespiration is an energy expensive process because it oxidizes NAPDH, the overall effect of photorespiration reduces the photon yield (Nogales et al., 2012). Alternatively, the excess energy can also be handled by increased flux in metabolic cycles operating in different part of metabolism. These cycles involve either the conversion of NADPH into NADH involving transhydrogenase-like reactions or hydrolysis of ATP. In this way, growth could be maintained at its optimal rate while preventing photo-damage.

### 6.1.2 Metabolic cycles for energy dissipation

Energy dissipating metabolic modes are sets of reactions operating in a metabolic system, whose net stoichiometry either hydrolyses ATP or oxidises a reductant while the overall net change of the process is only the absorbtion of energy with no net involvement of any other external metabolites other then $H_2O$ for ATP hydrolysis and $O_2$ for NADPH oxidation. They include pathways with reactions operating in cyclic manner such that the cycle dissipates energy but without any net anabolic or catabolic transformation (Gebauer et al., 2012). They can also be defined as a set of two oppositely directed reactions that operate to achieve no change other

than dissipation of energy (Fell, 1997). Thus defined, the metabolic cycles would be thermodynamically infeasible without pairing the cycle to a thermodynamically favorable process that can drive it, such as sets of reactions producing ATP (Sridharan et al., 2015). As illustrated in the Figure 6.2, a metabolic cycle occurs when two reactions in a network run simultaneously in opposite directions but without causing any net transfer of the metabolites. However, the net effect of the cycle transforms a cofactor (oxidising a reducing equivalent) thus dissipating energy in the form of heat (Stein and Blum, 1978; Schwender et al., 2004). For example, if chloroplastic fructose bisphosphatase and 6-phosphofructokinase operate simultaneously the net stoichiometry of them results in hydrolysis of ATP, as shown below.

```
F16BDEPHOS-RXN_Plas:
1 FRUCTOSE-16-DIPHOSPHATE_Plas -> 1 FRUCTOSE-6P_Plas + 1 Pi_Plas
~

6PFRUCTPHOS-RXN_Plas:
1 FRUCTOSE-6P_Plas + 1 ATP_Plas -> 1 ADP_Plas + 1 FRUCTOSE-16-DIPHOSPHATE_Plas
~
-----------------------------------------------------------------------

ATP_Plas -> ADP_Plas + Pi_Plas
```

### Biological function of metabolic cycles

Metabolic cycles are involved in regulation of different biochemical pathways in cellular metabolism (Qian and Beard, 2006). Because of their role in maintaining various metabolic functions such as cellular homeostasis, esterification and hydrolysis of fatty acids, metabolic cycles are also labeled as a potentially important motif of metabolic networks (Sridharan et al., 2015). Metabolic cycles involving carbohydrate turnover have been widely reported in plant tissue although with less clarity about the extent of their occurrence, mechanism of actions and functions (Alonso et al., 2005). Based on isotopic labeling experiments, Whittaker and

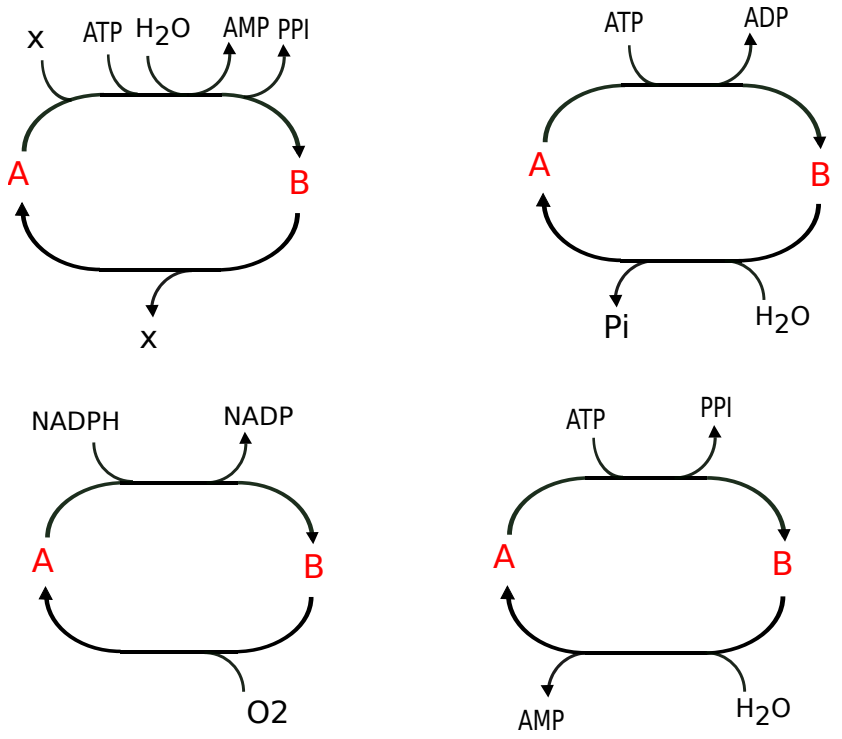



Figure 6.2: Patterns of occurrence of energy dissipating nodes. In all of the nodes represented, there is no net change in the production and consumption of the metabolites A and B but the cycles are driven either by hydrolysis of ATP with water or the oxidation of NADPH by $O_2$. Since the light reactions use photons to generate ATP with the release of water and NADPH with the release of $O_2$, these cycles can couple with the light reactions to absorb photons with no other net change.

Botha (1997) have suggested that between 12-40 percent of sucrose synthesised is re-hydrolysed to glucose and fructose in *Saccharum officinarum* (sugarcane). The process involves metabolic cycles which operate by the hydrolysis of ATP making it energetically expensive. Further, it has been proposed that removal of these cycles might lead to improved sucrose accumulation and agricultural yield (Rohwer and Botha, 2001).

### Approaches being used to identify metabolic cycles

One approach for identifying potential metabolic cycles in metabolism has been the use of elementary modes analysis. This has been used successfully in various small and medium sized models to identify energy dissipating cycles. Rohwer and Botha (2001) used the concept with a small model of sugarcane to find 14 elementary modes, 6 of which were metabolic cycles. Similarly, Teusink et al. (2006) used elementary mode analysis in a genome scale model (643 reactions) of *Lactobacillus plantarum* to identify 28 potential cycles consuming ATP. Another closely related method is the Petri net theory (Chaouiya, 2007), which is based on minimal T-invariants. The method was used by Koch et al. (2005) to identify substrate cycles in sucrose metabolism in potato tubers. Similarly, a medium sized model of *A. thaliana* was used to compute elementary modes to investigate the change in metabolism with the availability of light and further identify reactions involving redox and energy metabolism (Beckers et al., 2016).

Following sections will discuss the methods used in this analysis which are primarily based on a reaction correlation coefficient based metabolic tree (Poolman et al., 2007), recursive calculation of alternate solutions using mixed integer linear programming (Lee et al., 2000) and light scan analysis. Later the results and discussion from the investigation of energy dissipating mechanism will be presented.

## 6.2 Methodology

### 6.2.1 Identifying a photon absorbing core model

In order to carry out a focused analysis of energy dissipation mechanism and reduce the complexity of using GSMs, the models were condensed to photon absorbing

core model. For this all the exchange reactions were removed to get a closed system, with no mass transfer across the system boundary. Then the reactions that cannot carry flux at the steady state were identified (section 2.1.4) and removed. The remaining reactions were isolated as a subsystem and the correlation coefficients between all reactions were calculated. Finally, all reactions that have some degree of correlation with the photon consuming reaction (the manually added reaction to represent input of light in the system) were identified, and separated as a core model. These are primarily the reactions that are assumed to be active under light conditions, and are therefore, potentially involved in various energy dissipation processes.

### 6.2.2 Energy dissipating nodes in correlation tree

A reaction correlation coefficient based metabolic tree can be used to visualise and interpret the function of sets of reactions (Poolman et al., 2007). The set of reactions that cluster together in such tree are more likely to operate together and have similar biological significance compared to others in the adjoining clusters. Here the photosynthetic core model was used to generate a metabolic tree representing the relationship between reactions. In such a metabolic tree, the root node represents the complete model while each leaf node represents a reaction. Combinations of leaf nodes which descend from a common parent node represent a unique subsystem of reactions representing a metabolic module, referred here as sibling nodes. All sibling nodes were isolated as a separate subsystems and the net interconversions of metabolites in such systems were evaluated. The nodes in which the net stoichiometry results in either the hydrolysis of ATP or oxidation of reductants (functioning similarly to the nodes shown in Figure 6.2) were identified as energy dissipating nodes.

### 6.2.3 Computing energy dissipating modes using MILP

A module from IBM-CPLEX (see section 2.1.7) was used to solve mixed integer linear programming problems to find all combinations of reactions in the photon absorbing core model, as alternate solutions. First the MILP problem shown in Equation 6.2.1, was set with objective to minimise the sum of total reaction flux

in the model. A constraint was set on the photon transport reaction such that it can carry only positive flux, forcing the system to use some photon energy. The MILP problem was then repeatedly solved and resulting multiple optimal solutions collected as a dataset. Each alternate solution will consist of a combination of reactions that could use the photon to generate the energy components using either light cyclic or light non cyclic reactions, and consume it internally, thus acting as potential energy dissipation modes. Also, the minimisation as objective function will force the system to return a minimal possible solution set, making them more likely to operate as metabolic cycles. Further, as each solution set corresponds qualitatively to an elementary mode, the nullspace method was used to identify which of the alternate solutions are also elementary modes.

$$\begin{aligned} &\text{minimise} && : \ \mathbf{v} \\ &\text{subject to} && \begin{cases} \mathbf{Nv} = 0 \\ v_\nu > 0 \end{cases} \end{aligned} \tag{6.2.1}$$

The objective is to minimise the sum of the total reaction flux, $\mathbf{v}$, subject to the following constraints:

- $\mathbf{Nv} = 0$ defines steady-state similar to normal LP problem, $\mathbf{v}$ is the vector of all fluxes
- $v_\nu > 0$ defines the constrain to the photon transporter forcing it to carry positive flux.

### 6.2.4 Integration of proteomics data

Proteomics data, with amounts of protein expressed as counts for respective genes associated with reactions in the model (see section 2.1.8), under low and high light conditions, was used for comparison with modelling results. All the data used in this analysis was kindly provided by Dr. Chris Baker from the University of California, Berkley. Briefly, the data for low light condition was gathered from plants grown at 8 hour photo period exposed to 200 mm of photon $m^{-2}$ $s^{-1}$. Similarly, the data for high light condition was gathered from plants grown at 8 hour photo period exposed to 800 mm of photon $m^{-2}$ $s^{-1}$. The maximum rate of photosynthesis was found to be nearly 3 times higher under high light conditions. Since we are interested in finding out which of the reactions are more likely to be active under

high light conditions, the protein expression data under high light conditions was integrated with the model for comparative analysis.

Proteomics expression can be used as a qualitative indicator of the metabolic activity. Thus for each reaction, the corresponding genes were identified and mapped to their respective protein count. In the case of many to one or many to many relationships between genes and reactions, the gene with highest protein count was considered for the comparison. The energy dissipating nodes and modes having positive protein expression for all reactions involved in the node or mode, were identified and further evaluated for their biological significance.

### 6.2.5 Light scan analysis

To study the responses of reactions in the model to changing light conditions, equation 6.2.2 was repeatedly solved with equal increment of photon flux at fixed steps, until the the system was saturated and responses of the reactions remain unchanged. As the production of biomass by plants changes according to the intensity of light, this analysis sets a open flux bound in the biomass reactions to allow them to increase or decrease according to metabolic activity. Additional constraints as defined in (Poolman et al., 2013) were also included in the equation to represent biological conditions.

$$\begin{aligned} \text{minimise} \quad & : \ |\mathbf{v}| \\ \text{subject to} \quad & \begin{cases} \mathbf{Nv} = \mathbf{0} \\ \mathbf{v}_{\nu} = \nu \\ \mathbf{v}_{i..j} \geq t_{i..j} \\ \mathbf{v}_{\text{ATPase}} = ATPase \\ \mathbf{v}_{\text{LightNonCyc}} \geq \mathbf{v}_{\text{LightCyc}} \\ \mathbf{v}_{\text{Carboxylase}} + \mathbf{v}_{\text{Oxygenase}} = \mathbf{c} \\ \mathbf{v}_{\text{EDrxn}} = \mathbf{0} \end{cases} \end{aligned} \tag{6.2.2}$$

where, the additional constraints $\mathbf{v}_{i..j} \geq t_{i..j}$ are imposed to set a flux bound to biomass reactions such that they can carry flux great or equal to the amount measured experimentally. $\mathbf{v}_{\text{Carboxylase}} + \mathbf{v}_{\text{Oxygenase}} = \mathbf{c}$ sets a arbitrary constant flux $c$ to sum of RuBisCO reaction. Further, $\mathbf{v}_{\text{EDrxn}} = \mathbf{0}$ constrains a flux of zero to all unique reactions involved in energy dissipating modes identified from

MILP method. Energy dissipating reactions which are involved in central carbon metabolism or are essential to produce biomass components are excluded from this constraint. The dataset thus obtained was used to identify reactions which respond to the changing light conditions, as discussed in Section 4.2.2, and hence show the potential for absorbing additional light energy by changing biomass composition in the absence of other dissipating mechanisms.

## 6.3 Results from GSM of *A. thaliana*

Out of 2641 reactions in the *A. thaliana* GSM, 1038 were identified to be capable of carrying a steady state flux under light conditions and thus have some degree of correlation with the photon transport reaction. This set of reactions were separated as photon absorbing core model. The core model is presented in Appendix C.

### 6.3.1 Energy dissipating nodes - Correlation Tree Analysis

A metabolic tree was generated for the photon absorbing core model. The tree could be divided into 348 *sibling* nodes such that all the leaf nodes in the tree are attached to common immediate parent node (see Figure 6.3). In total 56 of these sibling nodes were identified as potential energy dissipating nodes. These nodes span the length of 2-4 reactions. 4 of the nodes were localised in the chloroplast, one in the mitochondria, one in the peroxisome and 49 in the cytosol.

#### Nodes with increased protein expression

9 of the nodes shown in Figure 6.4, have all of their reactions with increased protein expression under high light conditions. 27 of the remaining nodes have at least one reaction in them with high protein expression while no protein expression could be found for reactions involved in other 20 nodes. A comprehensive list of all the nodes with the function of reactions involved and their respective protein expression profile is presented in Appendix B Table B.1.

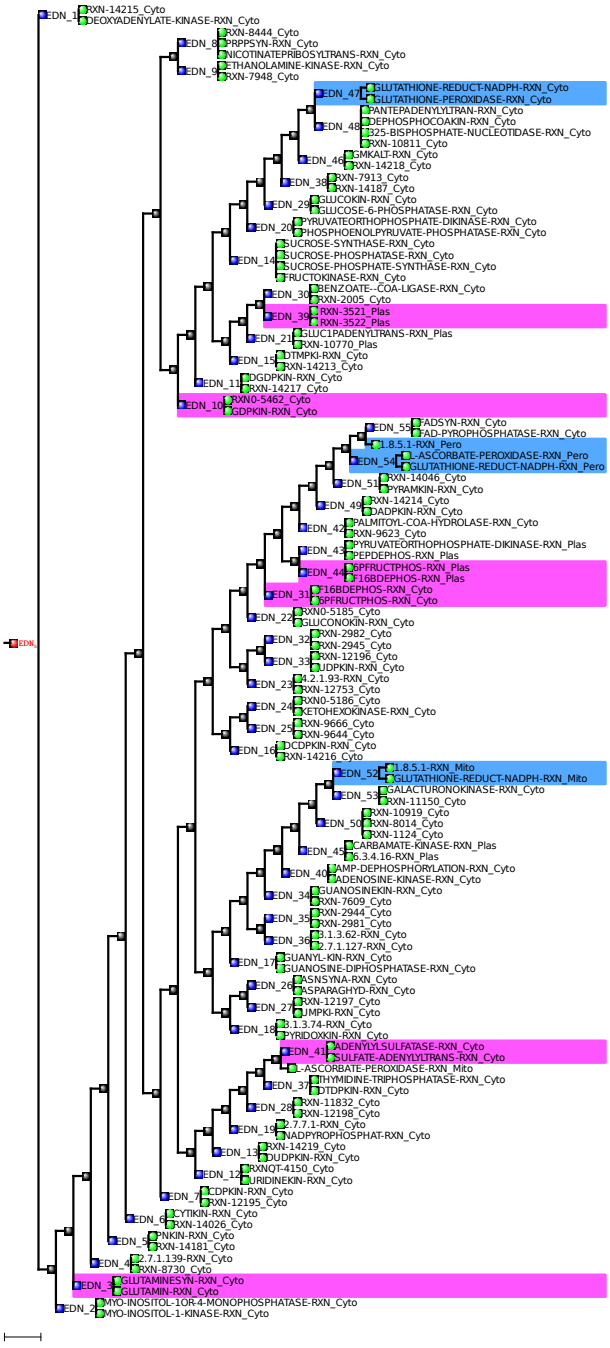



Figure 6.3: Metabolic tree of all the energy dissipating nodes. The nodes for which all the reactions have positive protein expression are highlighted, magenta for the nodes involved in ATP hydrolysis and blue for the nodes involved in oxidation of NADPH. Network diagram of these nodes are presented in Figure 6.4.

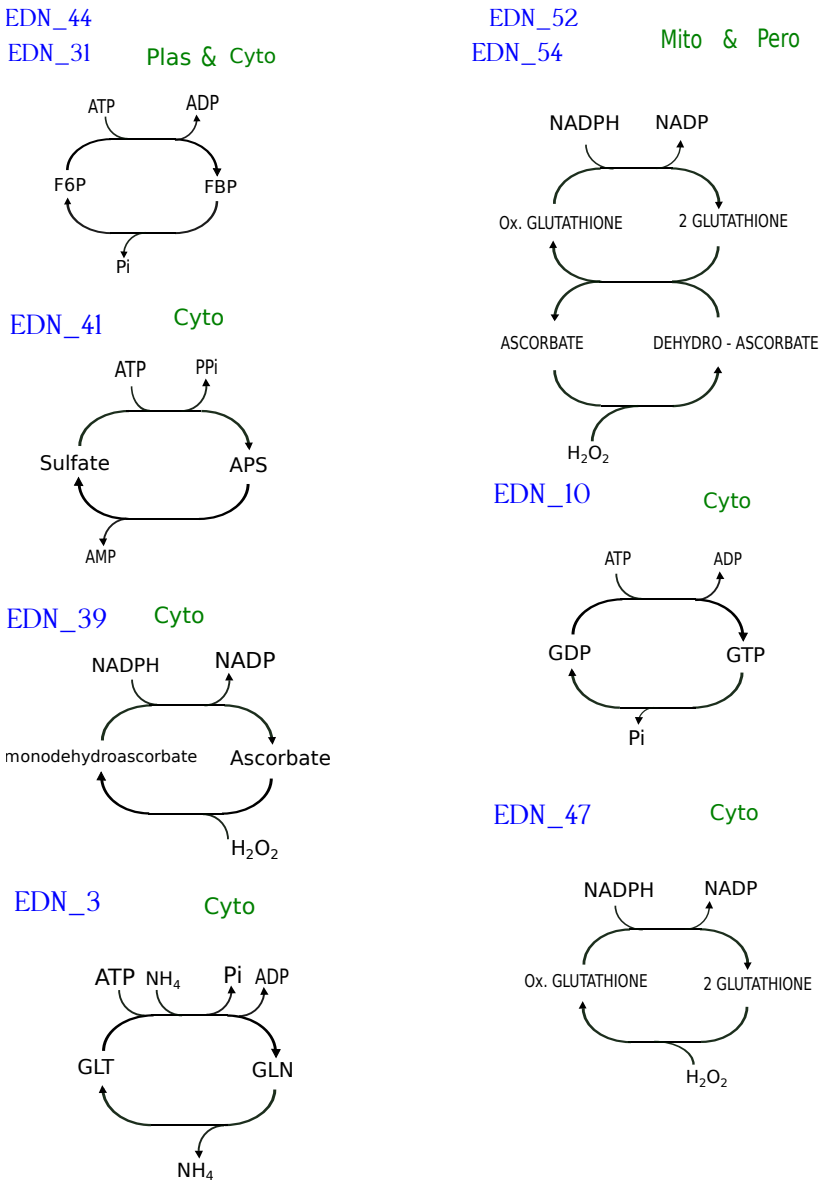



Figure 6.4: Energy dissipation nodes, identified form correlation analysis, in which all of reactions involved in the mode have increased protein expression in high light. The EDN number corresponds to the node number in the metabolic tree shown in Figure 6.3. In the redox cycles, the direct oxidant is $H_2O_2$, derived from utilisation of photosynthetic $O_2$ elsewhere in metabolism. Abbreviations Plas, Cyto, Mito and Pero are used on side of each nodes to indicate their compartmental localisation in chloroplast, cytosol, mitochondria and peroxisome respectively.

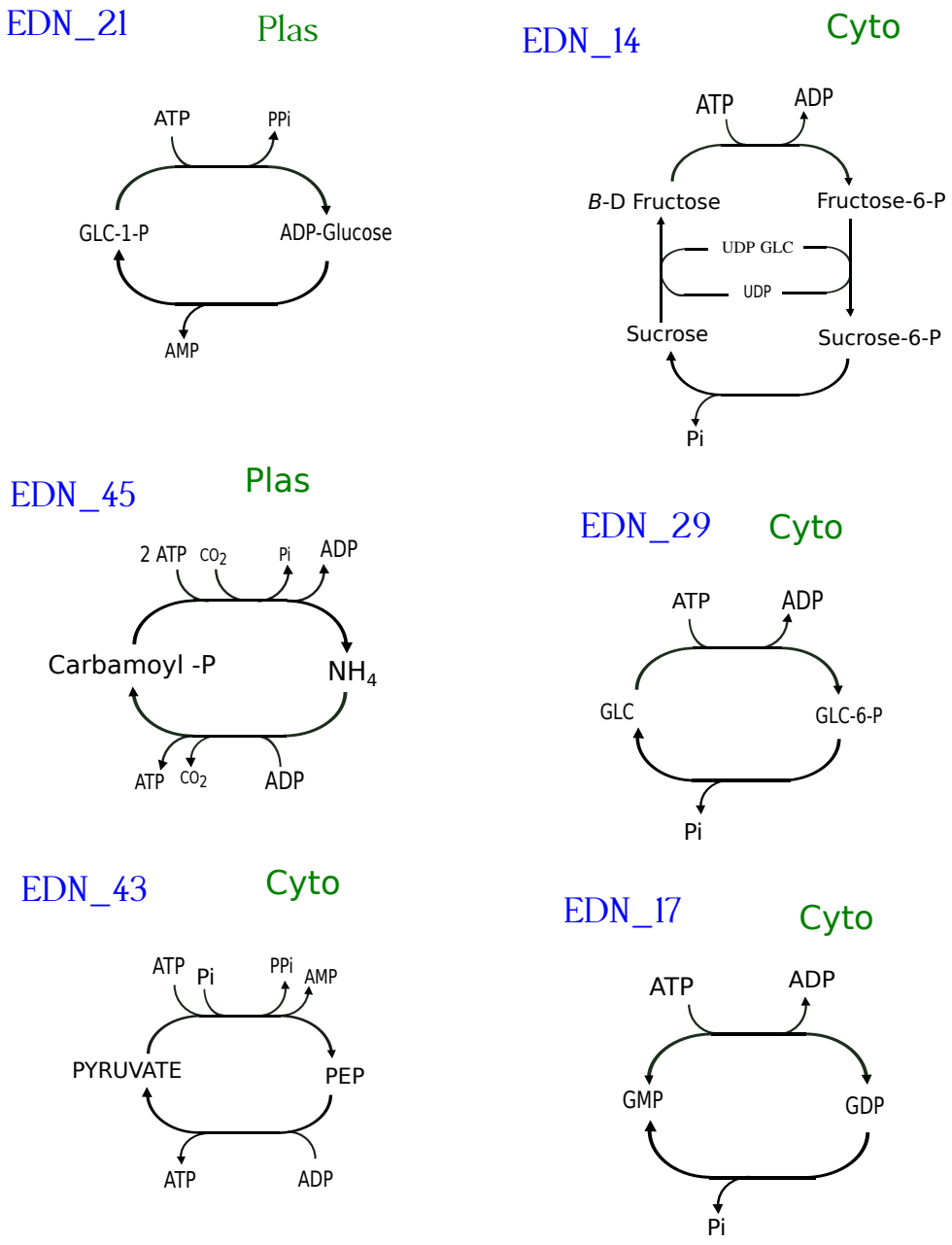



Figure 6.5: Energy dissipating nodes, identified form correlation analysis, for which at least one reaction has increased protein expression under high light conditions. Experimental evidence of their *in vivo* activity can be found in published literature.

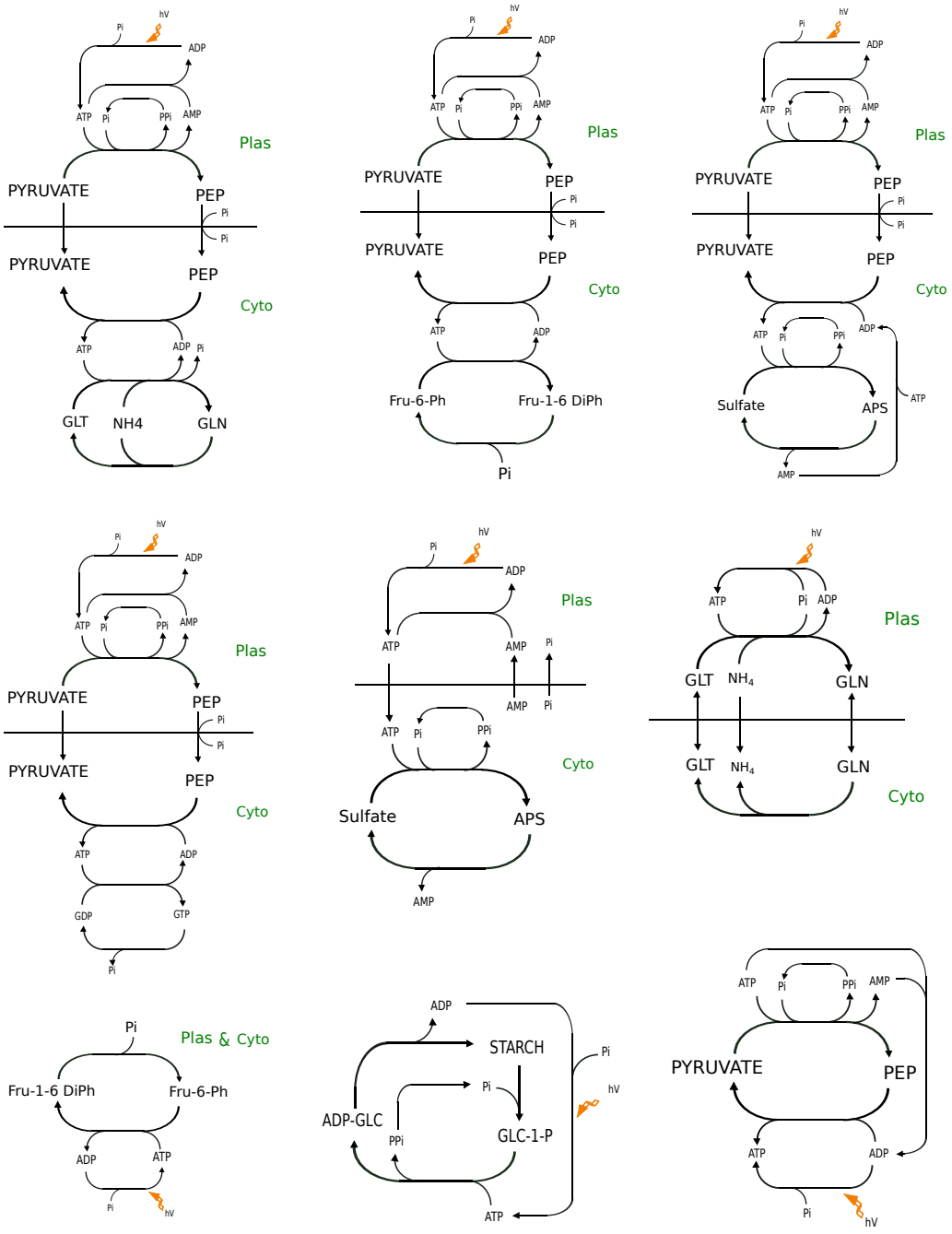



Figure 6.6: Energy dissipation modes, identified from MILP method, in which all of reactions involved in the mode have a positive protein expression. As all of the modes are also elementary modes, the cyclic light reaction producing ATP and subsequent dissipation in the metabolic cycles is explicitly shown.

### 6.3.2 Alternate energy dissipating modes - MILP

In total 231 potential energy dissipating modes that could consume photon and dissipate the energy were identified from the MILP method. 211 of these were also elementary modes. All of these modes were formed by combinations of 237 unique reactions and each involved between 4 to 27 reactions. These modes were localised either in cytosol, chloroplast or spanned across both of the compartments.

#### Modes with increased protein expression

In total 9 of the modes have all the reactions in them with increased protein expression under high light conditions. Figure 6.6 shows network diagrams of these 9 modes. Overall 53 reactions, out of 237 that participate in energy dissipating nodes, have a positive expression under high light conditions and 37 of them have a larger amount of protein as compared to low light conditions. A comprehensive list of all the reactions involved in these energy dissipating modes, frequency of their occurrence, their metabolic functions and respective protein expression profiles is presented in Appendix B Table B.2.

### 6.3.3 Light Scan Analysis - increase in biomass components

In total 198 reactions, out of 237 that are involved in one of the energy dissipating modes identified from MILP method, were set to a flux of zero. The remaining 39 were essential for biomass synthesis.

The response curves of some of the reactions which respond to changing light is shown in Figure 6.7. The whole plot can be divided into different regions according to the points at which some group of reactions saturates, a normal photosynthetic metabolism can be seen in region A. With further increase in photon flux the reactions involved in the starch synthesis pathway were turning on, as shown (region B). After the flux through the starch synthesis saturates, there is a brief increase in activity of glucose-6-phosphatase reaction in region C associated with increased flux in the GLC biomass transporter. The flux through sucrose phosphatase increases in region D, leading to subsequent increase in export of sucrose. The sucrose-6P used by the phosphatase reaction is produced by sucrose phosphatase

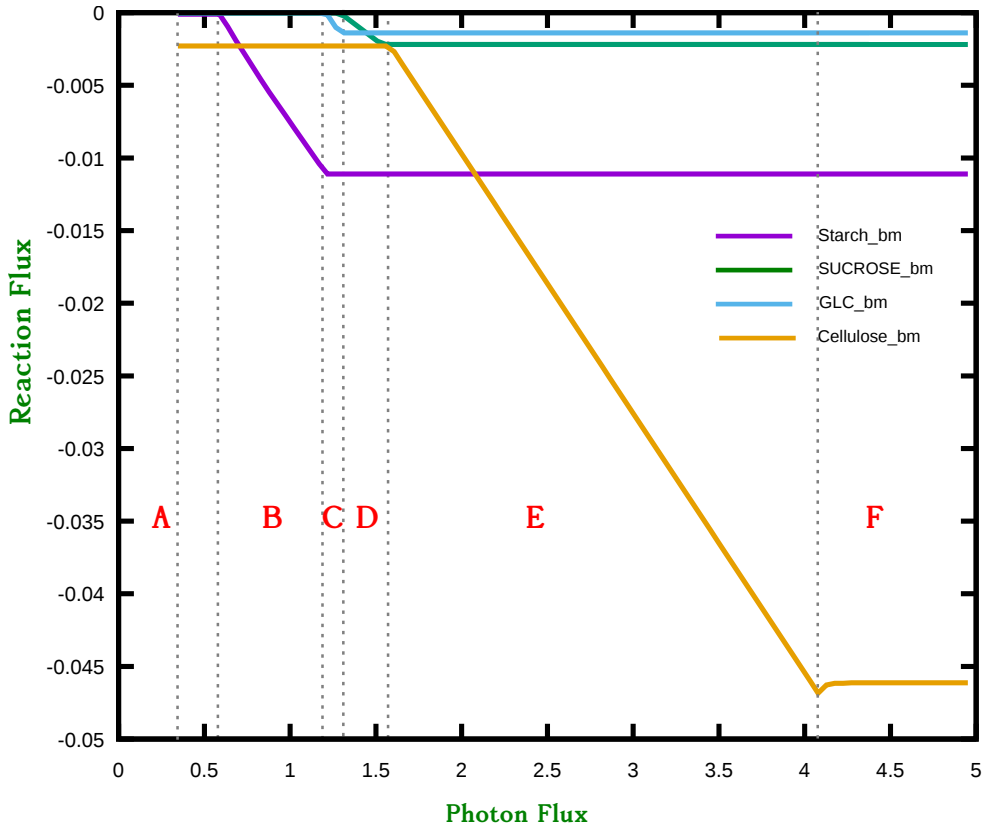



Figure 6.7: Results from light scan analysis in the Arabidopsis model, demonstrating increase in biomass components. Negative flux indicate export of the biomass components from the system. The plot can be divided into 6 regions. Region A represented normal metabolism with optimal photon flux. No solution is possible before this region. Region B, C, D and E represents the range of photon flux when the production of starch, glucose, sucrose and cellulose increases respectively. Production of all the biomass components saturates in region F. The basal level of synthesis of glycogen, sucrose and glucose in biomass in region A is non-zero but not observable at the scale of this graph.

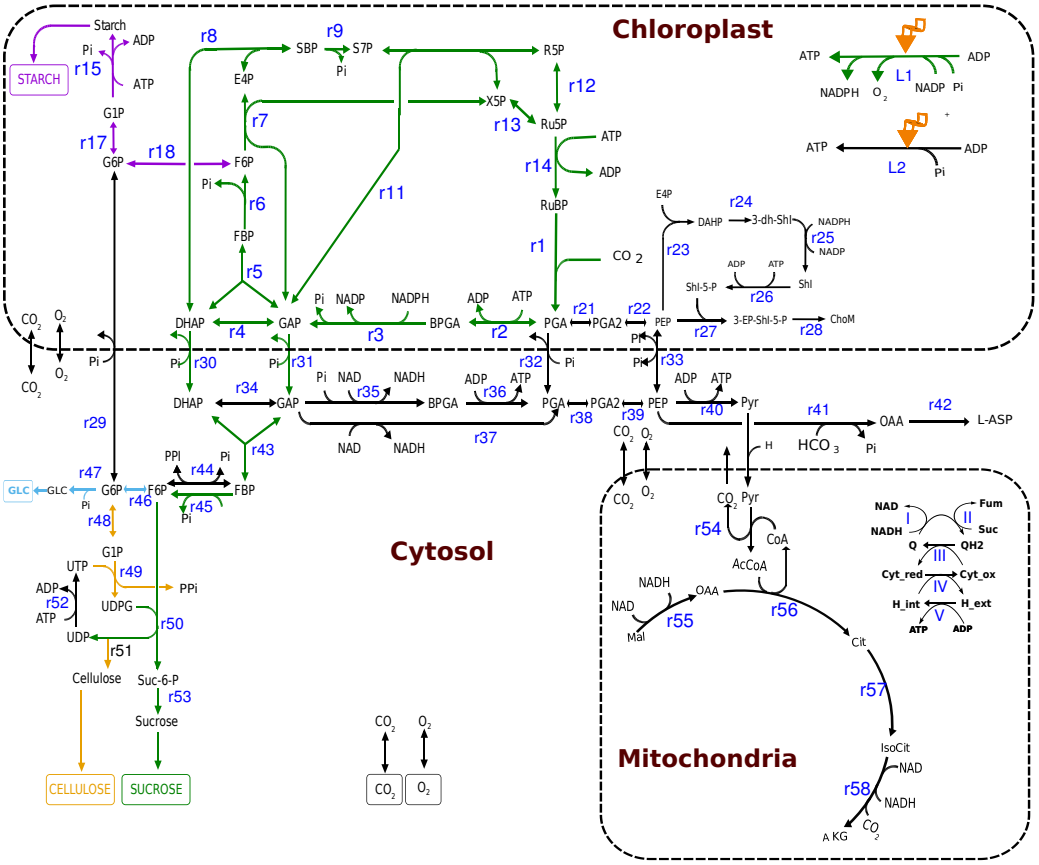



Figure 6.8: Network diagram showing the pathways involved in production of biomass components in Arabidopsis model. Limitation in the fluxes through the reactions involved in energy dissipating pathways improves the performance of central carbon metabolism thus leading to increase in production of biomass components. The pathways representing production of each biomass component are color coded as in the response curve shown in Figure 6.7, the Calvin cycle reactions are shown in green to represent increase in flux through region B to E. Reactions represented in black carry same flux over the range of photon fluxes increased.

reaction which uses fructose-6-phosphate as a substrate along with UDP glucose. In the region E, after the sucrose phosphatase reaction saturates, the UDP glucose is used by the UDP dependent cellulose synthase reaction to produce cellulose. However, the UDP glucose used by the cellulose synthase reaction is produced by UDP-glucose pyrophosphorylase which uses the glucose 1-phosphate along with UTP to produce UDP glucose. The glucose 1-phosphate used here is produced by the isomerase reaction. The fructose 6-phosphate is generated in the chloroplast by the Calvin cycle and transported to the cytosol to be used by glucose, sucrose and cellulose synthesis pathways thus increasing their export under increasing light conditions.

Removing reactions involved in energy dissipation mechanism helped to improve the function of central carbon metabolism thus contributing to increased production of biomass components. Further, it should be noted that even though 198 reactions actively participating in energy dissipation mechanism were removed from the model, processes such as the xanthophyll cycle and other essential reactions (including 39 reactions that were excluded from removal) are are still operating in the network thus preventing plants from the vulnerability of damage from high light.

## 6.4 Results from Chlamydomonas model

Out of total 1865 reactions in the *C. reinhardtii* model, 626 were separated as a photon absorbing core model.

### 6.4.1 Energy dissipating nodes - Correlation Tree Analysis

The metabolic tree was generated for the core model. The tree could be divided into 220 sibling nodes out of which 34 subsystem were identified as energy dissipating nodes. Out of these, 3 of them were localised in chloroplast, 2 in mitochondria and 29 in cytosol. Metabolic functions of the reactions in each mode and the part of metabolism they belong to are presented in Table 6.9.

Comparison of ED nodes obtained from *A. thaliana* and *C. reinhardtii* model shows that 12 of the nodes are common in both the models and are listed below. The first 3 nodes in the list were also found to have increased protein expression in proteomics data analysis in the *A. thaliana* model. These nodes are highlighted with green color in Figure 6.9. The remaining 8 modes are highlighted with light brown in Figure 6.9.

```
['F16BDEPHOS-RXN_Plas', '6PFRUCTPHOS-RXN_Plas']
['GLUTATHIONE-REDUCT-NADPH-RXN_Cyto', 'GLUTATHIONE-PEROXIDASE-RXN_Cyto']
['SULFATE-ADENYLYLTRANS-RXN_Cyto', 'ADENYLYLSULFATASE-RXN_Cyto']
['MYO-INOSITOL-1OR-4-MONOPHOSPHATASE-RXN_Cyto', 'MYO-INOSITOL-1-KINASE-RXN_Cyto']
['RXN-5647_Cyto', 'CHOLINE-KINASE-RXN_Cyto']
['RXN-12197_Cyto', 'UMPKI-RXN_Cyto']
['RXN-12195_Cyto', 'CDPKIN-RXN_Cyto']
['RXN0-5462_Cyto', 'GDPKIN-RXN_Cyto']
['RXN-12196_Cyto', 'UDPKIN-RXN_Cyto']
['PALMITOYL-COA-HYDROLASE-RXN_Cyto', 'RXN-9623_Cyto']
['2.7.1.139-RXN_Cyto', 'RXN-8730_Cyto']
```

### 6.4.2 Energy dissipating modes - MILP

In total 191 alternate solutions were identified out of which 177 were also elementary modes. These modes were formed by combinations of 242 unique reactions and localised either in the chloroplast, the cytosol or spanned across both of these compartments.

## 6.5 Discussion

The response of plants and algae to high light conditions requires adjustment in capacity to either use the absorbed excitation energy through photochemistry and or dissipate the excess energy. This investigation was particularly focused to comprehensively identify the full range of potential energy dissipating cycles occurring in different part of metabolism in GSMs. Using correlation coefficient analysis and MILP, it was possible to systematically identify all possible combinations of reactions in the model metabolic networks that could collectively function to dissipate excess energy by hydrolysis of ATP or oxidation of equivalent reductants.

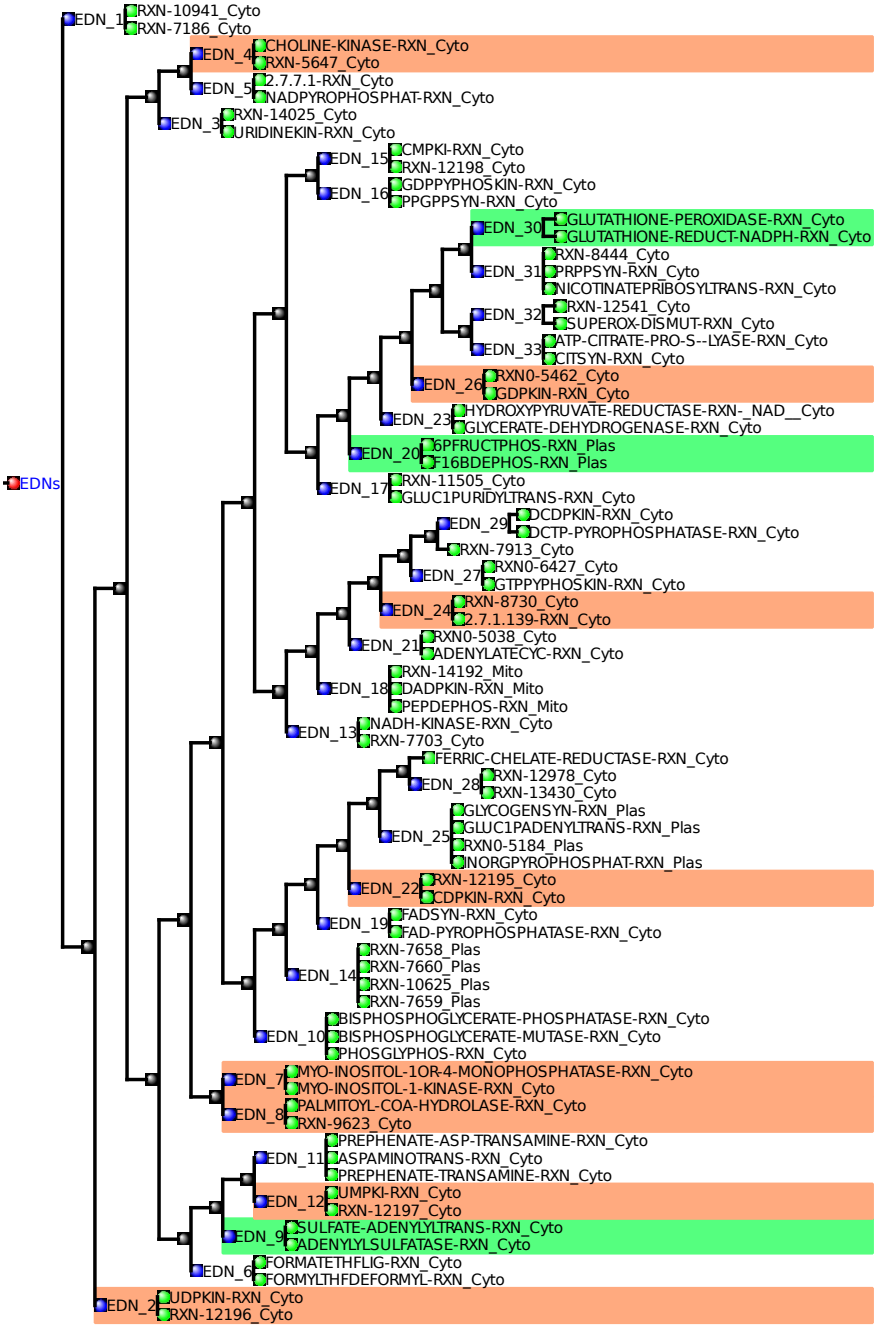



Figure 6.9: Metabolic tree of energy dissipating nodes in the *C. reinhardtii* model. All nodes that are highlighted are also present in the *A. thaliana* model. Nodes for which have positive protein expression in *A. thaliana* are highlighted in green.

These processes could operate in different cellular compartments and cover reactions from various textbook-defined metabolic pathways. For example the cycle involving F6P and FBP (Figure 6.4 EDN_44) is part of both glycolysis and gluconeogenesis, likewise the ascorbate-glutathione cycle (Figure 6.4 EDN_52) involving production and degradation of glutathione is associated with the metabolism of glutamate cysteine and glycine. In this regard, the analysis has been able to predict novel function and a close interaction between reactions with seemingly unrelated functions, simply based on their involvement in either consumption or production of energy components.

A common feature noted in energy dissipating cycles that are involved in the hydrolysis of ATP, identified from both methods, is that these nodes and modes are usually made up of two non-equilibrium reactions involving a kinase and a phosphatase for which a product of one is the substrate to the other. If both reactions in such a cycle are assumed to operate at the same rate then there is no net flux of the metabolites but the metabolic loss of energy as heat. Similarly, the cycles involving oxidation of reductants include a pair of oxidase and reductase reactions running in opposite direction with the net effect of oxidising NADPH.

The energy dissipating nodes identified from the correlation method are always localised in one single compartment while some of the modes identified from the MILP methods span two compartments and are connected by metabolite transporters. Some of the cycles identified from both the methods are more commonly known and have experimental evidence (Geigenberger and Stitt, 1991; Rohwer and Botha, 2001; Koch et al., 2005) of their *in vivo* activity while some others have been reported from computational analysis. The methods used here have been able to identify all those reported previously and further propose additional cycles with a potential function in energy dissipation. In total 7 out of 56 nodes in case of the *A. thaliana*, and 6 out of 34 in case of the *C. reinhardtii*, identified from the correlation method, have been reported previously in the literature. As it is difficult to extensively characterise all the cycles identified from two methods, proteomics data obtained under high light conditions was used here to qualitatively compare which of the cycles identified from the model are more likely to occur *in vivo*.

### 6.5.1 Experimental validation of results with proteomics data

The metabolic fluxes are the end result of the interplay of gene expression, protein concentration, protein kinetics and metabolite concentrations and thus represent the metabolic phenotype of the organism (Winter and Kromer, 2013). Integration of proteomics data into GSMs can give useful insight into conditional changes in the metabolic activity and hierarchical regulation of metabolic fluxes by representing their control over the activity of enzymes(Yizhak et al., 2010). In this regard, use of proteomics data to identify the reactions more responsive under high light condition is a logical choice. Moreover, it has been concluded that proteomics data constrain GSMs more close to a physiological state then other omics data (Großeholz et al., 2016). Similarly, the integration of state-specific protein data in a GSM have shown to improve prediction accuracy of the model (Simons et al., 2014). Since, the data used in this study has been generated under high light conditions, when photosynthetic rates were observed to be maximum, it is ideal to be used for investigation of potential energy dissipating modes under high light conditions.

By using the proteomics data as a basis for defining activity of reactions, have suggested a shift in metabolism that would not be predicted on the basis of observations of input and output rates alone. Thus the proteomics data analysis have been useful to suggest with more certainty that at least 9 of the nodes identified from the correlation analysis and 10 of the modes from the MILP analysis are likely to occur *in vivo*. Moreover, some of the modes identified from this analysis have already been validated experimentally.

Experimental evidence support the role of GLC-G6P cycle (EDN 29 Figure 6.3 and 6.5 ) in carbohydrate metabolism, for maintaining the level of glucose (Stein and Blum, 1978). The significance of the cycle was also demonstrated using the concept of elementary mode in a genome scale model (643 reactions) of *Lactobacillus plantarum* (Teusink et al., 2006). The 28 energy dissipating cycle identified in the analysis were attributed to their function in uncoupling catabolic ATP production and anabolic ATP consumption specially under the conditions when their level is high.

The cycling of F6P-FBP (EDN_31, EDN_44 Figure 6.3) is known for recycling metabolites in the glycolysis and has been observed experimentally in yeast cells (Hers and Hue, 1983). Similarly, the cycling of Pyr-PEP (EDN_43 Figure 6.3) was reported to have a role in carbon flux distribution mainly by connecting different pathways in carbon metabolism responsible for catabolism, anabolism and energy supply in bacteria (Sauer and Eikmanns, 2005).

The metabolic cycle of sucrose as shown in EDN_14 (Figure 6.3, 6.5) has previously been reported in different plant species. From an experiment in germinating castor-oil-plant seed, Geigenberger and Stitt (1991) have shown that the metabolic cycle involving simultaneous synthesis and degradation of sucrose allow the change in pool of sucrose with very minimal change in concentration of sucrose and other metabolites. Similarly, experimental work on sugarcane, reported by Rohwer and Botha (2001), showed a increase in conversion of hexose to sucrose, increasing its overall accumulation, in response to an increase in activity of the fructose uptake and decrease in activity of hexokinase reaction. Their analysis of a smaller model of sugarcane, also shows that 6 elementary modes, out of 14 total, were metabolic cycles involving sucrose metabolism, showing significant control over sucrose accumulation (Rohwer and Botha, 2001). Likewise recycling of sucrose involving, sucrose invertase, sucrose phosphate synthase, sucrose phosphatase has also been reported to play an important role in sucrose accumulation in potato tubers (Koch et al., 2005).

While the method used here for protein data integration has the merit that reactions associated with over-expressed proteins will tend to be favored over those with under-expressed genes, and its application appears to yield biologically meaningful results (described above), a complicating factor in this work is the presence of isozymes, multi-subunit enzymes and relatively sparse information concerning compartmental targeting of isozymes. These aspects could be considered for future developments of the study to further improve the predictability of the model.

### 6.5.2 Broad applicability of the method

Most of the methods reported in literature to determine metabolic cycles were primarily based on elementary mode analysis. But as pointed out previously, it is challenging and computationally expensive to compute all possible elementary mode in larger models. This challenge has been addressed in parts by targeting a specific metabolite or cofactor such as only those consuming ATP (Gebauer et al., 2012). Nevertheless the approach does not represent the relationship between different cofactors such as the modes which could consume both ATP and NADPH.

The MILP method used here overcomes this problem as it does not require a specified boundary to limit the number of reactions and or limit the type of reactions that could form a solution set. Thus the method is able to compute all possible combinations of reactions involving individual or combinations of cofactors. The modes identified here are of various length, some only hydrolysing ATP, some only oxidising NADPH and some doing both. Moreover, as the method is allowed to find a multiple solution set with the same or different objective function, it can be assumed that the computed solution set contains all possible combinations of solutions. Also, because minimisation is used as the objective function, the algorithm will tend to return a minimal set of reactions in a particular solution that are thus more suited to function as a metabolic cycle. Similarly, the correlation method does not require a specified boundary as to how many reactions could cluster. The program used here allows clustering of all reactions simply based on their correlation and thus they are more likely to have similar biological functions. The fact that all of these nodes only have between 2-4 reactions, makes them more suitable candidates to operate as metabolic cycles.

### 6.5.3 Possibilities of *in vivo* activity

All the energy dissipating modes identified in this study can be catagorised as being distinct from one another based on their activity and mode of operation. They represent a continuum of photoprotection responses and are critical to the maintenance of the photosynthetic apparatus under conditions of excess light. The operation of these modes could be controlled by various regulatory factors *in vivo*, and it is difficult to determine their activity and effectiveness simply based on

stoichiometric analysis. On the other hand, modes involving many reactions can be assumed to be inactive as a result of down-regulation of at least one enzyme involved. This assumption can also be supported by the fact that only few of the energy dissipating cycles in Arabidopsis (9 from each analysis) have reactions with positive protein expressions. Even if all enzymes are assumed to be active, the flux through these cycles could be small or negligible as a result of shared flux in a big network (Gebauer et al., 2012). In this regard, the ED nodes identified from the correlation analysis are more likely to occur *in vivo* as they only involve 2-3 reactions in the cycle.

Even though the biological functions of all the energy dissipation modes cannot be established without further experimental evidences, from the modelling point of view this investigation have successfully listed all potential candidates that could occur in different part of metabolism. This could be used as a guide to design new experiments. Moreover, the investigation demonstrates the applicability of two new methods developed here to study structural properties of a metabolic network and further explore its capabilities.

### 6.5.4 Increase in biomass composition after removing reactions involved in energy dissipation

The knowledge of mechanism that are employed by plants during environmental fluctuations can provide novel targets for improving crop yield (Raines, 2011). Experimental results that demonstrate improved growth or biomass production after engineering metabolic system involved with energy dissipation are being published regularly. More recently, Kromdijk et al. (2016) have shown that by increasing the amount of PSII subunit and accelerating the interconversion of violaxanthin and zeaxanthin improves the production of biomass by 15% in natural conditions. Similarly, manipulation of the xanthophyll cycle, involved in an energy dissipating cycle, has been shown to improve the adaptation to change in light intensity in the tobacco plant (Ruban, 2017).

In this analysis, after all the reactions involved in energy dissipating modes (all unique reactions involved in alternate solutions) were removed from the Arabidopsis model, a significant increase in the production of biomass was observed (Figure

6.7). This indicates that the 198 reactions in the case of Arabidopsis could be the potential knockout targets to help improve productivity in plants. As shown in the Figure 6.8, the biomass components that show increase in flux with increasing photon flux are all the glucose derivatives and are produced primarily using central carbon metabolic reactions. The observation strongly suggest the removal of the energy dissipating reactions from the model could improve the performance of central carbon metabolism and direct the metabolic flux to convert $CO_2$ into biomass.

Overall, this study presented two robust methods that were used to identify all possible energy dissipating metabolic cycles in the GSMs. Further, with the integration of proteomics data, the analysis have been able to propose 19 metabolic cycles which are most likely to occur *in vivo* and dissipate excess energy under high light conditions. Similarly, based on the light scan analysis results, the study propose that removal of reactions involved in energy dissipation mechanism helps to improve the performance of central carbon metabolism and hence improve the production of biomass in plants.

# Chapter 7

# General summary

## 7.1 Overview of results obtained

In this thesis, newly constructed GSMs of *A. thaliana* and *C. reinhardtii* were used to study different aspects of plant and algal photosynthetic metabolism. The models were developed using the latest annotations of the respective BioCyc databases, thus including large coverage of the genomes and hence they represent the metabolic characteristics of the organisms in more detail compared to their predecessors. Further, metabolic modelling approaches such as elementary mode analysis (Chapter 5) FBA (Chapter 5, 4 6), reaction correlation coefficient analysis (Chapter 6) and MILP (Chapter 6) were used to study metabolic fluxes under different light conditions to explore the potential of the metabolic networks in achieving important biological functions.

Chapter 3, describes the construction of fully compartmentalised GSM of *A. thaliana* and *C. reinhardtii*. An iterative process of curation was carried out with both the models to make sure that the direction and reversibility of all reactions comply with the law of mass and energy conservation. The models were checked for their ability to produce biomass components under both heterotrophic and phototrophic conditions in *A. thaliana* model and also under mixotrophic conditions in case of the *C. reinhardtii* model. The resulting *A. thaliana* model has 2588 reactions and 2481 metabolites, divided into 5 compartments according to the information obtained from Cheung et al. (2013). The *C. reinhardtii* model has 1858 reactions and 1931 metabolites, localised into 4 compartments based on

information inherited from Chang et al. (2011). Comparative study of the models with their predecessors have demonstrated that it is important to include newer genome annotations in the model reconstructions as they influence the structure of the resulting models and hence their predictive capacity. The analysis has also emphasised that only a small fraction of reactions in the model, including central carbon metabolism, are required to support its growth. Additionally, the chapter also describes the methods and results for experimental measurements of biomass components of Chlamydomonas under mixotrophic conditions, using a torus photobioreactor (Takache et al., 2012). The results show that the *C. reinhardtii* produces more biomass components under mixotrophic condition although the proportions of individual biomass components were comparable to its phototrophic growth.

Chapter 4 was focused on the study of metabolic responses of the models with different energy and carbon sources. The *A. thaliana* model was analysed under heterotrophic and phototrophic conditions. The analysis of reaction fluxes have demonstrated the capacity of the model to render naturally occurring regulatory mechanisms such a thioredoxin regulation of light active and dark active reactions, without explicitly setting such constraints. The analysis under changing light conditions highlighted the shift in control of energy metabolism from mitochondrial oxidative phosphorylation to chloroplastic photophosphorylation when the availability of light increased. Under such situations the redox transfer between the compartments was taken care by a shuttle mechanism. As the input of photon flux increased, the reactions involved in the energy dissipation mechanism were observed to carry more flux. An unique combination of reactions involving the xanthophyll cycle with the ascorbate glutathione cycle and a cycle involving interconversion of glutamate-glutamine were identified to be involved in oxidation of excess reductant energy under high light conditions. Moreover, the analysis led to a discovery of a pattern in the network that a group of reactions belonging to different parts of metabolism could combine together to function to utilise excess energy in the system. The observation exhibited a scope for further research in energy dissipating cycles in the network. The analysis of *C. reinhardtii* model under mixotrophic metabolism supports previous observations that the organism is capable of using both $CO_2$ and acetate to support its growth. The analysis in this chapter has also illustrated the importance of central carbon metabolism in

supporting the growth of organisms.

Chapter 5 was centered on the study of the effects of knock-out and deregulation of the Calvin cycle enzymes in the *A. thaliana* model. The FBA based analysis on the model revealed that the a feasible solution is possible after removal of reactions catalysed by SBPase, FBPase and GAPdh, indicating that such knockouts can be potentially viable *in vivo*. The observations from model analysis are consistent with the experimental results which show that, albeit with significantly impaired phenotypes, single knockout mutants of these enzymes are viable. Moreover, the analysis on the GSM was useful to investigate the rearrangement of the fluxes under the knockout conditions. The analyses have suggested alternate reactions that are capable of supporting the metabolism in the absence of the knocked out reactions. It has emphasised that SBPase and FBPase have the capacity to compensate each other's activity and thus the network is able to maintain all fluxes to biomass components even when one of them is removed. Moreover the analysis also proposed a novel role of the dark active, tranaldolase reaction, in day time metabolism. The reaction was not present in the model flux distributions under light conditions and is known to be active only under the dark condition. However, it was carrying a positive flux under both SBPase and FBPase knockout conditions. Notably under all knockout conditions, the major changes in the fluxes were observed among the Calvin cycle reactions in the chloroplast and some of their isoforms in the cytosol which further illustrates its importance in supporting photosynthetic metabolism. The deregulation of the SBPase and FBPase in dark metabolism suggested that such change will generate a lot of reducing potential in the system which can be lethal to the organism.

Chapter 6 followed up the investigation of metabolism under high light conditions undertaken in Chapter 4 explicitly focusing on identification of all possible combinations of reactions in the model that can potentially involve in energy dissipation mechanisms. For this, a photon absorbing core model was identified by filtering the reactions which showed positive correlation with the photon transporter. The core models for both *A. thaliana* and *C. reinhardtii* contained less than half of the reactions from their respective GSMs. All the reactions present in the core model are capable of carrying a steady state flux and thus are likely to respond under high light conditions. The aim was to identify the sets of reactions

that can operate simultaneously with net stoichiometry of either hydrolysing an ATP or oxidising equivalent reductants (NAD, NADPH). Two different approaches were then used to identify such energy dissipating cycles. In the first approach, a correlation coefficient based metabolic tree was generated. The reactions which are closely correlated and thus clustered together as children of a parent node were isolated and evaluated for their ability to operate as the energy dissipating node. In total, 56 such cycles were identified in the Arabidopsis model and 34 in the Chlamydomonas model. In the second approach, MILP was used to identify all possible alternate routes through the network that could consume a photon, generate energy through the activity of either cyclic or non-cyclic photophosphorylation and subsequently consume the ATP or NADPH generated without a net carbon conversion. In total, 231 such energy dissipating modes were identified in Arabidopsis model, 211 of which were also elementary modes. In the case of the Chalamydomonas model, 191 such energy dissipating modes were identified out of which 171 were also the elementary modes. Moreover, proteomics data obtained under high light conditions was integrated with the model to identify which of the nodes and modes identified from this study are more likely to occur *in vivo*. In total, 9 nodes and 10 modes identified from this analysis are found to have increased protein expression. Further, when the reactions involved in energy dissipation mechanisms identified from the analysis were removed, the model was able to predict an improvement in performance of the central carbon metabolism leading to an increase in production of biomass components.

Hence, while establishing some of the well known properties of photosynthesis, the study presented here has been able to predict new features of the metabolic networks and propose further studies which could help improve the productivity of plants and algae. Moreover, the study has also demonstrated the potential of using metabolic modelling tools in new areas of photosynthetic research.

## 7.2 Scope of the model and methods developed during this study

Analyses presented here have demonstrated that the newly constructed GSMs are capable of correctly representing well known biological phenomena such as thiore-

doxin regulation without explicitly setting such constraints. This shows their potential to be used in studying other regulatory and mechanistic properties of the organism. Accordingly, the ability of the model to correctly predict viability of Calvin cycle reactions knockout mutants demonstrate that the model can easily be adapted to study gene essentiality or mutational properties in other parts of the metabolic network. Similarly, since the models already represent the entire genome of the organism and contain all the reactions belonging to secondary metabolism it can easily adapted to study secondary metabolic properties by including additional metabolite exporters.

It would be worthwhile to design experiments to test the activity of transaldoase in day time metabolism, as predicted by the model. Moreover, additional experimental data such as biomass composition and other photosynthetic measurements can also be integrated in the model to more accurately predict phenotypic impact of such knockouts. Similarly, the proteomics data analysis has been useful to indicate which of the energy dissipating cycles identified from the modelling analysis are more likely to occur *in vivo*. However, this is not a proof that such mechanism exist in reality. It would be worthwhile to design new experiments to test which of the metabolic cycles identified here are active under real biological conditions. This would help develop newer understanding of photo-protection mechanism in these organisms.

Additionally, the modelling techniques employed during this study such as use of correlation coefficients to identify the photon absorbing core model can be used to generate core models to represent other metabolic characteristics. Likewise, the method and modules developed to identify energy dissipating cycles, can easily be adapted for similar analysis on other photosynthetic organisms.

## 7.3 Limitations of the model and future developments

Although the GSMs developed here cover detailed metabolic characteristics of the organisms and have been able to make interesting predictions, there are certain aspects which could be further improved. For example the subcellular localisa-

tions for both the models were adapted from their predecessors. Although there is much less experimental data to support accurate localisation of many reactions in the metabolism, more updated information for Arabidopsis has been published recently (Hooper et al., 2017) and can be adapted for future studies. Similarly, the tools such as PredAlgo (Tardif et al., 2012) can be used to further identify compartmental localisation in Chlamydomonas.

Metabolite transporters are an integral part of metabolism; however there are relatively fewer studies about the mechanisms of intercompartmental transporters in both plant and algal species. The analysis here has identified that transporters of energy components in particular influence the metabolic responses of the model. For example the ATP-ADP-Pi antiporter was observed to couple with ATP hydrolysing reactions and act as a spurious energy consuming cycle. Thus future development of the models can consider these issues in more detail.

Although, the models include secondary metabolic reactions more emphasis was given on the study of primary metabolism in this thesis. Thus, future development of models can also focus on improving these aspects. Moreover, the plant leaf is a complex organ and is composed of multiple tissues and cell types where each tissue type has different metabolic characterists. There have been studies which pieced together a tissue specific network of a plant species (Saha et al., 2011) but it lacks the representation of interactions between different cell types. Future developments of the models can focus on developing cell or tissue specific model and ultimately move towards developing whole organism models.

## 7.4 Conclusion

Research on plant and algal biology is aimed at improving their photosynthetic productivity to address increasing demand for food and energy supplies. Metabolic models have been useful tools to study behavior and complexity of biological systems. This study in particular has successfully demonstrated that metabolic models can be used as a framework to study metabolic phenotypes and generate predictions that will provide a guide to design new experiments with more certainty of what to expect. Similarly, the study has also highlighted how metabolic models can be effectively integrated with existing experimental data and complement

in developing our understanding of underlying properties of photosynthetic metabolism and further explore its potential to achieve new functions. Moreover, this study has been able to demonstrate that metabolic models are important tools for integrative, multidisciplinary research (Moejes et al., 2017) and can contribute in collaborative research projects like AccliPhot. It is hoped that the information covered by the models, the modelling approaches developed and the predictions made in this study will be translated into crop plants and help improve their productivity.

# Bibliography

Abu-Ghosh, S., Fixler, D., Dubinsky, Z. and Iluz, D. (2016), 'Flashing light in microalgae biotechnology', *Bioresour Technol.* **203**, 357–363.

Ågren, J., Oakley, C. G., McKay, J. K., Lovell, J. T. and Schemske, D. W. (2013), 'Genetic mapping of adaptation reveals fitness tradeoffs in Arabidopsis thaliana', *PNAS* **110**(52), 21077–21082.

Alonso, A. P., Vigeolas, H., Raymond, P., Rolin, D. and Dieuaide-Noubhani, M. (2005), 'A New Substrate Cycle in Plants. Evidence for a High Glucose-Phosphate-to-Glucose Turnover from in Vivo Steady-State and Pulse-Labeling Experiments with 13 C Glucose and 14 C Glucose', *Plant Physiol.* **138**, 2220–2232.

Anderson, L. E. (1981), 'Light inactivation of transaldolase in pea leaf chloroplasts', *Biochem Biophys Res Commun.* **99**(4), 1199–202.

Andrews, J. and Whitney, S. (2003), 'Manipulating ribulose bisphosphate carboxylase/oxygenase in the chloroplasts of higher plants.', *Arch Biochem Biophys.* **414**(2), 159–69.

Anoman, A. D., Flores-Tornero, M., Rosa-Telléz, S., Muñoz-Bertomeu, J., Segura, J. and Ros, R. (2016), 'The specific role of plastidial glycolysis in photosynthetic and heterotrophic cells under scrutiny through the study of glyceraldehyde-3-phosphate dehydrogenase', *Plant Signal Behav* **110**(3).

Anoman, A. D., Muñoz-Bertomeu, J., Rosa-Téllez, S., Flores-Tornero, M., Serrano, R., Bueso, E., Fernie, A. R., Segura, J. and Ros, R. (2015), 'Plastidial Glycolytic Glyceraldehyde-3-Phosphate Dehydrogenase Is an Important Deter-

minant in the Carbon and Nitrogen Metabolism of Heterotrophic Cells in Arabidopsis', *Plant Physiol.* **169**(3), 1619–1637.

Arabidopsis Genome Initiative (2000), 'Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.', *Nature* **408**(6814), 796–815.

Arnold, A. and Nikoloski, Z. (2014), 'Bottom-up Metabolic Reconstruction of Arabidopsis and Its Application to Determining the Metabolic Costs of Enzyme Production', *Plant Physiol.* **165**(3), 1380–1391.

Baltz, A., Dang, K.-V., Beyly, A., Auroy, P., Richaud, P., Cournac, L. and Peltier, G. (2014), 'Plastidial Expression of Type II NAD(P)H Dehydrogenase Increases the Reducing State of Plastoquinones and Hydrogen Photoproduction Rate by the Indirect Pathway in Chlamydomonas reinhardtii1.', *Plant Physiol* **165**(3), 1344–1352.

Bassham, J. (2003), 'Mapping the carbon reduction cycle: a personal retrospective', *Photosynth Res* **76**(1-3), 37–52.

Bassham, J., Benson, A. and Calvin, M. (1950), 'The path of carbon in photosynthesis.', *J Biol Chem* **185**(2), 781–787.

Beckers, V., Dersch, L. M., Lotz, K., Melzer, G., Bläsing, O. E., Fuchs, R., Ehrhardt, T. and Wittmann, C. (2016), 'In silico metabolic network analysis of Arabidopsis leaves', *BMC Systems Biology* **10**(102), 1–19.

Behre, J., Wilhelm, T., von Kamp, A., Ruppin, E. and Schuster, S. (2008), 'Structural robustness of metabolic networks with respect to multiple knockouts', *J Theor Biol* **252**(3), 433–41.

Benson, A. (2002), 'Following the path of carbon in photosynthesis: a personal story.', *Photosynth Res* **73**(1-3), 29–49.

Beste, D. J. V., Hooper, T., Stewart, G., Bonde, B., Avignone-Rossa, C., Bushell, M. E., Wheeler, P., Klamt, S., Kierzek, A. M. and McFadden, J. (2007), 'GSMN-TB: a web-based genome-scale network model of *Mycobacterium tuberculosis* metabolism', *Genome Biology* **8**(5).

Bevan, M. and Walsh, S. (2005), 'The Arabidopsis genome: A foundation for plant research', *Genome Res* **15**, 1632–1642.

Boyle, N. R. and Morgan, J. A. (2009), 'Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii.', *BMC Systems Biology* **3**, 4.

Brautigam, A. and Weber, A. P. (2011), 'Do Metabolite Transport Processes Limit Photosynthesis?', *Plant Physiol* **155**(1), 43–48.

Brower, V. (2008), 'Back to Nature: Extinction of Medicinal Plants Threatens Drug Discovery', *JNCI J Natl Cancer Inst* **100**(12), 838–839.

Buchanan, B. B. (1980), 'Role of Light in the Regulation of Chloroplast Enzymes', *Ann. Rev. of Plant Physol* **31**, 341–374.

Buchanan, B. B. (1991), 'Regulation of CO2 assimilation in oxygenic photosynthesis: The ferredoxin/thioredoxin system: Perspective on its discovery, present status, and future development', *Arch Biochem Biophys* **288**(1), 1–9.

Buchanan, B. B. and Balmer, Y. (2005), 'Redox regulation: a broadening horizon.', *Annu Rev Plant Biol* **56**, 187–220.

Burris, J. E. (1981), 'Effects of oxygen and inorganic carbon concentrations on the photosynthetic quotients of marine algae.', *Marine Biol.* **65**(3), 215–219.

Bustos, D. M., Bustamante, C. A. and Iglesias, A. A. (2008), 'Involvement of non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase in response to oxidative stress.', *J Plant Physiol* **165**(4), 456–461.

Caillau, M. and Quick, W. P. (2005), 'New insights into plant transaldolase', *Plant J.* **43**, 1–16.

Carlson, R. and Srienc, F. (2004), 'Fundamental E.coli biochemical pathways for biomass and energy production: Identification of reactions.', *Biotechnol. Bioeng.* **85**(1), 1–19.

Carrari, F., Nunes-Nesi, A., Gibon, Y., Lytovchenko, A., Loureiro, M. E. and Fernie, A. R. (2003), 'Reduced Expression of Aconitase Results in an Enhanced Rate of Photosynthesis and Marked Shifts in Carbon Partitioning in Illuminated Leaves of Wild Species Tomato', *Plant Physiol* **133**(3), 1322–1335.

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q.,

Paley, S., Subhraveti, P., Weaver, D. S. and Karp, P. D. (2016), 'The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases.', *Nucleic Acids Res* **44**(D1), D471–80.

Caspi, R., Foerster, H., Fulcher, C. A., Hopkinson, R., Ingraham, J., Kaipa, P., Krummenacker, M., Paley, S., Pick, J., Rhee, S. Y., Tissier, C., Zhang, P. and Karp, P. D. (2006), 'MetaCyc: a multiorganism database of metabolic pathways and enzymes.', *Nucleic Acids Res* **34**(Database issue), D511–6.

Çakr, T. and Khatibipour, M. J. (2014), 'Metabolic Network Discovery by Top-Down and Bottom-Up Approaches and Paths for Reconciliation', *Front Bioeng Biotechnol.* **2**(62), 1–11.

Chance, E. M., Curtis, A. R., Jones, I. P. and Kirby, C. R. (1977), *FACSIMILE: a Computer Program for Flow and Chemistry Simulation, and General Initial Value Problems.* AERE–R8775, HMSO London.

Chang, R. L., Ghamsari, L., Manichaikul, A., Hom, E. F. Y., Balaji, S., Fu, W., Shen, Y., Hao, T., Palsson, B. Ã., Salehi-Ashtiani, K. and Papin, J. A. (2011), 'Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism.', *Mol Syst Biol* **7**, 518.

Chaouiya, C. (2007), 'Petri net modelling of biological networks.', *Briefings in Bioinformatics* **8**(4), 210–219.

Chapman, S. P., Paget, C. M., Johnson, G. N. and Schwartz, J.-M. (2015), 'Flux balance analysis reveals acetate metabolism modulates cyclic electron flow and alternative glycolytic pathways in Chlamydomonas reinhardtii', *Front Plant Sci.* **6**(474), 1–14.

Chatterjee, A. and Kundu, S. (2015), 'Revisiting the chlorophyll biosynthesis pathway using genome scale metabolic model of Oryza sativa japonica.', *Sci Rep.* **5**(14975).

Cheung, C. M., Poolman, M. G., Fell, D. A., Ratcliffe, R. G. and Sweetlove, L. J. (2014), 'A Diel Flux Balance Model Captures Interactions between Light and Dark Metabolism during Day-Night Cycles in C3 and Crassulacean Acid Metabolism Leaves', *Plant Physiol.* **165**(2), 917–929.

Cheung, C. M., Ratcliffe, R. G. and Sweetlove, L. J. (2015), 'A Method of Accounting for Enzyme Costs in Flux Balance Analysis Reveals Alternative Pathways and Metabolite Stores in an Illuminated Arabidopsis Leaf.', *Plant Physiol* **169**(3), 1671–1682.

Cheung, C. Y. M., Williams, T. C. R., Poolman, M. G., Fell, D. A., Ratcliffe, R. G. and Sweetlove, L. J. (2013), 'A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions.', *Plant J* **75**(6), 1050–61.

Chiadmi, M., Navaza, A., Miginiac-Maslow, M., Jacquot, J. P. and Cherfils, J. (1999), 'Redox signalling in the chloroplast: structure of oxidized pea fructose-1,6-bisphosphate phosphatase.', *EMBO J.* **18**(23), 6809–6815.

Cogne, G., gen, M. R., Bockmayr, A., Titica, M., Dussap, C.-G., Cornet, J.-F. and Legrand, J. (2011), 'A Model-Based Method for Investigating Bioenergetic Processes in Autotrophically Growing Eukaryotic Microalgae: Application to the Green Algae Chlamydomonas reinhardtii', *Amc Ins. of Chem. Eng* **27**, 631–640.

Cornish-Bowden, A. and Hofmeyr JH, C. M. (2002), 'Stoicheiometric analysis in studies of metabolism', *Biochem Soc Trans* **30**(2), 43–46.

Davis, M. C., Fiehn, O. and Durnford, D. G. (2013), 'Metabolic acclimation to excess light intensity in Chlamydomonas reinhardtii.', *Plant Cell Environ.* **36**(7), 1391–405.

de Oliveira Dal'Molin, C. G., Quek, L.-E., Palfreyman, R. W., Brumbley, S. M. and Nielsen, L. K. (2010), 'AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis.', *Plant Physiol* **152**(2), 579–89.

de Oliveira Dal'Molin, C. G., Quek, L.-E., Palfreyman, R. W. and Nielsen, L. K. (2011), 'AlgaGEM–a genome-scale metabolic reconstruction of algae based on the Chlamydomonas reinhardtii genome.', *BMC Genomics* **12 Suppl 4**, S5.

del Arco, A. and Satrústegui, J. (2005), 'New mitochondrial carriers: an overview', *Cell Mol Life Sci.* **62**(19-20), 2204–27.

Demmig-Adams, B. and Adams, W. W. (1996), 'The role of xanthophyll cycle carotenoids in the protection of photosynthesis.', *Trends in Plant Science* **1**, 21–26.

Ding, F., Wang, M., Zhang, S. and Ai, X. (2016), 'Changes in SBPase activity influence photosynthetic capacity, growth, and tolerance to chilling stress in transgenic tomato plants.', *Sci Rep.* **6**, 1–14.

Edgerton, M. D. (2009), 'Increasing Crop Productivity to Meet Global Needs for Feed, Food, and Fuel', *Plant Physiol.* **149**(1), 7–13.

Famili, I. and Palsson, B. O. (2003), 'Systemic metabolic reactions are obtained by singular value decomposition of genome-scale stoichiometric matrices', *J Theor Biol* **224**(1), 87–96.

Farquhar, G. D., von Caemmerer, S. and Berry, J. A. (2001), 'Models of photosynthesis.', *Plant Physiol* **125**(1), 42–45.

Feist, A. M. and Palsson, B. A. (2008), 'The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli.', *Nat Biotechnol* **26**(6), 659–67.

Fell, D. (1997), *Understanding the Control of Metabolism*, Portland Press, London.

Fell, D. A., Poolman, M. G. and Gevorgyan, A. (2010), 'Building and analysing genome-scale metabolic models.', *Biochem Soc Trans* **38**(5), 1197–201.

Fell, D. A. and Small, J. R. (1986), 'Fat synthesis in adipose tissue. An examination of stoichiometric constraints.', *Biochem J* **238**(3), 781–6.

Fernie, A. R., Carrari, F. and Sweetlove, L. J. (2004), 'Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport.', *Curr Opin Plant Biol.* **7**(3), 254–61.

Field, B. and Randerson, F. (1998), 'Primary production of the biosphere: integrating terrestrial and oceanic components', *Science* **281**(5374), 237–40.

Flügge, U.-I., Häusler, R. E., Ludewig, F. and Gierth, M. (2011), 'The role of transporters in supplying energy to plant plastids.', *J Exp Bot.* **2011**(7), 2381–92.

Flugge, U. I. and Heldt, H. (1984), 'The phosphate-triose phosphate-phos- phoglycerate translocator of the chloroplast', *Trends Biochem Sci* **9**, 530–533.

Flugge, U.-I. and Heldt, H. W. (1991), 'Metabolite Translocators of the Chloroplast Envelope', *Annu. Rev. Plant Physiol, Plnt. Mol. Biol.* **42**, 129–144.

Foyer, C. H. and Noctor, G. (2005), 'Oxidant and antioxidant signalling in plants: a re-evaluation of the concept of oxidative stress in a physiological context.', *Plant Cell Environ* **28**(8), 1056–1071.

Gagneur, J. and Klamt, S. (2004), 'Computation of elementary modes: a unifying framework and the new binary approach', *BMC Bioinformatics* **5**(175).

Garfinkel, D. (1968), 'A Machine–Independent Language for the Simulation of Complex Chemical and Biochemical Systems.', *Comput. Biomed. Res.* **2**, 31–44.

Gebauer, J., Schuster, S., de Figueiredo, L. F. and Kaleta, C. (2012), 'Detecting and investigating substrate cycles in a genome-scale human metabolic network', *FEBS J.* **279**, 3192–3202.

Geigenberger, P. and Stitt, M. (1991), 'A "futile" cycle of sucrose synthesis and degradation is involved in regulating partitioning between sucrose, starch and respiration in cotyledons of germinating Ricinus communis L. seedlings when phloem transport is inhibited.', *Planta.* **185**(1), 81–90.

Geiger, D. R. and Servaites, J. C. (1994), 'Diurnal Regulation of Photosynthetic Carbon Metabolism in C3 Plants', *Annu Rev Plant Physiol Plant Mol Biol* **45**, 235–56.

Gevorgyan, A., Poolman, M. G. and Fell, D. A. (2008), 'Detection of stoichiometric inconsistencies in biomolecular models.', *Bioinformatics* **24**(19), 2245–51.

Goss, R. and Jakob, T. (2010), 'Regulation and function of xanthophyll cycle-dependent photoprotection in algae.', *Photosynth Res* **106**(1-2), 103–22.

Grobbelaar, J. U. (1991), 'The Influence of Light/Dark Cycles in Mixed Algal Cultures on their Productivity ', *Bioresource Technology* **38**, 189–194.

Großeholz, R., Koh, C.-C., Veith, N., Fiedler, T., Strauss, M., Olivier, B., Collins, B. C., Schubert, O. T., Bergmann, F., Kreikemeyer, B., Aebersold, R. and Kummer, U. (2016), 'Integrating highly quantitative proteomics and genome-scale metabolic modeling to study pH adaptation in the human pathogen Enterococcus faecalis', *Nat. Sys. Biol. and Appln.* **2**, 1–9.

Gutiérrez, R. A., Shasha, D. E. and Coruzzi, G. M. (2005), 'Systems Biology for the Virtual Plant', *Plant Physiol.* **138**(2), 550–554.

Habenicht, A., Hellman, U. and Cerff, R. (1994), 'Non-phosphorylating GAPDH of Higher Plants is a Member of the Aldehyde Dehydrogenase Superfamily with no Sequence Homology to Phosphorylating GAPDH', *Journal of Molecular Biology* **237**(1), 165–171.

Haferkamp, I. (2007), 'The diverse members of the mitochondrial carrier family in plants', *FEBS Lett.* **581**(12), 2375–2379.

Harrison, E. P., Olcer, H., Lloyd, J. C., Long, S. P. and Raines, C. A. (2001), 'Small decreases in SBPase cause a linear decline in the apparent RuBP regeneration rate, but do not affect Rubisco carboxylation capacity.', *J. Exp. Bot.* **52**(362), 1779–1784.

Hartman, H. (2013), Genome-scale metabolic modelling of Salmonella and Lactobacillus species, PhD thesis, Oxford Polytechnic.

Hay, J. and Schwender, J. (2014), 'Flux Variability Analysis: Application to Developing Oilseed Rape Embryos Using Toolboxes for Constraint-Based Modeling', *Methods Mol Biol.* **1090**, 301–16.

Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Smal, I. and Millar, A. H. (2007), 'SUBA: the Arabidopsis Subcellular Database', *Nucl Acids Res* **35**(D213), 1–6.

Heber, U. and Santarius, K. A. (1970), 'Direct and indirect transfer of ATP and ADP across the chloroplast envelope.', *Z Naturforsch B.* **25**(7), 718–28.

Heifetz, P. B., Förster, B., Osmond, C. B., Giles, L. J. and Boynton, J. E. (2000), 'Effects of acetate on facultative autotrophy in Chlamydomonas reinhardtii assessed by photosynthetic measurements and stable isotope analyses.', *Plant Physiol.* **122**(4), 1439–45.

Heineke, D., Riens, B., Grosse, H., Hoferichter, P., Peter, U., Flugge, U. I. and Heldt, H. W. (1991), 'Redox Transfer across the Inner Chloroplast Envelope Membrane.', *Plant Physiol* **95**(4), 1131–7.

Heinrich, R. and Schuster, S. (1996), *The regulation of cellular systems*, Chapman & Hall, London, England.

Heldt, H.-W. and Piechulla, B. (2011*a*), *Plant Biochemisty*, Fourth Edition, Elsevier Academic Press, UK.

Heldt, H.-W. and Piechulla, B. (2011*b*), *Plant Biochemisty*, Fourth Edition, Elsevier Academic Press.

Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M. and Stevens, R. L. (2010), 'High-throughput generation, optimization and analysis of genome-scale metabolic models.', *Nat Biotec* **28**, 977–982.

Hers, H. and Hue, L. (1983), 'Gluconeogenesis and related aspects of glycolysis.', *Annu Rev Biochem.* **52**, 617–53.

Hofmeyr, J. H. (1986), 'Steady-state modelling of metabolic pathways: A guide for the prospective simulator', *Comput. Appl. Biosci.* **2**, 5–11.

Holzhütter, H.-G. (2006), 'The generalized flux-minimization method and its application to metabolic networks affected by enzyme deficiencies.', *Biosystems* **83**(2-3), 98–107.

Hooper, C. M., Castleden, I. R., Tanz, S. K., Aryamanesh, N. and Millar, A. H. (2017), 'SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations.', *Nucl Acids Res* **45**(D1064–D1074), 1–11.

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P. and Kummer, U. (2006), 'COPASI a COmplex PAthway SImulator', *Bioinformatics* **22**(24), 3067–3074.

Howard, T. P., Metodiev, M., Lloyd, J. C. and Raines, C. A. (2008), 'Thioredoxin-mediated reversible dissociation of a stromal multiprotein complex in response to changes in light availability.', *Proc Natl Acad Sci U S A* **105**(10), 4056–61.

Imam, S., Schäuble, S., Valenzuela, J., de Lomana, A. L. G., Carter, W., Price, N. D. and Baliga, N. S. (2015), 'A refined genome-scale reconstruction of Chlamydomonas metabolism provides a platform for systems-level analyses', *Plant J.* **84**(6), 1239–56.

Ito, J., Parsons, H. T. and Heazlewood, J. L. (2014), 'The Arabidopsis cytosolic proteome: the metabolic heart of the cell', *Front. Plant Sci.* **5**(212), 1–8.

Jacquot, J.-P., Lancelin, J.-M. and Meyer, Y. (1997), 'Thioredoxins: structure and function in plant cells', *New Phytologist* **136**(4), 543–570.

Jamshidi, N. and Palsson, B. O. (2007), 'Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the in silico strain iNJ661 and proposing alternative drug targets', *BMC Systems Biology* **1**(26), 1–20.

Johnson, X. and Alric, J. (2013), 'Model simulations are consistent with published experimental data, which have shown that ICL activity is essential for efficient mixotrophic growth (Plancke et al., 2014). ', *Eukaryot Cell.* **12**(6), 776–93.

Johnson, X., Steinbeck, J., Dent, R. M. ., Takahashi, H., Richaud, P., Ozawa, S.-I., Laura Houille-Vernes, Petroutsos, D., Rappaport, F., Grossman, A. R., Niyog i, K. K., Hippler, M. and Alric, J. (2014), 'Proton gradient regulation 5-mediated cyclic electron flow under ATP- or redox-limited conditions: a study of DeltaATpase pgr5 and DeltarbcL pgr5 mutants in the green alga Chlamydomonas reinhardtii', *Plant Physiol.* **165**(1), 438–52.

Joliot, P. and Joliot, A. (2002), 'Cyclic electron transfer in plant leaf', *Proc. Natl. Acad. Sci. USA* **99**(15), 10209–10214.

Joliot, P. and Joliot, A. (2005), 'Quantification of cyclic and linear flows in plants', *Proc. Natl. Acad. Sci. USA* **102**(13), 4913–4918.

Jorg Stelling, Uwe Sauer, Zoltan Szallasi, Francis J. Doyle and John Doyle (2004), 'Robustness of Cellular Functions', *Cell* **118**, 675–685.

Joyard, J. and McCormick, S. (2010), 'Plant Systems Biology', *Plant Physiol* **152**(2), 401.

Kaleta, C., de Figueiredo, L. F., Behre, J. and Schuster, S. (2009), 'EFMEvolver: Computing Elementary Flux Modes in Genome-scale Metabolic Networks.', *Proceeding, Conference on Bioinformatics 2009* **157**, 179–189.

Kammerer, B., Fischer, K., Hilpert, B., Schubert, S., Michael Gutensohn, Weber, A. and Flügge, U.-I. (1998), 'Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose 6-phosphate/phosphate antiporter.', *Plant Cell.* **10**(1), 105–117.

Karp, P. D., Latendresse, M. and Caspi, R. (2011), 'The Pathway Tools Pathway Prediction Algorithm', *Stand Genomic Sci.* **5**(3), 424–429.

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O. and Lewis, N. E. (2016), 'BiGG Models: A platform for integrating, standardizing and sharing genome-scale models', *Nucleic Acids Res.* **4**(44), D515–D522.

Kinoshita, H., Nagasaki, J., Yoshikawa, N., Yamamoto, A., Takito, S., Kawasaki, M., Sugiyama, T., Miyake, H., Weber, A. P. M. and Taniguchi, M. (2011), 'The chloroplastic 2-oxoglutarate/malate transporter has dual function as the malate valve and in carbon/nitrogen metabolism', *Plant J* **65**, 15–26.

Klamt, S., Saez-Rodriguez, J. and Gilles, E. D. (2007), 'Structural and functional analysis of cellular networks with CellNetAnalyzer', *BMC Syst Biol* **1**(2), 1–13.

Klamt, S. and Stelling, J. (2002), 'Combinatorial complexity of pathway analysis in metabolic networks.', *Molecular Biology Reports* **29**(1-2), 233–236.

Klamt, S. and Stelling, J. (2003), 'Two approaches for metabolic pathway analysis?', *TIBS* **21**(2), 64–69.

Klingenberg, M. (2008), 'The ADP and ATP transport in mitochondria and its carrier', *Biochim Biophys Acta* **1778**(10), 1978–2021.

Knappe, S., Flügge, U.-I. and Fischer, K. (2003), 'Analysis of the Plastidic phosphate translocator Gene Family in Arabidopsis and Identification of New phosphate translocator-Homologous Transporters, Classified by Their Putative Substrate-Binding Site', *Plant Physiol.* **131**(3), 1178–1190.

Koch, I., Junker, B. H. and Heiner, M. (2005), 'Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber.', *Bioinformatics* **21**(7), 1219–1226.

Kramer, D. M. and Evans, J. R. (2010), 'The importance of energy balance in improving photosynthetic productivity.', *Plant Physiol* **155**(1), 70–78.

Kromdijk, J., Głowacka, K., Leonelli, L., Gabilly, S. T., Iwai, M., Niyogi, K. K. and Long, S. P. (2016), 'Improving photosynthesis and crop productivity by accelerating recovery from photoprotection.', *Science* **354**(6314), 857–861.

Kruger, E. L. and Volin, J. C. (2006), 'Reexamining the empirical relation between plant growth and leaf photosynthesis.', *Funct. Plant Biol* **33**, 421 – 429.

Ksas, B., Becuwe, N., Chevalier, A. and Havaux, M. (2015), 'Plant tolerance to excess light energy and photooxidative damage relies on plastoquinone biosynthesis', *Nat. Sci. Rep.* **5**(10919), 1–16.

Kumar, V. S., Dasika, M. S. and Maranas, C. D. (2007), 'Optimization based automated curation of metabolic reconstructions.', *BMC Bioinformatics* **8**(212).

Lakshmanan, M., Cheung, C. Y. M., Mohanty, B. and Lee, D.-Y. (2016), 'Modeling Rice Metabolism: From Elucidating Environmental Effects on Cellular Phenotype to Guiding Crop Improvement', *Front Plant Sci* **7**, 1795.

Lakshmanan, M., Lim, S.-H., Mohanty, B., Kim, J. K., Ha, S.-H. and Lee, D.-Y. (2015), 'Unraveling the Light-Specific Metabolic and Regulatory Signatures of Rice through Combined in Silico Modeling and Multiomics Analysis 1', *Plant Physiol* **169**, 3002–3020.

Laloi, M. (1999), 'Plant mitochondrial carriers: an overview', *Cell Mol Life Sci.* **56**(11-12), 918–44.

Lee, S.-K., Jeon, J.-S., Bornke, F., Voll, L., Cho, J.-I., Goh, C. H., Jeong, S. W., Park, Y., Kim, S. J., Choi, S. B., Miyao, A., Hirochika, H., An, G., Cho, M. H., Bhoo, S. H., Sonnewald, U. and Hahn, T.-R. (2008), 'Loss of cytosolic fructose-1,6-bisphosphatase limits photosynthetic sucrose synthesis and causes severe growth retardations in rice (Oryza sativa)', *Plant, Cell and Environment* **31**, 1851–1863.

Lee, S., Phalakornkule, C., Domach, M. M. and Grossmann, I. E. (2000), 'Recursive MILP model for finding all the alternate optima in LP models for metabolic networks.', *Computers and Chemical Engineering* **24**, 711–716.

Lefebvre, S., Lawson, T., Zakhleniuk, O. V., Lloyd, J. C. and Raines, C. A. (2005), 'Increased Sedoheptulose-1,7-Bisphosphatase Activity in Transgenic Tobacco Plants Stimulates Photosynthesis and Growth from an Early Stage in Development', *Plant Physiol.* **138**, 451–460.

Liao, Q., Li, L., Chen, R. and Zhu, X. (2014), 'A novel photobioreactor generating the light/dark cycle to improve microalgae cultivation.', *Bioresour Technol.* **161**, 186–91.

Linka, N. and Weber, A. P. (2010), 'Intracellular Metabolite Transporters in Plants', *Mol Plant* **3**(1), 21–53.

Liu, X.-L., Yu, H.-D., Guan, Y., Li, J.-K. and Guo, F.-Q. (2012), 'Carbonylation and Loss-of-Function Analyses of SBPase Reveal Its Metabolic Interface Role in Oxidative Stress, Carbon Assimilation, and Multiple Aspects of Growth and Development in Arabidopsis', *Molecular Plant* **5**(5), 1082–1099.

Lloyd, J. C., Raines, C. A., John, U. P. and Dyer, T. A. (1991), 'The chloroplast FBPase gene of wheat: structure and expression of the promoter in photosynthetic and meristematic cells of transgenic tobacco plants.', *Mol Gen Genet.* **225**(2), 209–16.

Logan, D. C. (2006), 'The mitochondrial compartment', *J Exp Bot.* **57**(6), 1225–1243.

López-Calcagno, P. E., Abuzaid, A. O., Lawson, T. and Raines, C. A. (2017), 'Arabidopsis CP12 mutants have reduced levels of phosphoribulokinase and impaired function of the Calvin–Benson cycle', *J. Exp. Bot.* pp. 1–14.

Lunn, J. E. (2007), 'Compartmentation in plant metabolism', *J. Exp. Bot.* **58**, 35–47.

Lutz, M. (2001), *Programming Python*, 2nd edn, O'Reilly & Associates.

Lutz, M. and Ascher, D. (1999), *Learning Python*, 1st edn, O'Reilly & Associates.

Maarleveld, T. R., Boele, J., Bruggeman, F. J. and Teusink, B. (2014), 'A Data Integration And Visualization Resource for the Metabolic Network of Synechocystis sp. PCC 6803 ', *Plant Physiol.* **164**(3), 1–11.

Machado, D. and Herrgård, M. (2014), 'Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism', *PLoS Comp. Biol* **10**(10), 1–12.

Mackenzie, S. and McIntosh, L. (1999), 'Higher Plant Mitochondria', *Plant Cell.* **11**(4), 571–586.

Makino, A., Miyake, C. and Yokota, A. (2002), 'Physiological functions of the water-water cycle (Mehler reaction) and the cyclic electron flow around PSI in rice leaves.', *Plant Cell Physiol.* **43**(9), 1017–26.

Manichaikul, A., Ghamsari, L., Hom, E. F. Y., Lin, C., Murray, R. R., Chang, R. L., Balaji, S., Hao, T., Shen, Y., Chavali, A. K., Thiele, I., Yang, X., Fan, C., Mello, E., Hill, D. E., Vidal, M., Salehi-Ashtiani, K. and Papin, J. A. (2009), 'Metabolic network analysis integrated with transcript verification for sequenced genomes.', *Nat Methods* **6**(8), 589–92.

Marri, L., Zaffagnini, M., Collin, V. and Trost, P. (2009), 'Prompt and Easy Activation by Specific Thioredoxins of Calvin Cycle Enzymes of Arabidopsis thaliana Associated in the GAPDH/CP12/PRK Supramolecular Complex', *Mol Plant* **2**(2), 259–69.

Martins, S. C. V., Araújo, W. L., Tohge, T., Fernie, A. R. and DaMatta, F. M. (2014), 'In high-light-acclimated coffee plants the metabolic machinery is adjusted to avoid oxidative stress rather than to benefit from extra light enhancement in photosynthetic yield.', *PLoS ONE* **9**(4), e94862.

Masakapalli, S. K., Lay, P. L., Huddleston, J. E., Pollock, N. L., Kruger, N. J. and Ratcliffe, G. (2010), 'Subcellular flux analysis of central metabolism in a heterotrophic Arabidopsis cell suspension using steady-state stable isotope labeling.', *Plant Physiol.* **152**(2), 602–619.

Masclaux-Daubresse, C., Reisdorf-Cren, M., Pageau, K., Lelandais, M., Grandjean, O., Kronenberger, J., Valadier, M.-H., Feraud, M., Jouglet, T. and Suzuki,

A. (2006), 'Glutamine Synthetase-Glutamate Synthase Pathway and Glutamate Dehydrogenase Play Distinct Roles in the Sink-Source Nitrogen Cycle in Tobacco', *Plant Physiol.* **140**(2), 444–456.

Maxwell, D. P., Wang, Y. and McIntosh, L. (1999), 'The alternative oxidase lowers mitochondrial reactive oxygen production in plant cells.', *Proc Natl Acad Sci U S A* **96**(14), 8271–6.

May, P., Christian, J.-O., Kempa, S. and Walther, D. (2009), 'ChlamyCyc: an integrative systems biology database and web-portal for Chlamydomonas reinhardtii', *BMC Genomics* **10**(209), 1–11.

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., MarAchal-Drouard, L., Marshall, W. F., Qu, L.-H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S. M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.-L., Cognat, V., Croft, M. T., Dent, R., Dutcher, S., FernA¡ndez, E., Fukuzawa, H., GonzA¡lez-Ballester, D., GonzA¡lez-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P. A., Lemaire, S. D., Lobanov, A. V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J. V., Moseley, J., Napoli, C., Nedelcu, A. M., Niyogi, K., Novoselov, S. V., Paulsen, I. T., Pazour, G., Purton, S., Ral, J.-P., Rian, D. M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S. L., Allmer, J., Balk, J., Bisova, K., Chen, C.-J., Elias, M., Gendler, K., Hauser, C., Lamb, M. R., Ledford, H., Long, J. C., Minagawa, J., Page, M. D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A. M., Yang, P., Ball, S., Bowler, C., Dieckmann, C. L., Gladyshev, V. N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R. T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y. W., Jhaveri, J., Luo, Y., Martnez, D., Ngau, W. C. A., Otillar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I. V., Rokhsar, D. S. and Grossman, A. R. (2007), 'The Chlamydomonas genome reveals the evolution of key animal and plant functions.', *Science* **318**(5848), 245–50.

Mettler, T., Mühlhaus, T., Hemme, D., Schöttler, M.-A., Rupprecht, J., Idoine,

A., Veyel, D., Pal, S. K., Yaneva-Roder, L., Winck, F. V., Sommer, F., Vosloh, D., Seiwert, B., Erban, A., Burgos, A., Arvidsson, S., Schönfelder, S., Arnold, A., Günther, M., Krause, U., Lohse, M., Kopka, J., Nikoloski, Z., Mueller-Roeber, B., Willmitzer, L., Bock, R., Schroda, M. and Stitt, M. (2014), 'Systems Analysis of the Response of Photosynthesis, Metabolism, and Growth to an Increase in Irradiance in the Photosynthetic Model Organism Chlamydomonas reinhardtii.', *Plant Cell* **26**(6), 2310–2350.

Michelet, L., Zaffagnini, M., Morisse, S., Sparla, F., Perez-Perez, M. E., Francia, F., Danon, A., Marchand, C. H., Fermani, S., Trost, P. and Lemaire, S. D. (2013), 'Redox regulation of the Calvin–Benson cycle: something old, something new', *Front Plant Sci* **4**(470).

Mintz-Oron, S., Meir, S., Malitsky, S., Ruppin, E., Aharoni, A. and Shlomi, T. (2011), 'Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity.', *Proc Natl Acad Sci U S A* **109**(1), 339–344.

Moejes, F. W., Matuszyńska, A., Adhikari, K., Bassi, R., Cariti, F., Cogne, G., Dikaios, I., Falciatore, A., Finazzi, G., Flori, S., Goldschmidt-Clermont, M., Magni, S., Maguire, J., Monnier, A. L., Müller, K., Poolman, M., Singh, D., Spelberg, S., Stella, G. R., Succurro, A., Taddei, L., Urbain, B., Villanova, V., Zabke, C. and Ebenhöh, O. (2017), 'A systems-wide understanding of photosynthetic acclimation in algae and higher plants', *J Exp Bot.* **erx**(137), 1–15.

Moretti, S., Martin, O., Tran, T. V. D., Bridge, A., Morgat, A. and Pagni, M. (2016), 'MetaNetX / MNXref – reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks', *Nucleic Acids Res* **4**(44(D1)), 523–6.

Morgan, B. J. T. and Ray, A. P. G. (1995), 'Non-uniqueness and inversions in cluster analysis', *Applied Statistics* **44**(11), 117–34.

Mueller, L. A., Zhang, P. and Rhee, S. Y. (2003), 'AraCyc: a biochemical pathway database for Arabidopsis.', *Plant Physiol* **132**(2), 453–60.

Müller, P., Li, X.-P. and Niyogi, K. K. (2001), 'Non-Photochemical Quenching. A Response to Excess Light Energy', *Plant Physiol.* **125**(4), 1558–1566.

Munoz-Bertomeu, J., Cascales-Miñana, B., Alaiz, M., Segura, J. and Ros, R. (2010), 'A critical role of plastidial glycolytic glyceraldehyde-3-phosphate dehydrogenase in the control of plant metabolism and development.', *Plant Signal Behav* **5**(1), 67–69.

Munoz-Bertomeu, J., Cascales-Minana, B., Mulet, J. M., Baroja-Fernandez, E., Pozueta-Romero, J., Kuhn, J. M., Segura, J. and Ros, R. (2009), 'Plastidial glyceraldehyde-3-phosphate dehydrogenase deficiency leads to altered root development and affects the sugar and amino acid balance in Arabidopsis.', *Plant Physiol* **151**(2), 541–58.

Neuhaus, H. E., Henrichs, G. and Scheibe, R. (1993), 'Characterization of glucose-6- phosphate incorporation into starch by isolated intact cauliflower-bud plastids.', *Plant Physiol* **101**(2), 573–578.

Neuhaus, H. E. and Schulte, N. (1996), 'Starch degradation in chloroplasts isolated from C3 or CAM (crassulacean acid metabolism)-induced Mesembryanthemum crystallinum L.', *Biochem J.* **318**, 945–953.

Neuhaus, H. and Wagner, R. (2000), 'Solute pores, ion channels, and metabolite transporters in the outer and inner envelope membranes of higher plant plastids.', *Biochim Biophys Acta.* **1465**(1), 307–323.

Niewiadomski, P., Knappe, S., Geimer, S., Fischer, K., Schulz, B., Unte, U. S., Rosso, M. G., Ache, P., Flügge, U.-I. and Schneider, A. (2005), 'The Arabidopsis plastidic glucose 6-phosphate/phosphate translocator GPT1 is essential for pollen maturation and embryo sac development.', *Plant Cell* **17**(3), 760–75.

Nikkanen, L. and Rintamäki, E. (2014), 'Thioredoxin-dependent regulatory networks in chloroplasts under fluctuating light conditions.', *Philos Trans R Soc Lond B Biol Sci.* **369**(1640).

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. Ø. and Thiele, I. (2012), 'Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis', *Proc Natl Acad Sci U S A* **109**, 2678–83.

Noguchi, K. and Yoshida, K. (2008), 'Interaction between photosynthesis and respiration in illuminated leaves.', *Mitochondrion.* **8**(1), 87–99.

Oberhardt, M. A., Palsson, B. Ø. and Papin, J. A. (2009), 'Applications of genome-scale metabolic reconstructions', *Mol Syst Biol.* **5**.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H and Kanehisa M (1999), 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Res.* **27**, 29–34.

Olcer, H., Lloyd, J. C. and Raines, C. A. (2001), 'Photosynthetic Capacity Is Differentially Affected by Reductions in Sedoheptulose-1,7-Bisphosphatase Activity during Leaf Development in Transgenic Tobacco Plants', *Plant Physol* **125**, 982–989.

Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., Blankenship, R. E., Bock, R., Croce, R., Hanson, M. R., Hibberd, J. M., Long, S. P., Moore, T. A., Moroney, J., Niyog, K. K., Parry, M. A. J., Peralta-Yahya, P. P., Prince, R. C., Redding, K. E., Spalding, M. H., van Wijk, K. J., Vermaas, W. F. J., von Caemmerer, S., Weber, A. P. M., Yeates, T. O., Yuan, J. S. and Zhu, X. G. (2015), 'Redesigning photosynthesis to sustainably meet global food and bioenergy demand.', *Proc Natl Acad Sci U S A* **112**(28), 8529–8536.

Orth, J. D. and Palsson, B. Ø. (2010), 'Systematizing the Generation of Missing Metabolic Knowledge', *Biotechnol Bioeng.* **107**(3), 403–412.

Orth, J. D., Thiele, I. and Palsson, B. Ø. (2010), 'What is flux balance analysis?', *Nat. Biotec.* **28**, 245–248.

Padmasree, K., Padmavathi, L. and Raghavendra, A. (2002), 'Essentiality of mitochondrial oxidative metabolism for photosynthesis: optimization of carbon assimilation and protection against photoinhibition.', *Crit. Rev. in Biochem. and Mol. Bio* **376**(2), 71–119.

Papin, J. A., Price, N. D., Wiback, S. J., Fell, D. A. and Palsson, B. O. (2003), 'Metabolic pathways in the post-genomic era.', *Trends Biochem. Sci.* **28**(5), 250–258.

Park, D. J. M. and Wright, B. E. (1973), 'Metasim, a General Purpose Metabolic Simulator for Studying Cellular Transformations.', *Comput. Progm. Biomed.* **3**, 10–26.

Pearcy, R. W. (1990), 'Sunflecks and Photosynthesis in Plant Canopies', *Annu Rev Plant Physiol Plant Mol Biol* **414**, 421–53.

Perrière, G. and Gouy, M. (1996), 'WWW-query: An on-line retrieval system for biological sequence banks.', *Biochimie.* **78**(5), 364–9.

Pey, J., Villar, J. A., Tobalina, L., Rezola, A., Carrasco, J. M. G., Beasley, J. E. and Planes, F. J. (2015), 'TreeEFM: calculating elementary flux modes using linear optimization in a tree-based algorithm.', *Bioinformatics* **31**(6), 897–904.

Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J., Montero, F. and Schuster, S. (1999), 'Metatool: for studying metabolic networks.', *Bioinformatics* **15**(3), 251–257.

Philippar, K. and Soll, J. (2007), 'Intracellular Transport: Solute Transport in Chloroplasts, Mitochondria, Peroxisomes and Vacuoles, and Between Organelles', *Plant Solute Transport* pp. 133–192.

Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J. and Schomburg, D. (2016), 'BRENDA in 2017: new perspectives and new tools in BRENDA', *Nucl Acids Res* **45**(D1), 380–388.

Plancke, C., Vigeolas, H., Hohner, R., Roberty, S., Emonds-Alt, B., Larosa, V., Willamme, R., Duby, F., Dhali, D. O., Thonart, P., Hiligsmann, S., Franck, F., Eppe, G., Cardol, P., Hippler, M. and Remacle, C. (2014), 'Lack of isocitrate lyase in Chlamydomonas leads to changes in carbon metabolism and in the response to oxidative stress under mixotrophic growth.', *Plant J.* **77**(3), 404–17.

Plaxton, W. C. (1996), 'The organization and regulation of plant glycolysis.', *Plant Physiol.* **47**, 185–214.

Pohlmeyer, K., Paap, B. K., Soil, J. and Wedel, N. (1996), 'CP12: a small nuclear-encoded chloroplast protein provides novel insights into higher-plant GAPDH evolution.', *Plant Mol Biol* **32**(5), 969–78.

Polukhina, I., Fristedt, R., Dinc, E., Cardol, P. and Croce, R. (2016), 'Carbon Supply and Photoacclimation Cross Talk in the Green Alga Chlamydomonas reinhardtii.', *Plant Physiol.* **172**(3), 1494–1505.

Poolman, M. G. (2006), 'ScrumPy: metabolic modelling with Python.', *Syst Biol (Stevenage)* **153**(5), 375–8.

Poolman, M. G., Assmus, H. E. and Fell, D. A. (2004), 'Applications of metabolic modelling to plant metabolism', *J.Exp.Bot.* **55**(400), 1177–1186.

Poolman, M. G., Fell, D. A. and Raines, C. A. (2003), 'Elementary modes analysis of photosynthate metabolism in the chloroplast stroma.', *Eur J Biochem* **270**(3), 430–9.

Poolman, M. G., Fell, D. A. and Thomas, S. (2000), 'Modelling photosynthesis and its control', *J. Exp. Bot* **51**(1), 319–328.

Poolman, M. G., Kundu, S., Shaw, R. and Fell, D. A. (2013), 'Responses to light intensity in a genome-scale model of rice metabolism.', *Plant Physiol* **162**(2), 1060–72.

Poolman, M. G., Kundu, S., Shaw, R. and Fell, D. A. (2014), 'Metabolic trade-offs between biomass synthesis and photosynthate export at different light intensities in a genome–scale metabolic model of rice', *Front Plant Sci.* **5**(656), 1–7.

Poolman, M. G., Miguet, L., Sweetlove, L. J. and Fell, D. A. (2009), 'A genome-scale metabolic model of Arabidopsis and some of its properties.', *Plant Physiol* **151**(3), 1570–81.

Poolman, M. G., Sebu, C., Pidcock, M. K. and Fell, D. A. (2007), 'Modular decomposition of metabolic systems via null-space analysis', *Journal of Theoretical Biology* **249**, 691–705.

Qian, H. and Beard, D. (2006), 'Metabolic futile cycles and their functions: a systems analysis of energy and control.', *Syst Biol (Stevenage)* **153**(4), 192–200.

Radrich, K., Tsuruoka, Y., Dobson, P., Gevorgyan, A., Swainston, N., Baart, G. and Schwartz, J.-M. (2010), 'Integration of metabolic databases for the reconstruction of genome-scale metabolic networks.', *BMC Syst Biol* **4**, 114.

Raines, C. A. (2003), 'The Calvin cycle revisited: anaylsis of antisense plants.', *Photosynthesis Research* **75**(1), 1–10.

Raines, C. A. (2006), 'Transgenic approaches to manipulate the environmental responses of the C3 carbon fixation cycle', *Plant Cell Environ* **29**(3), 331–339.

Raines, C. A. (2011), 'Increasing Photosynthetic Carbon Assimilation in C3 Plants to Improve Crop Yield: Current and Future Strategies.', *Plant Physiol.* **155**(1), 36–42.

Raines, C. A., Harrison, E. P., Olcer, H. and Lloyd, J. C. (2000), 'Investigating the role of the thiol-regulated enzyme sedoheptulose-1,7-bisphosphatase in the control of photosynthesis', *Physol. Plant.* **110**, 303 – 308.

Raines, C. A., Lloyd, J. C., Willingham, N. M., Potts, S. and Dyer, T. A. (1992), 'cDNA and gene sequences of wheat chloroplast sedoheptulose-1,7-bisphosphatase reveal homology with fructose-1,6-bisphosphatases.', *Eur J Biochem.* **205**(3), 1053–9.

Raman, K. and Chandra, N. (2009), 'Flux balance analysis of biological systems: applications and challenges', *Briefing in Bioinf* **10**(4), 435–449.

Rausch, C. and Bucher, M. (2002), 'Molecular mechanisms of phosphate transport in plants. ', *Planta* **216**(1), 23–37.

Reiser, J., Linka, N., Lemke, L., Jeblick, W. and Neuhaus, H. E. (2004), 'Molecular Physiological Analysis of the Two Plastidic ATP/ADP Transporters from Arabidopsis', *Plant Physiol* **136**, 3524–3536.

Renberg, L., Johansson, A. I., Shutova, T., Stenlund, H., Aksmann, A., Raven, J. A., Gardeström, P., Moritz, T. and Samuelsson, G. (2010), 'A metabolomic approach to study major metabolite changes during acclimation to limiting CO2 in Chlamydomonas reinhardtii', *Plant Physiol.* **154**(1), 187–96.

Renne, P., Dreben, U., Hebbeker, U., Hille, D., gge U-I, F., Westhoff, P. and APM, W. (2003), 'The Arabidopsis mutant dct is deficient in the plastidic glutamate/malate translocator DiT2', *Plant J* **35**, 316–331.

Rittmann, B. E. (2008), 'Opportunities for renewable bioenergy using microorganisms.', *Biotechnol Bioeng* **100**(2), 203–12.

Rohwer, J. M. and Botha, F. C. (2001), 'Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data.', *Biochem. J.* **1**(358), 437–445.

Rojas-González, J. A., Soto-Súarez, M., García-Díaz, Á., Romero-Puertas, M. C., Sandalio, L. M., Mérida, Á., Thormählen, I., Geigenberger, P., Serrato, A. J. and Sahrawy, M. (2015), 'Disruption of both chloroplastic and cytosolic FBPase genes results in a dwarf phenotype and important starch and metabolite changes in Arabidopsis thaliana', *J Exp Bot* **66**(9), 2673–89.

Ruban, A. V. (2017), 'Plant science: Crops on the fast track for light', *Nature* **541**(7635), 36–37.

Saha, R., Suthers, P. F. and Maranas, C. D. (2011), 'Zea mays iRS1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism.', *PLoS One* **6**(7), e21784.

Sahrawy, M., Avila, C., Chueca, A., Canovas, F. M. and Lopez-Gorge, J. (2004), 'Increased sucrose level and altered nitrogen metabolism in Arabidopsis thaliana transgenic plants expressing antisense chloroplastic fructose-1,6-bisphosphatase.', *J Exp Bot.* **55**(408), 2495–2503.

Salisbury, F. B. and Ross, C. W. (1992), 'Plant physiology', *Plant physiology* pp. 225–248.

Saraste, M. (1999), 'Oxidative phosphorylation at the fin de siècle.', *Science* **283**(5407), 1488–93.

Sauer, U. and Eikmanns, B. J. (2005), 'The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria.', *FEMS Microbiol Rev.* **29**(4), 765–94.

Sauro, H. (1993), 'SCAMP: a general-purpose metabolic simulator and metabolic control analysis program.', *Comp. Appl. Biosci.* **9**(4), 441–450.

Scaife, M. A., Nguyen, G. T. D. T., Rico, J., Lambert, D., Helliwell, K. E. and Smith, A. G. (2015), 'Establishing Chlamydomonas reinhardtii as an industrial biotechnology host.', *Plant J* **82**(3), 532–46.

Scheibe, R. (1991), 'Redox-modulation of chloroplast enzymes : a common principle for individual control.', *Plant Physiol.* **96**(1), 1–3.

Scheibe, R. (2004), 'Malate valves to balance cellular energy supply.', *Physiol Plant* **120**(1), 21–26.

Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmanian, S., Kang, J., Hyduke, D. R. and Palsson, B. (2011), 'Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0', *Nat Protoc.* **6**, 1290–1307.

Schilling, C. H., Schuster, S., Palsson, B. O. and Heinrich, R. (1999), 'Metabolic Pathway Analysis: Basic Concepts and Scientific Applications in the Post-genomic Era', *Biotechnol Prog* **15**(3), 296–303.

Schilling, C., Letscher, D. and Palsson, B. (2000), 'Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway oriented perspective.', *J.Theor.Biol.* **203**, 229–248.

Schilling, C. and Palsson, B. (1998), 'The underlying pathway structure of biochemical reaction networks.', *PNAS* **95**, 4193–4198.

Schmidt, R., Waschina, S., Boettger-Schmidt, D., Kost, C. and Kaleta, C. (2015), 'Computing autocatalytic sets to unravel inconsistencies in metabolic network reconstructions.', *Bioinformatics* **31**(3), 373–381.

Schurmann, P. and Jacquot, J.-P. (2000), 'Plant thioredoxin system revisited.', *Annu Rev Plant Physiol Plant Mol Biol* **51**, 371–400.

Schuster, S., Dandekar, T. and Fell, D. (1999), 'Detection of Elementary Flux Modes in Biochemical Networks: a Promising Tool for Pathway Analysis and Metabolic Engineering.', *Trends. Biotech.* **17**(2), 53–60.

Schuster, S. and Fell, D. (2007), *Bioinformatics - From Genomes to Therapies*, Vol. 2, Wiley-VCH, chapter 20, pp. 755–805.

Schuster, S. and Hilgetag, C. (1994), 'On Elementary Flux Modes in Biochemical Systems at Steady State.', *J.Biol.Syst.* **2**, 165–182.

Schuster, S., Pfeiffer, T., Moldenhauer, F., Koch, I. and Dandekar, T. (2002), 'Exploring the pathway structure of metabolism: decomposition into subnetworks and application to M.pneumoniae.', *Bioinformatics* **18**(2), 351–361.

Schwacke, R., Schneider, A., van der Graaff, E., Fischer, K., Catoni, E., Desimone, M., Frommer, W. B., Flügge, U.-I. and Kunze, R. (2003), 'ARAMEMNON, a novel database for Arabidopsis integral membrane proteins.', *Plant Physiol.* **131**(1), 16–26.

Schwender, J., Goffman, F., Ohlrogge, J. B. and Shachar-Hill, Y. (2004), 'Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds.', *Nature* **432**(7018), 779–82.

Seaver, S. M. D., Gerdes, S., Frelin, O., Lerma-Ortiz, C., Bradbury, L. M. T., Zallot, R., Hasnain, G., Niehaus, T. D., Yacoubi, B. E., Pasternak, S., Olson, R., Pusch, G., Overbeek, R., Stevens, R., de Crécy-Lagard, V., Ware, D., Hanson, A. D. and Henry, C. S. (2014), 'High-throughput comparison, functional annotation, and metabolic modeling of plant genomes using the PlantSEED resource', *Proc. Natl. Acad. Sci. USA* **111**(26), 1–6.

Serrato, A. J., Barajas-Lopez, J. d. D., Chueca, A. and Sahrawy, M. (2009), 'Changing sugar partitioning in FBPase-manipulated plants', *J. Exp. Bot.* **60**(10), 2923–2931.

Serrato, A. J., Yubero-Serrano, E. M., Sandalio, L. M., Munoz-Blanco, J., Chueca, A., Cabllero, J. L. and Sahrawy, M. (2009), 'cpFBPaseII, a novel redox-independent chloroplastic isoform of fructose-1,6-bisphosphatase', *Plant Cell Environ.* **32**(7), 811–827.

Shahzad, K. and Loor, J. J. (2012), 'Application of Top-Down and Bottom-up Systems Approaches in Rumi- nant Physiology and Metabolism', *Curr Genomics.* **13**(5), 379–94.

Simkin, A. J., Lopez-Calcagno, P. E., Davey, P. A., Headland, L. R., Lawson, T., Timm, S., Bauwe, H. and Raines, C. A. (2016), 'Simultaneous stimulation of sedoheptulose 1,7-bisphosphatase, fructose 1,6-bisphophate aldolase and the photorespiratory glycine decarboxylase-H protein increases CO 2 assimilation, vegetative biomass and seed yield in Arabidopsis', *Plant Biotechnol J* pp. 1–12.

Simkin, A. J., McAusland, L., Headland, L. R., Lawson, T. and Raines, C. A. (2015), 'Multigene manipulation of photosynthetic carbon assimilation increases CO2 fixation and biomass yield in tobacco', *J Exp Bot* **66**(13), 4075–4090.

Simons, M., Saha, R., Amiour, N., Kumar, A., Guillard, L., Clément, G., Miquel, M., Li, Z., Mouille, G., Lea, P. J., Hirel, B. and Maranas, C. D. (2014), 'Assessing the metabolic impact of nitrogen availability using a compartmentalized maize leaf genome-scale model.', *Plant Physiol.* **166**(3), 1659–1674.

Smallbone, K. and Simeonidis, E. (2009), 'Flux balance analysis: A geometric perspective.', *J Theor Biol.* **258**(2), 311–315.

Soh, K. and Hatzimanikatis, V. (2010), 'DREAMS of metabolism.', *Trends Biotechnol.* **10**, 501–8.

Soulies, A., Legrand, J., Marec, H. and Pruvost, J. (2016), 'Investigation and Modeling of the Effects of Light Spectrum and Incident Angle on the Growth of Chlorella vulgaris in Photobioreactors', *Biotechnol Prog.* **32**(2), 247–61.

Sridharan, G. V., Ullah, E., Hassoun, S. and Lee, K. (2015), 'Discovery of substrate cycles in large scale metabolic networks using hierarchical modularity.', *BMC Systems Biology* **9**, 5.

Stein, R. B. and Blum, J. J. (1978), 'On the Analysis of Futile Cycles in Metabolism', *J. Theor. Biol.* **72**, 487–522.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. and Gilles, E. D. (2002), 'Metabolic network structure determines key aspects of functionality and regulation', *Nature* **420**, 190–193.

Steuer, R., Nesi, A. N., Fernie, A. R., Gross, T., Blasius, B. and Selbig, J. (2007), 'From structure to dynamics of metabolic pathways: application to the plant mitochondrial TCA cycle', *Bioinformatics* **23**(11), 1378–1385.

Stitt, M., Lilley, R. M. and Heldt, H. W. (1982), 'Adenine nucleotide levels in the cytosol, chloroplasts, and mitochondria of wheat leaf protoplasts.', *Plant Physiol* **70**(4), 971–977.

Stitt, M. and Sonnewald, U. (1995), 'Regulation of Metabolism in Transgenic Plants', *Annu Rev Plant Physiol Plant Mol Biol* **46**, 341–368.

Stitt, M., Wirtz, W., Gerhardt, R., Heldt, H. W., Spencer, C., Walker, D. and Foyer, C. (1985), 'A comparative study of metabolite levels in plant leaf material in the dark', *Planta* **166**(3), 354–364.

Sung, S.-J. S., Xu, D.-P., Galloway, C. M. and Jr., C. C. B. (1988), 'A reassessment of glycolysis and gluconeogenesis in higher plants.', *Physiol. Plant* **72**, 650–54.

Sweetlove, L. J., Nielsen, J. and Fernie, A. R. (2016), 'Engineering central metabolism - a grand challenge for plant biologists', *Plant J* .

Takache, H., Pruvost, J. and Cornet, J.-F. (2012), 'Kinetic modeling of the photosynthetic growth of Chlamydomonas reinhardtii in a photobioreactor.', *Biotechnol Prog* **28**(3), 681–92.

Tamoi, M., Nagaoka, M., Miyagawa, Y. and Shigeoka, S. (2006), 'Contribution of fructose-1,6-bisphosphatase and sedoheptulose-1,7-bi sphosphatase to the photosynthetic rate and carbon flow in the Calvin cycle in transgenic plants ', *Plant and Cell Physiology* .

Taniguchi, M. and Miyake, H. (2012), 'Redox-shuttling between chloroplast and cytosol: integration of intra-chloroplast and extra-chloroplast metabolism', *Curr. Opin. Plant Biol.* **15**(3), 252–260.

Taniguchi, M., Taniguchi, Y., Kawasaki, M., Takeda, S., Kato, T., Sato, S., Tabata, S., Miyake, H. and Sugiyama, T. (2002), 'Identifying and characterizing plastidic 2-oxoglutarate/malate and dicarboxylate transporters in Arabidopsis thaliana', *Plant Cell Physiol* **43**, 706–717.

Tardif, M., Atteia, A., Specht, M., Cogne, G., Rolland, N., Brugière, S., Hippler, M., Ferro, M., Bruley, C., Peltier, G., Vallon, O. and Cournac, L. (2012), 'PredAlgo: A New Subcellular Localization Prediction Tool Dedicated to Green Algae', *Mol Biol Evol* **29**(12), 3625–3639.

Tcherkez, G., Mahé, A., Gauthier, P., Mauve, C., Gout, E., Bligny, R., Cornic, G. and Hodges, M. (2009), 'In folio respiratory fluxomics revealed by 13C isotopic labeling and H/D isotope effects highlight the noncyclic nature of the tricarboxylic acid "cycle" in illuminated leaves.', *Plant Physiol* **151**(2), 620–630.

Terry, K. (1986), 'Photosynthesis in modulated light: quantitative dependence of photosynthetic enhancement on flashing rate.', *Biotechnol Bioeng.* **28**(7), 988–95.

Teusink, B., Wiersma, A., Molenaar, D., Francke, C., de Vos, W. M., Siezen, R. J. and Smid, E. J. (2006), 'Analysis of growth of Lactobacillus plantarum WCFS1 on a complex medium using a genome-scale metabolic model', *J Biol Chem* **281**(52), 40041–8.

Thiele, I. and Palsson, B. O. (2010), 'A protocol for generating a high-quality genome-scale metabolic reconstruction.', *Nat Protoc.* **5**(1), 93–121.

Töpfer, N., Caldana, C., Grimbs, S., Willmitzer, L., Fernie, A. R. and Nikoloski, Z. (2013), 'Integration of genome-scale modeling and transcript profiling reveals metabolic pathways underlying light and temperature acclimation in Arabidopsis.', *Plant Cell* **25**(4), 1197–1211.

Trinh, C., Unrean, P. and Srienc, F. (2008), 'Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses', *Appl Environ Microbiol.* **74**, 3634–3643.

Trinh, C., Wlaschin, A. and Srienc, F. (2009), 'Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism', *Appl Environ Microbiol.* **81**, 813–826.

van Heck, R. G. A., Ganter, M., dos Santos, V. A. P. M. and Stelling, J. (2016), 'Efficient Reconstruction of Predictive Consensus Metabolic Network Models', *PLoS Comp. Biol* **12**(8), 1–21.

Varma, A. and Palsson, B. O. (1993), 'Metabolic capabilities of Escherichia coli: I. synthesis of biosynthetic precursors and cofactors.', *J Theor Biol* **165**(4), 477–502.

Voss, I., Sunil, B., Scheibe, R. and AS., R. (2013), 'Emerging concept for the role of photorespiration as an important part of abiotic stress response.', *Plant Biol (Stuttg)* **15**(4), 713–22.

Wagner, C. (2004), 'Nullspace Approach to Determine the Elementary Modes of Chemical Reaction Systems', *J. Phys. Chem. B* **108**, 2425–2431.

Walters, R. G. (2005), 'Towards an understanding of photosynthetic acclimation.', *J. Exp. Bot.* **56**(411), 435–47.

Wang, J., Yu, Q., Xiong, H., Wang, J., Chen, S., Yang, Z. and Dai, S. (2016), 'Proteomic Insight into the Response of Arabidopsis Chloroplasts to Darkness', *PLoS One* **11**(5), 1–24.

Watson, M. R. (1986), 'A discrete model of bacterial metabolism', *Comp. Appl. Biosci.* **2**(1), 23–27.

Weber, A. P. and Linka, N. (2011), 'Connecting the Plastid: Transporters of the Plastid Envelope and Their Role in Linking Plastidial with Cytosolic Metabolism', *Annu Rev Plant Biol.* **62**, 53–77.

Weise, S. E., Weber, A. P. M. and Sharkey, T. D. (2004), 'Maltose is the major form of carbon exported from the chloroplast at night.', *Planta.* **218**(3), 474–82.

Whittaker, A. and Botha, F. C. (1997), 'Carbon Partitioning during Sucrose Accumulation in Sugarcane Internodal Tissue.', *Plant Physiol.* **115**(4), 1651–1659.

Wilhelm, C. and Selmar, D. (2011), 'Energy dissipation is an essential mechanism to sustain the viability of plants: The physiological limits of improved photosynthesis', *J Plant Physiol* **168**((2)), 79–87.

Winter, G. and Kromer, J. O. (2013), 'Fluxomics – connecting 'omics analysis and phenotypes', *Environmental Microbiology* **15**, 1901–1916.

Yeung, M., Thiele, I. and Palsson, B. O. (2007), 'Estimation of the number of extreme pathways for metabolic networks', *BMC Bioinformatics* **8**.

Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E. and Shlomi, T. (2010), 'Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model', *Bioinformatics* **26**(12), 255–260.

Yuan, H., Cheung, C. Y. M., Hilbers, P. A. J. and van, N. A. W. (2016), 'Flux Balance Analysis of Plant Metabolism: The Effect of Biomass Composition and Model Structure on Model Predictions', *Front. Plant Sci.,* **7**, 537.

# Appendix A

# Knockout investigation - experimental work

Homozygous mutant lines of each of the enzymes (SBPase, FBPase and GAPdh) were identified and knockout was performed using T-DNA insertion method. The growth of each mutants were observed and the phenotypic characters of each mutants were studied. Figure A.1 shows the comparative growth of wild type and mutant plants.

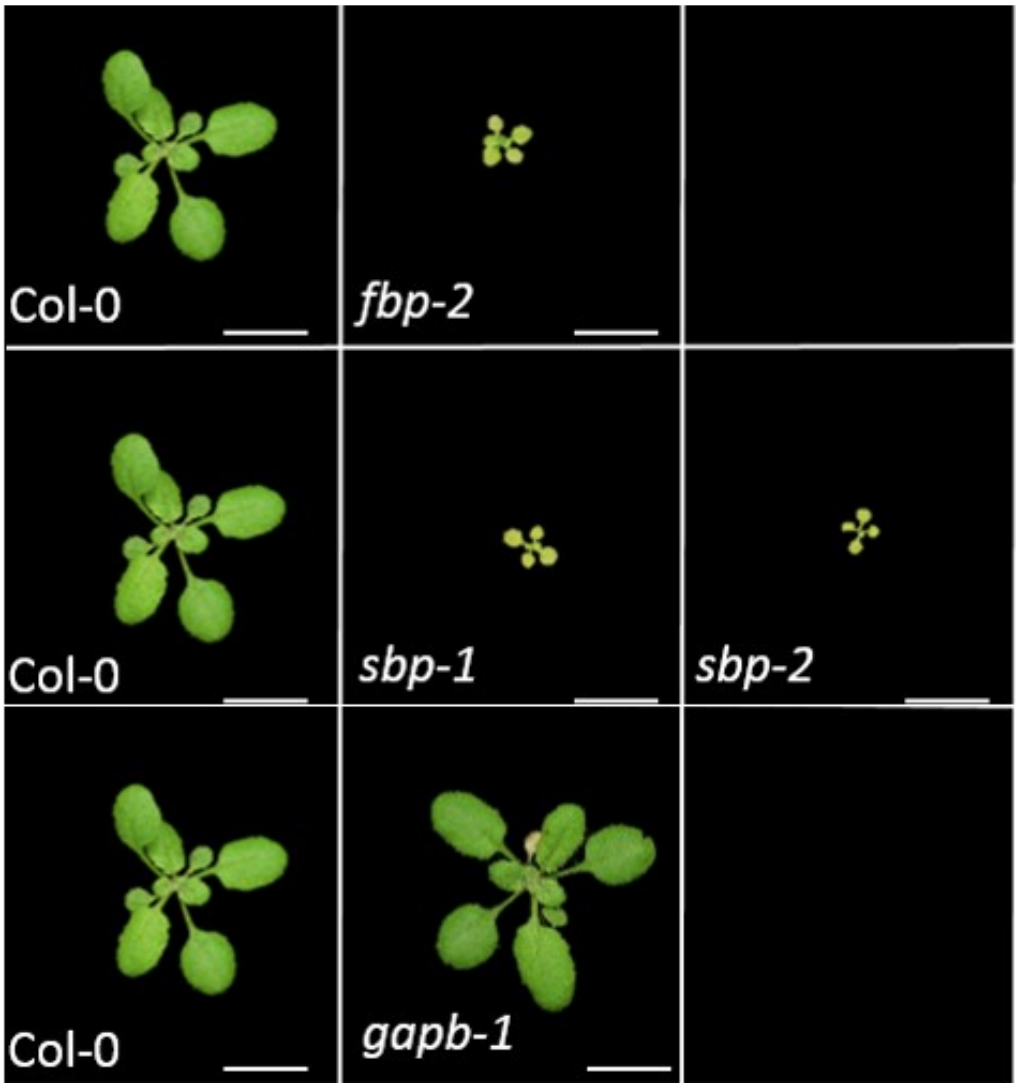



Figure A.1: Experimental results for the knock-out Arabidopsis lines. All plants are homozygous for the TDNA insertion. Each panel shows pictures of 22 day-old plants. Scale bar 1cm. Col-0 is Columbia wild types, fbp-2 is FBPase mutnant KO line 2, sbp-1 and sbp-2 are SBPase KO mutant line 1 and 2 respectively, gapb-1 is G3Pdh KO line 1. Pictures kindly provided by Martina Zanella, ETH Zurich

# Appendix B

# Energy dissipating reactions

Table B.1: List of reactions involved in energy dissipating nodes (separated by lines) identified from correlation coefficient method (see Section 6.3.1) and their functions based on BioCyc annotations. Prot Exp, states whether protein expression for the reaction increase; Sto, states if the net stoichiometry of the node is involved in hydrolysis of ATP or oxidation of NADPH.

| **Reactions** | **Function** | **Prot Exp** | **Sto** |
|---|---|---|---|
| F16BDEPHOS-RXN_Plas | Calvin cycle | Yes | ATP |
| 6PFRUCTPHOS-RXN_Plas | glycolysis | Yes | |
| GLUTATHIONE-REDUCT-NADPH-RXN_Pero | glutathione redox reactions | Yes | NADPH |
| 1.8.5.1-RXN_Pero | hydrogen peroxide detoxification | Yes | |
| L-ASCORBATE-PEROXIDASE-RXN_Pero | ascorbate metabolism | Yes | |
| SULFATE-ADENYLYLTRANS-RXN_Cyto | sulfation pathway | Yes | ATP |
| ADENYLYLSULFATASE-RXN_Cyto | sulfur metabolism | Yes | |
| PEPDEPHOS-RXN_Plas | glycolysis | Yes | ATP |
| PYRUVATEORTHOPHOSPHATE-DIKINASE-RXN_Plas | glutamine biosynthesis | Yes | |
| RXN-3522_Plas | hydrogen peroxide detoxification | Yes | ATP |
| RXN-3521_Plas | hydrogen peroxide detoxification | Yes | |
| GLUTATHIONE-REDUCT-NADPH-RXN_Cyto | glutathione redox reactions | Yes | NADPH |
| GLUTATHIONE-PEROXIDASE-RXN_Cyto | glutathione redox reactions | Yes | |
| GLUTATHIONE-REDUCT-NADPH-RXN_Mito | glutathione redox reactions | Yes | |
| 1.8.5.1-RXN_Mito | hydrogen peroxide detoxification | Yes | NADPH |
| L-ASCORBATE-PEROXIDASE-RXN_Mito | None | Yes | |
| F16BDEPHOS-RXN_Cyto | Calvin cycle | Yes | |
| 6PFRUCTPHOS-RXN_Cyto | glycolysis | Yes | ATP |
| GLUTAMIN-RXN_Cyto | glutamine amidotransferase | Yes | ATP |
| GLUTAMINESYN-RXN_Cyto | glutamate-glutamine shuttle | Yes | |
| RXN0-5462_Cyto | guanosine metabolism | Yes | ATP |
| GDPKIN-RXN_Cyto | guanosine ribonucleotides biosynthesis | No | |
| FRUCTOKINASE-RXN_Cyto | sucrose mobilization | Yes | |
| SUCROSE-PHOSPHATASE-RXN_Cyto | sucrose metabolism | No | ATP |

| | | | |
|---|---|---|---|
| SUCROSE-PHOSPHATE-SYNTHASE-RXN_Cyto | sucrose metabolism | No | |
| SUCROSE-SYNTHASE-RXN_Cyto | sucrose degradation | Yes | |
| RXN-12196_Cyto | pyrimidine biosynthesis | No | ATP |
| UDPKIN-RXN_Cyto | pyrimidine biosynthesis | Yes | |
| RXN-14217_Cyto | guanosine metabolism | No | ATP |
| DGDPKIN-RXN_Cyto | guanosine deoxyribonucleotides biosynthesis I | Yes | |
| RXN-1124_Cyto | ferulate metabolism | No | |
| RXN-10919_Cyto | ferulate metabolism | No | ATP |
| RXN-8014_Cyto | ferulate and sinapate biosynthesis | Yes | |
| SULFATE-ADENYLYLTRANS-RXN_Cyto | sulfation pathway | Yes | ATP |
| ADENYLYLSULFATASE-RXN_Cyto | sulfur metabolism | No | |
| RXN-12195_Cyto | pyrimidine biosynthesis | No | |
| CDPKIN-RXN_Cyto | pyrimidine biosynthesis | Yes | ATP |
| RXN-10811_Cyto | CoA metabolism | No | ATP |
| DEPHOSPHOCOAKIN-RXN_Cyto | CoA biosynthesis | No | |
| 325-BISPHOSPHATE-NUCLEOTIDASE-RXN_Cyto | CoA metabolism | Yes | |
| PANTEPADENYLYLTRAN-RXN_Cyto | CoA biosynthesis | No | |
| AMP-DEPHOSPHORYLATION-RXN_Cyto | nucleotide metabolism | No | ATP |
| ADENOSINE-KINASE-RXN_Cyto | adenine and adenosine salvage II | Yes | |
| CARBAMATE-KINASE-RXN_Plas | ornithine-citrulline shuttle | No | ATP |
| 6.3.4.16-RXN_Plas | Krebs ornithine cycle | Yes | |
| DUDPKIN-RXN_Cyto | nucleotide metabolism | Yes | ATP |
| RXN-14219_Cyto | nucleotide metabolism | No | |
| DADPKIN-RXN_Cyto | purine deoxyribonucleosides salvage | Yes | ATP |
| RXN-14214_Cyto | RXN-14195 | No | |
| DCDPKIN-RXN_Cyto | pyrimidine deoxyribonucleotide phosphorylation | Yes | ATP |
| RXN-14216_Cyto | RXN-14198 | No | |
| 2.7.7.1-RXN_Cyto | None | No | ATP |
| NADPYROPHOSPHAT-RXN_Cyto | Preiss-Handler salvage pathway | Yes | |
| GLUCOKIN-RXN_Cyto | hexokinase | Yes | |
| THYMIDINE-TRIPHOSPHATASE-RXN_Cyto | thymine metabolism | No | ATP |
| DTDPKIN-RXN_Cyto | pyrimidine deoxyribonucleotide phosphorylation | Yes | |
| RXN0-5462_Cyto | nucleotide metabolism | No | |
| GDPKIN-RXN_Cyto | guanosine ribonucleotides biosynthesis | Yes | ATP |
| RXN-10770_Plas | starch metabolism | No | |
| GLUC1PADENYLTRANS-RXN_Plas | starch biosynthesis | Yes | ATP |
| MYO-INOSITOL-1OR-4-MONOPHOSPHATASE | myo-inositol biosynthesis | Yes | ATP |
| MYO-INOSITOL-1-KINASE-RXN_Cyto | lipid-independent phytic acid biosynthesis | No | |
| PHOSPHOENOLPYRUVATE-PHOSPHATASE-RXN_Cyto | pyruvate metabolism | No | ATP |
| PYRUVATEORTHOPHOSPHATE-DIKINASE-RXN_Cyto | glutamine biosynthesis | Yes | |

Table B.2: A comprehensive list of all reactions involved in energy dissipating modes identified from MILP method (see section 6.3.2), their function based on BioCyc annotations and, where available, respective protein counts under high and low light conditions. Md. Inv., No. of modes the reaction is involved in; ProtLo, Protein counts under low light conditions; ProtHi, Protein counts under high light conditions.

| **Reaction** | **Md. Inv.** | **Function** | **ProtLo** | **ProtHi** |
|---|---|---|---|---|
| **Plastid** | | | | |
| LightCyc | 168 | Cyclic phosphorylation | | |
| ADENYL-KIN-RXN | 166 | PWY-7219 | 12 | 63.84 |
| INORGPYROPHOSPHAT-RXN | 158 | | 28.8 | 50.16 |
| PYRUVATEORTHOPHOSPHATE-DIKINASE-RXN | 141 | glutamine biosynthesis | 9.12 | 18.24 |
| LightNonCyc | 77 | Non cyclic phosphorylation | | |
| 1.2.1.13-RXN | 60 | Calvin cycle | 319.2 | 453.72 |
| PHOSGLYPHOS-RXN | 54 | glycolysis | 242.4 | 403.56 |
| GLUTATHIONE-REDUCT-NADPH-RXN | 25 | GLUT-REDOX-PWY, PWY-4081 | 13.68 | 31.92 |
| RXN-13185 | 25 | Xanthophyll cycle | 7.2 | 9.12 |
| 1.8.5.1-RXN | 25 | hydrogen peroxide detoxification | 108 | 148.2 |
| GLUC1PADENYLTRANS-RXN | 23 | PWY-622 | 28.8 | 191.52 |
| CARBAMATE-KINASE-RXN | 18 | PWY-7060, CITRULLINE-DEG-PWY | | |
| F16BDEPHOS-RXN | 18 | Calvin cycle | 124.8 | 310.08 |
| GLUTAMINESYN-RXN | 17 | glutamine - glutamate pathway | 244.8 | 300.96 |
| RXN-13193-(NADP) | 17 | | 4.56 | 4.56 |
| 6PFRUCTPHOS-RXN | 16 | glycolysis | 6.84 | 6.84 |
| RXN-10770 | 14 | | | |
| 6.3.4.16-RXN | 14 | Krebs ornithine cycle | 15.96 | 15.96 |
| GLY3KIN-RXN | 13 | photorespiration | | |
| GLYCOGENSYN-RXN | 9 | PWY-622 | 6.27 | 25.08 |
| GLUCOKIN-RXN | 9 | GLUCOSE1PMETAB-PWY, PWY-7238, PWY0-1182 | 4.56 | 4.56 |
| RXN-13193-(NAD) | 8 | | 4.56 | 4.56 |
| GAPOXNPHOSPHN-RXN | 6 | glycolysis | 319.2 | 590.52 |
| PEPDEPHOS-RXN | 5 | glycolysis | 16.72 | 91.2 |
| RXN0-5184 | 5 | | 38.76 | 93.48 |
| CYSTATHIONINE-BETA-SYNTHASE-RXN | 4 | transsulfuration pathway | | |
| PHOSPHOGLUCMUT-RXN | 4 | GLUCOSE1PMETAB-PWY, PWY-3801, PWY-622, PWY-7238, PWY-7343, PWYQT-4466 | 47.88 | 47.88 |
| PHOSPHORIBULOKINASE-RXN | 4 | Calvin cycle | 213.6 | 357.96 |
| RXN0-5224 | 4 | cyanate catabolism | 698.4 | 1019.16 |
| CYSTATHIONINE-BETA-LYASE-RXN | 4 | | 45.6 | 79.8 |
| CARBPSYN-RXN | 4 | PWY-7060, PWY-5686, ARGSYN-PWY, ARGSYNBSUB-PWY | 15.96 | 15.96 |
| RXN0-5114 | 3 | SERSYN-PWY | | |
| PSERTRANSAM-RXN | 3 | SERSYN-PWY | | |
| GLUTAMATE-SYNTHASE-NADH-RXN | 3 | glutamate biosynthesis from glutamine | | |
| PGLYCDEHYDROG-RXN | 3 | SERSYN-PWY | 21.6 | 114 |
| RXN-2141 | 2 | | | |
| MALTODEXGLUCOSID-RXN | 2 | | | |
| TYRAMINOTRANS-RXN | 2 | | | |
| PGLUCISOM-RXN | 2 | glycolysis | 3.42 | 6.84 |
| RIBULOSE-BISPHOSPHATE-CARBOXYLASE-RXN | 2 | Calvin cycle | 2846.4 | 7462.44 |

| | | | | |
|---|---|---|---|---|
| TRIOSEPISOMERIZATION-RXN | 2 | glycolysis | 112.8 | 278.16 |
| PREPHENATEDEHYDROG-RXN | 2 | | | |
| RXN-961 | 2 | photorespiration | 2846.4 | 7462.44 |
| GPH-RXN | 2 | photorespiration | 4.8 | 9.12 |
| PREPHENATE-TRANSAMINE-RXN | 2 | PWY-3461, PWY-3462 | | |
| F16ALDOLASE-RXN | 2 | glycolysis | 530.4 | 1035.12 |
| 6PGLUCONOLACT-RXN | 2 | OXIDATIVEPENT-PWY | | |
| GLU6PDEHYDROG-RXN | 2 | OXIDATIVEPENT-PWY | 2.28 | 2.28 |
| RXN-1827 | 2 | | 9.12 | 27.36 |
| RXN-5682 | 2 | pretyrosine pathway | | |
| 6PGLUCONDEHYDROG-RXN | 2 | OXIDATIVEPENT-PWY | | |
| 3PGAREARR-RXN | 1 | glycolysis | 16.8 | 95.76 |
| 2PGADEHYDRAT-RXN | 1 | glycolysis | 45.6 | 189.24 |

**Cyto**

| | | | | |
|---|---|---|---|---|
| PEPDEPHOS-RXN | 130 | glycolysis | 16.72 | 91.2 |
| 1.2.1.9-RXN | 54 | glycolysis | 14.4 | 104.88 |
| INORGPYROPHOSPHAT-RXN | 53 | | 28.8 | 50.16 |
| ADENYL-KIN-RXN | 44 | PWY-7219 | 12 | 63.84 |
| UMPKI-RXN | 33 | | | |
| RXN-10745-(NADP) | 31 | | | |
| RXN-7381 | 28 | | | |
| RXN-12199 | 28 | pyrimidine biosynthesis | | |
| RXN-7401 | 28 | UDP-L-arabinose biosynthesis | | 2.28 |
| RXN-3522 | 26 | hydrogen peroxide detoxification | 21.6 | 184.68 |
| GLUTATHIONE-REDUCT-NADPH-RXN | 17 | GLUT-REDOX-PWY, PWY-4081 | 13.68 | 31.92 |
| GLUTATHIONE-PEROXIDASE-RXN | 15 | DETOX1-PWY-1, PWY-4081 | 148.8 | 216.6 |
| RXN-3521 | 15 | hydrogen peroxide detoxification | 134.4 | 371.64 |
| 3-PHOSPHOGLYCERATE-PHOSPHATASE-RXN | 13 | | | |
| ASPARAGHYD-RXN | 12 | asparagine degradation 1 | | |
| PRPPSYN-RXN | 11 | 5-phosphoribosyl 1-pyrophosphate biosynthesis | | |
| ADENOSINE-KINASE-RXN | 11 | PWY-6619 | 33.6 | 75.24 |
| RXN-8444 | 11 | Preiss-Handler salvage pathway | | |
| F16BDEPHOS-RXN | 11 | Calvin cycle | 124.8 | 310.08 |
| NICOTINATEPRIBOSYLTRANS-RXN | 11 | Preiss-Handler salvage pathway | | |
| CATAL-RXN | 11 | removal of superoxide radicals | 208.8 | 380.76 |
| RXN-3541 | 11 | | 2.4 | 2.4 |
| ADENOSYLHOMOCYSTEINASE-RXN | 10 | activated methyl cycle, SAM cycle | 93.6 | 373.92 |
| S-ADENMETSYN-RXN | 10 | S-adenosylmethionine biosynthesis | 33.6 | 72.96 |
| 2.7.1.90-RXN | 10 | glycolysis | | 6.84 |
| CDPKIN-RXN | 8 | pyrimidine biosynthesis | 74.4 | 171 |
| ATP-PYROPHOSPHATASE-RXN | 8 | | | |
| ATPase | 8 | | | |
| PANTEPADENYLYLTRAN-RXN | 7 | CoA biosynthesis | | |
| RXN-12200 | 7 | pyrimidine biosynthesis | | |
| 325-BISPHOSPHATE-NUCLEOTIDASE-RXN | 7 | | 7.2 | 18.24 |
| HOMOCYSTEINE-S-METHYLTRANSFERASE-RXN | 7 | methionine biosynthesis from homoserine II | | |
| ASNSYNB-RXN | 7 | ASPARAGINE-BIOSYNTHESIS | 2.28 | 2.28 |
| DEPHOSPHOCOAKIN-RXN | 7 | CoA biosynthesis | | |
| RXN-10811 | 7 | | | |
| GLUTAMIN-RXN | 6 | ornithine and citrulline biosynthesis | 7.41 | 31.92 |
| GMKALT-RXN | 6 | PWY-7224 | | |
| UDPKIN-RXN | 6 | pyrimidine biosynthesis | 74.4 | 171 |
| GLUTAMINESYN-RXN | 6 | glutamine - glutamate pathway | 244.8 | 300.96 |
| RXN-11832 | 6 | pyrimidine ribonucleotide/ribonucleoside metabolism | | |
| RXN-11811 | 6 | | | |
| GLUCOSE-6-PHOSPHATASE-RXN | 6 | | | |
| CMPKI-RXN | 5 | | | |
| GUANYL-KIN-RXN | 5 | PWY-7221 | | |

| | | | | |
|---|---|---|---|---|
| SULFATE-ADENYLYLTRANS-RXN | 5 | sulfation pathway | 16.8 | 41.04 |
| ADENYLYLSULFATASE-RXN | 5 | | | 2.28 |
| ASNSYNA-RXN | 5 | Asparagine synthetase (glutamine-hydrolysing) | | |
| RXN-14026 | 4 | pyrimidine nucleotides dephosphorylation | | |
| GDPKIN-RXN | 4 | PWY-7221 | 26.4 | 102.6 |
| RXN-8092 | 4 | | | |
| RXN-9276-(NADP) | 4 | | | |
| RXN0-3962 | 4 | | | |
| RXN-15514-(NADP) | 4 | | | |
| DTDPKIN-RXN | 4 | PWY0-166, PWY-7187, PWY-6545, PWY-7197, PWY-7184 | 74.4 | 171 |
| DTMPKI-RXN | 4 | PWY0-166, PWY-7187, PWY-6545, PWY-7197, PWY-7184 | | |
| RXN-7913 | 4 | PWY-7197 | | |
| PHOSPHOENOLPYRUVATE-PHOSPHATASE-RXN | 4 | | | |
| DGDPKIN-RXN | 4 | PWY-7224, PWY-7226 | 74.4 | 171 |
| RXN-9275 | 4 | PWY-5287 | | |
| DCDPKIN-RXN | 4 | PWY0-166, PWY-6545, PWY-7197, PWY-7184 | 74.4 | 171 |
| RXN0-383 | 4 | | | |
| DEOXYADENYLATE-KINASE-RXN | 4 | PWY-7224 | | |
| DADPKIN-RXN | 4 | PWY-7227, PWY-7224 | 74.4 | 171 |
| 1.5.3.12-RXN | 4 | PWY-5287 | | |
| CTPSYN-RXN | 3 | pyrimidine biosynthesis | | |
| RXN-2005 | 3 | benzoic acid biosynthesis | | |
| RXN-9666 | 3 | | | |
| RXN-2945 | 3 | IAA-amide conjugate biosynthesis | | |
| RXN-12753 | 3 | PWY-6938 | | |
| FAD-PYROPHOSPHATASE-RXN | 3 | | | |
| RXN-2981 | 3 | IAA biosynthesis from amide-conjugates | | 2.28 |
| FADSYN-RXN | 3 | vitamin B2 biosynthesis | | |
| RXN-2944 | 3 | IAA-amide conjugate biosynthesis | | |
| 2.7.7.1-RXN | 3 | | | |
| RXN-9623 | 3 | | | |
| RXN-12198 | 3 | pyrimidine biosynthesis | | |
| RXN-2982 | 3 | IAA biosynthesis from amide-conjugates | | |
| RXN-14218 | 3 | RXN-14208 | | |
| BENZOATE–COA-LIGASE-RXN | 3 | benzoate oxidation | | |
| RXN0-5462 | 3 | RXN-14201 | | 2.28 |
| 4.2.1.93-RXN | 3 | PWY-6938 | | |
| RXN-9644 | 3 | PWY-5995 | | |
| NADPYROPHOSPHAT-RXN | 3 | Preiss-Handler salvage pathway | 8.305714286 | 38.76 |
| PALMITOYL-COA-HYDROLASE-RXN | 3 | | | |
| URIDINEKIN-RXN | 3 | PWY-7193 | | |
| CYTIDEAM2-RXN | 3 | PWY-7193, PWY0-163, PWY-6556 | | |
| MMUM-RXN | 3 | SMM cycle | | |
| METHIONINE-S-METHYLTRANSFERASE-RXN | 3 | SMM cycle | 4.56 | 9.12 |
| RXN0-384 | 2 | | | |
| RXN-12304 | 2 | | | |
| RXN-14201 | 2 | | | |
| RXN0-1441 | 2 | | 8.305714286 | 38.76 |
| RXNQT-4150 | 2 | PWYQT-4445 | | |
| RXN-14140 | 2 | | | |
| L-ASCORBATE-PEROXIDASE-RXN | 2 | | 134.4 | 371.64 |
| URKI-RXN | 2 | PWY0-163, PWYQT-4445 | | |
| RXN-12196 | 2 | pyrimidine biosynthesis | | |
| 2PGADEHYDRAT-RXN | 2 | glycolysis | 45.6 | 189.24 |
| 2.7.7.35-RXN | 2 | | | 2.28 |
| DCTP-PYROPHOSPHATASE-RXN | 2 | PWY-7206 | | |
| RXN-11625 | 2 | farnesylcysteine detoxification pathway | | |

| | | | | |
|---|---|---|---|---|
| RXN0-5107 | 2 | | | |
| RXN0-385 | 2 | | | |
| RXN-12195 | 2 | pyrimidine biosynthesis | | |
| 1.8.5.1-RXN | 2 | hydrogen peroxide detoxification | 108 | 148.2 |
| RXN-14215 | 1 | RXN-14195 | | |
| RXN-14208 | 1 | | | |
| RXN-14195 | 1 | | | |
| ETHANOLAMINE-KINASE-RXN | 1 | PWY4FS-6, PWY-3385 | | |
| RXN-14214 | 1 | RXN-14195 | | |
| CHOLINE-KINASE-RXN | 1 | CDP-choline pathway | | |
| GUANOSINEKIN-RXN | 1 | PWY-6618 | | |
| RXN-14213 | 1 | RXN-14200 | | |
| PYRIDOXKIN-RXN | 1 | vitamin B6 salvage | | |
| OHMETPYRKIN-RXN | 1 | HMP salvage | | |
| PYRIMSYN3-RXN | 1 | HMP-PP biosynthesis | | |
| 2.7.1.139-RXN | 1 | lipid-dependent phytic acid biosynthesis | | |
| RXN-15513 | 1 | glycolysis | 16.8 | 95.76 |
| DUDPKIN-RXN | 1 | PWY0-166 | 74.4 | 171 |
| 3.1.3.74-RXN | 1 | vitamin B6 salvage (plants) | | |
| MYO-INOSITOL-1-KINASE-RXN | 1 | lipid-independent phytic acid biosynthesis | | |
| 3PGAREARR-RXN | 1 | glycolysis | 16.8 | 95.76 |
| GALACTURONOKINASE-RXN | 1 | UDP-D-galacturonate biosynthesis (salvage pathway) | | |
| GLUCOKIN-RXN | 1 | GLUCOSE1PMETAB-PWY, PWY-7238, PWY0-1182 | 4.56 | 4.56 |
| RXN-11150 | 1 | vitamin c biosynthesis | | |
| RXNQT-4191 | 1 | vitamin B1 biosynthesis IV | | |
| RXN-14181 | 1 | vitamin B6 salvage (plants) | | |
| RXN-14046 | 1 | vitamin B6 salvage (plants) | | |
| RXN-14219 | 1 | RXN-14199 | | |
| AMP-DEPHOSPHORYLATION-RXN | 1 | | | |
| PYRAMKIN-RXN | 1 | vitamin B6 salvage | | |
| RXN-14217 | 1 | RXN-14208 | | |
| RXN-7609 | 1 | PWY-6607, PWY-6606 | | |
| GUANOSINE-DIPHOSPHATASE-RXN | 1 | RXN-14201 | | |
| RXN-14200 | 1 | | | |
| RXN-14187 | 1 | RXN-14198 | | |
| 4.1.1.32-RXN | 1 | PWY66-399 | | |
| RXN-15581 | 1 | D-serine hydro-lyase | | |
| 3.1.3.62-RXN | 1 | PWY-6364 | | |
| RXN-8899 | 1 | | | |
| RXN-12197 | 1 | pyrimidine biosynthesis | | |
| THYMIDINE-TRIPHOSPHATASE-RXN | 1 | RXN-14200 | | |
| RXN-14216 | 1 | RXN-14198 | | |
| CYTIKIN-RXN | 1 | PWY-7193 | | |
| PNKIN-RXN | 1 | vitamin B6 salvage | | |
| GLUCONOKIN-RXN | 1 | PWY-5530 | | |
| 2.7.1.127-RXN | 1 | PWY-6364 | | |
| THIAZOLSYN3-RXN | 1 | PWY-7353, | | |
| RXN0-5186 | 1 | | | |
| RXN-7948 | 1 | | | |
| MYO-INOSITOL-1OR-4-MONOPHOSPHATASE-RXN | 1 | PWY-2301 | 2.4 | 31.92 |
| PEPCARBOXYKIN-RXN | 1 | GLUCONEO-PWY | | |
| THIAMINASE-RXN | 1 | PWY-7356 | | |
| RXN-5647 | 1 | PWY-3385 | | |
| KETOHEXOKINASE-RXN | 1 | | | |
| 6PFRUCTPHOS-RXN | 1 | glycolysis | 6.84 | 6.84 |
| RXN-14139 | 1 | | | |
| RXN-8730 | 1 | PWY-6364 | | |
| THI-P-SYN-RXN | 1 | thiamine diphosphate biosynthesis | | |
| RXN0-5185 | 1 | | | |
| RXN-14198 | 1 | | | |
| PYRUVATEORTHOPHOSPHATE-DIKINASE-RXN | 1 | glutamine biosynthesis | 9.12 | 18.24 |

| 5.1.1.18-RXN | 1 | D serine biosynthesis |
|---|---|---|

| Plas_tx | |
|---|---|
| Photon_tx | 231 |
| PEP_Pi_tx | 135 |
| PYRUVATE_tx | 132 |
| GAP_3PGA_tx | 54 |
| O2_tx | 54 |
| Pi_H+_tx | 45 |
| ATP_tx | 42 |
| AMP_tx | 41 |
| ADP_tx | 15 |
| GLYCERATE_tx | 13 |
| GLN_GLT_tx | 12 |
| NH3_tx | 11 |
| GAP_Pi_tx | 8 |
| G6P_Pi_tx | 5 |
| 3PGA_Pi_tx | 5 |
| GLC_tx | 5 |
| SER_tx | 1 |

# Appendix C

# Model files and Python code

## C.1 Contents of the attached CD.

Attached CD contains following directories:

### Models

GSMs of Arabidopsis and Chlamydomonas in ScrumPy format.

### Curation files

Corrected reactions, list of reactions and metabolites substituted and removed; Corrections.spy, Substitutions.spy, Unwanted.py, respectively. Separate directories for both models.

### Knockout Investigation

KO_Data.data, with data set of all knockout investigation results and KO_Core.py, python code used for analysis

### Photon Absorbing Core models

Photon absorbing core models of Arabidopsis and Chlamydomonas

### Energy Dissipation modes-analysis

Correlation Analysis, MILP - Python codes used for correlation coefficient and MILP methods respectively.

### Biomass Data

Spreadsheet with experimentally generated biomass data for Chlamydomonas