

Genome scale metabolic models of plant tissues

C. Y. Maurice Cheung

Trinity term 2013

D.Phil. Thesis

University of Oxford

Department of Plant Sciences

Systems Biology Doctoral Training Centre

Keble College

Supervisors: Dr. L. J. Sweetlove, Prof. R. G. Ratcliffe and Dr. M. G. Poolman

Abstract

Genome scale metabolic models of plant tissues

C. Y. Maurice Cheung, Keble College

Submitted for the degree of D.Phil., Trinity term 2013

The aim of this thesis was to explore the use of genome-scale metabolic models to predict metabolic fluxes in plant tissues. Results from this thesis showed that the application of constraint-based modelling, namely flux balance analysis, to an Arabidopsis genome-scale metabolic model gave accurate predictions of metabolic fluxes in heterotrophic cell culture and in photosynthetic leaves. Two major factors important for the accuracy of model predictions were highlighted from the study: 1) the inclusion of energetic costs for transports and cellular maintenance in terms of ATP and NADPH; 2) consideration of the interactions between light and dark metabolism in modelling photosynthetic leaves.

This study began with the construction of a well-curated and compartmented genome-scale metabolic model of Arabidopsis. Using the model, cellular maintenance costs in a heterotrophic cell culture under control and two stress conditions were estimated in terms of ATP and reductant usage. The results suggested that the cells were not stressed under hyperosmotic conditions. Comparisons between model predictions and experimentally estimated flux maps showed that the inclusion of transport and maintenance costs was important for obtaining accurate model flux predictions.

To model leaf metabolism over a day-night cycle, a diel modelling framework was developed which took into account the interactions between light and dark metabolism. Numerous known features of metabolism in a C_3 leaf were predicted such as the nocturnal accumulation of citrate utilised for diurnal glutamate and glutamine synthesis and the operation of an incomplete TCA cycle during the day. Using the Arabidopsis genome-scale metabolic model and the diel modelling framework, the operation of the CAM cycle was predicted as a direct consequence of blocking the CO_2 exchange with the external air during the day to simulate closure of the stomata. Comparisons between model predictions of C_3 and various subtypes of CAM leaves suggested that photon and nitrogen use efficiencies are unlikely to be the driving forces for the evolution of CAM plants under the current atmospheric CO_2 concentration.

Finally, the model was utilised to predict the changes in metabolic fluxes, in particular fluxes through various routes of alternative electron flow, in a C_3 leaf with varying light intensity, nitrogen availability and at different stages of leaf development. From the model flux predictions, it was shown that constraint-based modelling can be utilised to elucidate the distinct metabolic roles of enzymes in different subcellular compartments and the tissue-specific use of distinct forms of enzymes with different coenzyme specificities.

Acknowledgements

I would like to thank my supervisors Dr. Lee Sweetlove and Prof. R. George Ratcliffe for their advice and encouragement throughout the course of my DPhil project, and Prof. David Fell and Dr. Mark Poolman for their guidance on metabolic modelling and technical help on modelling in ScrumPy.

Many thanks to everyone at the Department of Plant Sciences, in particular Prof. Andrew Smith for his advice on CAM, Dr. Nick Kruger and Dr. Mark Fricker for their comments on my work at various stages of my DPhil project, Aliah Hawari, Clare Woods and Merja Rossi for discussions on flux analysis and related applications, and Tom Williams and Shyam Masakapalli, former members of the Department, for sharing their experimental data and discussions on flux analysis.

I would also like to thank everyone, including visitors, in the Cell Systems Modelling Group, especially Hassan Hartman, Thomas Pfau and Sudip Kundu for discussions on genome-scale metabolic modelling, various modelling techniques and ScrumPy, and the Systems Biology Doctoral Training Centre, particularly Dr. Gail Preston, for providing training in research skills.

I am also grateful for the financial support provided by the Clarendon Fund and Sloane Robinson Scholarship.

Last but not least, I would like to offer my deepest gratitude to my family for their constant support and encouragement over the years, and to Louise Tan for her love, understanding and encouragement in the past few years.

Abbreviations

3PGA	3-phosphoglycerate
AEF	alternative electron flow
ATP	adenosine 5'-triphosphate
CAM	crassulacean acid metabolism
CoA	coenzyme A
DNA	deoxyribonucleic acid
DW	dry weight
ETC	electron transport chain
FBA	flux balance analysis
Fd	ferredoxin
FVA	flux variability analysis
G6P	glucose 6-phosphate
GAP	glyceraldehyde 3-phosphate
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GOGAT	glutamine oxoglutarate aminotransferase / glutamate synthase
GS	glutamine synthetase
GTP	guanosine 5'-triphosphate
KEGG	Kyoto encyclopedia of genes and genomes
MDH	malate dehydrogenase
ME	malic enzyme
MFA	metabolic flux analysis
MS	mass spectrometry
NAD ⁺	nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NADP ⁺	nicotinamide adenine dinucleotide phosphate
NADPH	reduced nicotinamide adenine dinucleotide phosphate

NMR	nuclear magnetic resonance
OAA	oxaloacetate
OPPP	oxidative pentose phosphate pathway
PEP	phosphoenolpyruvate
RAD	relative average deviation
RNA	ribonucleic acid
RSD	relative standard deviation
Rubisco	ribulose 1,5-bisphosphate carboxylase oxygenase
RuBP	ribulose 1,5-bisphosphate
TCA	tricarboxylic acid

Contents

1	General Introduction	10
1.1	Statement of aims	10
1.2	Plant metabolic network	10
1.3	Quantifying metabolic fluxes.....	13
1.4	Metabolic modelling	15
1.5	Objective functions in flux balance analysis.....	18
1.6	Multiple optimal flux solutions.....	22
1.7	Genome-scale metabolic models.....	24
1.8	Modelling metabolism of specific tissues and cell types	27
1.9	Summary	30
2	Modelling methods.....	31
2.1	Modelling software package – ScrumPy	31
2.2	Construction of genome-scale metabolic model using ScrumPy	31
2.2.1	Extracting reactions from AraCyc	32
2.2.2	Incorporating changes into the raw model	32
2.2.3	Defining the direction and reversibility of reactions	33
2.2.4	Compartmentation of the model	33
2.2.5	Model formats	34
2.3	Construction of a diel metabolic model	35
2.4	Identification of stoichiometric inconsistencies	36
2.5	Performing flux balance analysis in ScrumPy	36
2.6	Flux minimisation with random weighting factors	39
2.6.1	Variability of fluxes in multiple solutions from flux minimisation with random weighting factors	41
2.6.2	Quantifying the relatedness between reactions by the Pearson correlation coefficient	41
2.6.3	Calculating flux solutions with varying constraints using flux minimisation with random weighting factors	42
2.7	Pareto optimisation	43
2.8	Flux variability analysis	44
2.9	Comparison between FBA flux solution and ¹³ C-MFA flux map	45

2.10	Quantitative comparison between flux ranges calculated from the diel model and the single steady-state model.....	46
2.11	Calculations of ATP and NADPH requirements in leaves during the day from flux solutions	47
3	Reconstruction of a high-quality Arabidopsis genome-scale metabolic model	50
3.1	Introduction	50
3.1.1	Structural metabolic modelling	50
3.1.2	Reconstruction of genome-scale metabolic models.....	52
3.1.3	Compartmentation in models of eukaryotic organisms.....	54
3.2	Results	56
3.2.1	Construction of a draft genome-scale metabolic model from AraCyc database.....	56
3.2.1.1	Extracting the list of metabolic reactions from AraCyc database into ScrumPy format.....	56
3.2.1.2	Resolving problems with ill-defined metabolites in the AraCyc database.....	57
3.2.1.3	Removal of reactions and corrections of reaction stoichiometry	59
3.2.2	Conversion of a draft model into a functional metabolic model	61
3.2.2.1	Definition of inputs and outputs reactions.....	61
3.2.2.2	Definition of the internal and external metabolites	61
3.2.2.3	Model testing for metabolic functions.....	63
3.2.3	Compartmentation of the Arabidopsis genome-scale metabolic model	64
3.2.3.1	Assigning reactions to different subcellular compartments	64
3.2.3.2	Addition of metabolite transport reactions between subcellular compartments	64
3.2.4	Definition of proton gradients across membranes, metabolite transport costs, and energetic costs for other cellular processes in the model	65
3.2.4.1	Definition of proton gradients across membranes	65
3.2.4.2	Definition of metabolite transport costs.....	67
3.2.4.3	Energetic costs for other cellular processes.....	67

3.2.5	Identification of errors and inconsistencies in the structure of the model.....	69
3.2.5.1	Testing for stoichiometric inconsistency in the model	69
3.2.5.2	Identification and removal of internal reaction cycles producing ATP and NAD(P)H	69
3.2.5.3	Removal of transhydrogenase cycles transferring reducing power from NADH to NADP ⁺	71
3.2.6	Model statistics	72
3.3	Discussion	73
3.3.1	Improvements and limitations of the updated Arabidopsis genome-scale metabolic model	73
3.3.2	Compartmentation in plants genome-scale metabolic models	75
3.3.3	Energetic costs for transport and other cellular processes in plant flux-balance models	77
3.3.4	Summary	79
4	Predictive accuracy of a flux balance model of heterotrophic Arabidopsis metabolism	83
4.1	Introduction	83
4.1.1	Using flux balance analysis to predict metabolic fluxes.....	83
4.1.2	Comparisons between FBA predictions and experimentally measured fluxes in plant models	84
4.1.3	Cellular maintenance costs and other energetic costs in flux balance models	85
4.2	Results	88
4.2.1	The effect of reversibility of NADP-dependent dehydrogenase on flux predictions	88
4.2.2	Estimating maintenance costs	89
4.2.3	The effect of transport and maintenance costs on flux predictions	91
4.2.4	The effect of objective functions on flux predictions	92
4.2.5	Predicting fluxes and estimating maintenance costs in heterotrophic cell culture under stress conditions	94
4.3	Discussion	95
4.3.1	Accounting for energy expenditure in a flux balance model	95

4.3.2	The impact of different constraints and objective functions on flux predictions	99
4.3.3	Impact of hyperosmotic stress and elevated temperature on maintenance costs	101
4.3.4	Summary	102
5	Modelling metabolism of C ₃ and CAM leaves over a day-night cycle	113
5.1	Introduction	113
5.1.1	The use of flux balance analysis in modelling photosynthetic metabolism	113
5.1.2	C ₃ and Crassulacean acid metabolism (CAM) photosynthesis	114
5.2	Results	117
5.2.1	Diel modelling framework for modelling leaf metabolism	117
5.2.2	Comparison of predictions from diel modelling framework and single steady-state model	120
5.2.3	Predictions of metabolic fluxes in a mature C ₃ leaf over a day-night cycle	122
5.2.4	Predictions of metabolic fluxes in a mature CAM leaf over a day-night cycle	125
5.2.5	Quantitative comparisons between C ₃ and various subtypes of CAM	130
5.3	Discussion	131
5.3.1	Applications of diel metabolic modelling framework in FBA studies	131
5.3.2	Relating model flux prediction of a C ₃ leaf with known biochemistry	134
5.3.3	Predicting the metabolism in a CAM leaf using the diel modelling framework	136
5.3.4	Comparison of model predictions of C ₃ and CAM leaves	140
5.3.5	Summary	143
6	Modelling photosynthetic metabolism in response to varying environmental and physiological conditions	155
6.1	Introduction	155

6.1.1	Importance of alternative electron flow in photosynthetic metabolism	155
6.1.2	Modelling photosynthetic metabolism under different conditions	157
6.1.3	Physiological constraints in FBA studies of photosynthetic metabolism.....	158
6.2	Results	160
6.2.1	The use of random weighting factors in minimisation of total flux as an objective function	160
6.2.2	Changes in leaf metabolism under varying light intensity.....	162
6.2.3	Changes in leaf metabolism under varying nitrogen availability	165
6.2.4	The effect of growth on photosynthetic metabolism	167
6.2.5	Exploring the flexibility of leaf metabolism	170
6.3	Discussion	173
6.3.1	The use of flux minimisation with random weighting factors in analysing FBA models	173
6.3.2	Flux predictions of alternative electron flow and energy-dissipating reactions	175
6.3.3	Model predictions suggest distinct functions for enzymes in ammonia assimilation	178
6.3.4	Summary	180
7	General discussion.....	190
7.1	Summary of results.....	190
7.2	Scope of FBA in modelling plant metabolism.....	192
7.3	Limitations of FBA in modelling plant metabolism	195
7.4	Future directions in flux balance modelling of plant metabolism	199
8	Literature cited	204
9	Appendix	222
	List of appendices	222
	Appendix 1.....	223
	Appendix 2	225
	Appendix 3	228

1 General introduction

1.1 Statement of aims

The aim of this thesis is to explore the use of genome-scale metabolic models to predict metabolic fluxes in different plant tissues. Within this overall aim, several questions were addressed. Firstly, which modelling parameters are important in giving accurate flux prediction? Secondly, to what extent can heterotrophic and phototrophic metabolism be predicted from a genome-scale metabolic model? Thirdly, what novel insights can be inferred from the model predictions?

1.2 Plant metabolic network

Plant metabolism is both interesting and important to study. From a modelling point of view, one of the intriguing features of plant metabolism is its complexity. The structure of plant metabolic networks is shaped by a combination of multiple factors. Being autotrophic, all metabolites in plant metabolic networks can be synthesised based solely on simple inorganic compounds such as carbon dioxide, nitrate, sulphate and phosphate etc. Being sessile organisms, plants must adapt to short-term variations in abiotic and biotic conditions by regulating metabolic activities and by synthesising a wide-range of secondary metabolites for protection against abiotic stresses and/or interacting with other organisms. Being eukaryotic, the metabolic network of each cell is separated into multiple subcellular compartments, which gives rise to parallel metabolic routes that are inter-connected by the transport of metabolites across compartments. Being multi-cellular, the

metabolic networks in different cell types must interact with each other, both spatially and temporally, to ensure the integrity of the metabolic system of the whole organism. These multiple layers of complexity make the study of plant metabolic networks both fascinating and challenging.

Besides the curiosity in disentangling the complexity of plant metabolic networks, a deeper understanding of plant metabolism is important in addressing major global problems in relation to both humankind and the environment. The gross primary production of terrestrial plants by the carbon fixation through photosynthesis is the largest global carbon flux (Beer et al., 2010). As a key part of the global carbon cycle, a small change in the rate of plant photosynthesis could lead a large shift in the global carbon balance which in turn would have a major impact on the global climate system. The connection of plant metabolic networks to the wider biogeochemical cycles extends beyond the carbon cycle. Plant uptake of nitrogen from the soil and plant-microbe interactions in the rhizosphere are key processes in the global nitrogen cycle (Neff et al., 2003; Näsholm et al., 2009). As plants are one of the key components of the Earth system, the knowledge of the inner workings of plant metabolic networks and its interactions with other biogeochemical processes is crucial in understanding the Earth system as a whole.

In addition to the importance of plants in the Earth system, the survival of human beings relies on the products of plant metabolic networks in the form of nutrition. With the continuing increase in human populations, the search for approaches to improve crop yields is becoming ever more important. The manipulations of plant metabolic networks have yielded some success in

improving the C₃ cycle (Raines, 2011; Maurino and Weber, 2013). For example, an introduction of the bacterial glycolate pathway, a photorespiratory bypass, into *Arabidopsis* resulted in increases in CO₂ assimilation and biomass production (Kebeish et al., 2007). A more ambitious engineering project is the development of C₄ rice by introducing a higher-capacity photosynthetic mechanism, the C₄ cycle, into rice to improve yield (Hibberd et al., 2008; Zhu et al., 2010; von Caemmerer et al., 2012). This requires a large reprogramming of the fluxes through the metabolic network, not only in the carbon fixation pathway but also in other parts of the metabolic networks such as reactions for energy and redox metabolism. In addition to increase in quantity, the improvement of nutritional qualities of crops is also important for human health. Metabolic engineering has been shown to be successful in increasing the contents of a variety of micronutrients, including provitamin A, folate, vitamin E, ascorbate and beta-carotene, in different crop plants (Bhullar and Gruissem, 2013) with the best known example being Golden Rice which produces β-carotene (pro-vitamin A) to help combat vitamin A deficiency (Ye et al., 2000; Paine et al., 2005). Besides nutrition, progress is being made in increasing the production of plant natural products with applications such as colorants, flavours, fragrances, natural pesticides and medicine (Wu and Chappell, 2008; Lim and Bowles, 2012). Rational engineering of metabolism for desirable metabolic phenotypes requires an understanding of the structure and the control of plant metabolic networks. While the synthesis of secondary metabolites can often be studied in isolation due to the relatively few connections with the rest of the metabolic network, a network level understanding of the control of metabolism is

required to engineer central metabolism, as the manipulation of which influences a large part of the metabolic network (Sweetlove and Ratcliffe, 2011). In recent years, efforts in understanding the operation of central metabolism have focused on the quantification of metabolic fluxes at a network level.

1.3 Quantifying metabolic fluxes

Historically, the redistribution of radioisotopes was used to trace metabolic fluxes. With the advance of experimental techniques and the development of related software, metabolic flux analysis (MFA) has been widely used, first in micro-organisms and then in eukaryotes, to characterise and quantify fluxes through the metabolic system (Niklas et al., 2010). For heterotrophic plant tissues, steady-state MFA is the method of choice for analysing metabolic fluxes (Kruger and Ratcliffe, 2009). The system of interest is incubated with a precursor labelled with a stable isotope, typically ^{13}C (Roscher et al., 2000), and the metabolic system is then allowed to reach metabolic and isotopic steady-state (i.e. no change in metabolic flux and isotopic labelling). As the labelled precursor is metabolised through the metabolic system, the label is incorporated into other metabolites in the network. The extent and positions of labelling in each metabolite depends on relative fluxes through different branches of the system, meaning that metabolic fluxes can be inferred from the labelling patterns of the metabolites. After the metabolites are extracted, the labelling profile can be measured using mass spectrometry (MS) and/or nuclear magnetic resonance (NMR) methods (Ratcliffe and Shachar-Hill, 2006). Using available software, e.g. 13CFLUX (Wiechert et al., 2001),

¹³CFLUX2 (Weitzel et al., 2013), FiatFlux (Zamboni et al., 2005), OpenFlux (Quek et al., 2009), a flux map that best describe the experimentally measured labelling patterns can be generated from a metabolic model (Kruger and Ratcliffe, 2009). Steady-state MFA has been applied to more than eight different plant systems (Schwender, 2008), mostly isolated plant cells or tissues under constant culture conditions, and has yielded insights into plant metabolism. In developing oilseed rape embryos, for example, ¹³C-MFA was applied to prove a previously undescribed mode of metabolism involving the use of Rubisco without the Calvin cycle to increase the carbon use efficiency (Schwender et al., 2004; Allen et al., 2009). In heterotrophic *Arabidopsis* cell suspension culture, ¹³C-MFA was used to reveal the changes in metabolic fluxes under a variety of different conditions (Williams et al., 2008; Williams et al., 2010; Masakapalli et al., 2013). However, the application of steady-state MFA to study plant metabolism is limited to heterotrophic or mixotrophic tissues in culture because the only carbon source for phototrophic tissues is CO₂, which is a one-carbon compound with which steady-state positional labelling is not possible (Roscher et al., 2000). This limitation was recently overcome by measuring the time-dependent labelling patterns which was shown to be able to estimate a limited set of metabolic fluxes in illuminated *Arabidopsis* leaves (Szecowka et al., 2013).

While MFA has been widely applied to a number of different plant systems, its implementation requires a substantial commitment of time and effort (Kruger and Ratcliffe, 2009). An alternative and complementary approach for studying metabolic fluxes is metabolic modelling. With the rapid increase in

computing power and developments in the theory and software of metabolic modelling in the past decades, it has become possible to build and analyse complex metabolic networks *in silico* (Covert et al., 2001). The process of metabolic modelling can be divided into two steps, model construction and model analysis. Once a model is constructed and curated, analysis can be done more rapidly than is possible experimentally to investigate the properties of the metabolic network. Metabolic modelling can provide novel insights into the capabilities and behaviour of the system (Feist and Palsson, 2008) as well as being a complementary approach to experimental investigation by generating hypotheses for experimental investigation.

1.4 Metabolic modelling

Metabolic modelling can be broadly separated into two categories, kinetic modelling and structural modelling. In a kinetic model, each metabolic reaction is defined by its stoichiometry and a kinetic rate equation with multiple kinetic parameters. With a defined set of initial values for metabolite concentrations, it is possible to determine both the time-course and steady-state behaviour of the metabolic system. Kinetic modelling is powerful in that the steady-state and time-dependent changes of reaction rates and metabolite concentrations can be stimulated. In addition, with a kinetic model, metabolic control analysis can be applied to quantify the distribution of control over the metabolic system (Fell, 1997). Many plant metabolic pathways including central carbon metabolism and secondary metabolism have been investigated using kinetic modelling (Rios-Esteva and Lange, 2007; Schallau and Junker, 2010; Rohwer, 2012). However, the advantages

of kinetic modelling come with some drawbacks which stem from the large number of kinetic parameters involved and the uncertainties in the correct form for the rate equations. The lack of reliable data about kinetic parameters and the form of rate equations of many enzymes limits the use of kinetic modelling to investigations of relatively small sub-systems within the metabolic network.

In contrast to kinetic models, reactions in structural metabolic models are defined purely in terms of their stoichiometry without any kinetic information. Although the exclusion of kinetic information reduces the predictive power of the model in terms of metabolite concentrations and the time-dependent and regulatory behaviour of the metabolic system, structural modelling allows an examination of metabolic routes across a much larger network. A structural metabolic model can be represented in the form of a stoichiometry matrix, $\mathbf{N}$, with each column representing a reaction and each row representing a metabolite. Each element of the stoichiometry matrix represents the stoichiometry of the metabolite in the corresponding reaction where the production and consumption of metabolites are indicated by positive and negative values respectively. Once a stoichiometrically balanced metabolic model is built, structural analysis can be carried out to predict metabolic fluxes and examine the metabolic capability of the model. The basis of all structural analysis of metabolic models is the steady-state assumption where the rate of production of all internal metabolites equals to their rate of consumption. This can be represented by the equation $\mathbf{N} \cdot \mathbf{v} = 0$, where $\mathbf{N}$ is the stoichiometry matrix of the model and $\mathbf{v}$ is a vector of metabolic fluxes known as a flux mode. In most cases, the number of reactions is greater than

the number of metabolites which means the steady-state solution for metabolic flux modes is underdetermined, i.e. there is more than one possible flux modes that satisfies the steady-state assumption.

One commonly used structural metabolic modelling technique for analysing a metabolic network is elementary mode analysis (Schuster et al., 2000). An elementary (flux) mode can be described as a minimum subset of reactions from the system that can operate at steady-state, i.e. a flux mode in which no non-zero elements can be changed to zero without violating the steady-state assumption. The set of all elementary modes corresponds to the complete metabolic capability of the system. Elementary mode analysis has been applied to model metabolism in the chloroplast with the conclusion that the reductive pentose phosphate pathway and oxidative pentose phosphate pathway, in contrast to traditional view of being separate systems, are complementary and overlapping components of the same system (Poolman et al., 2003). Despite the success of elementary mode analysis in enumerating the capabilities of a metabolic network, it comes with a major limitation in that the application of elementary mode analysis on large metabolic models becomes computationally infeasible due to combinatorial explosion.

To analyse large-scale metabolic models, the most widely used modelling approach is flux balance analysis (FBA). As with other structural modelling approaches, FBA assumes the metabolic system to be at a steady-state. In contrast to elementary mode analysis, which enumerates all the metabolic capability of the entire metabolic system, FBA is a constraint based approach which identifies a subset of flux modes (or flux distributions) that are

biologically relevant. This makes FBA scalable to large-scale metabolic models. In addition to the steady-state constraint, the flux solution space, in which feasible flux solutions lie, can be further restricted by physicochemical constraints such as the rates of substrate consumption and biomass synthesis. In steady-state metabolic flux analysis (MFA), information in isotopic labelling of metabolites is used to determine the *in vivo* flux solution within the flux solution space. In the case of FBA, a biologically relevant flux solution can be identified by assuming that the metabolic network has evolved towards an evolutionary optimum. This allows the search for a biologically relevant flux solution within a bounded solution space to be formulated as an optimisation problem. If the function that is being optimised, known as an objective function, is linear, then the optimisation problem becomes a linear programming problem (Kauffman et al., 2003).

1.5 Objective functions in flux balance analysis

The most commonly used objective function for FBA is the biomass objective function which maximises the synthesis of biomass components in correct proportions for the known biomass composition of the cell (Feist and Palsson, 2010). Despite the common use of the biomass objective function, which is equivalent to maximising molar yield given a fixed amount of substrate, it is by no means universally applicable (Schuster et al., 2008). It was argued that fast but low-yield biochemical pathways can be evolutionarily favoured based on experimental evidence and evolutionary game theory. To address the issue of objective functions in FBA, a systematic evaluation of objective functions in *E. coli* showed that cells grown under different conditions were

best described by different objective functions (Schuetz et al., 2007), supporting the view that the biomass objective function cannot be applied universally. Maximisation of biomass and maximisation of ATP yield were the best objective functions for cells under nutrient scarcity in continuous cultures whereas cells experiencing unlimited growth on glucose in batch cultures was best described by maximization of the ATP yield per flux unit. An alternative approach for investigating the use of objective functions is to use optimisation algorithms to search for the best objective function that gives flux solution which matches best to experimental flux data (Burgard and Maranas, 2003; Gianchandani et al., 2008). The conclusion from these studies is that the synthesis of biomass components was the major component in the objective function which best describes the metabolic fluxes in *E. coli* and *S. cerevisiae*.

While biomass synthesis, i.e. growth, of a cell grown in culture can be directly related to its evolutionary fitness, this concept cannot be applied to cells in multi-cellular organisms because of the division of labour between different cell types. Each metabolically active cell in a multi-cellular organism has one or more metabolic functions, and these must involve a flow of material into and out of the metabolic system. A commonly used objective function in models of multi-cellular organisms is maximisation of molar yield. This assumes the metabolic system is regulated to maximising the metabolic efficiency of conversion from inputs to outputs. In a model of barley seed, the objective function of maximisation of growth was applied while the input of sucrose is constrained (Grafahrend-Belau et al., 2009). Given that carbon is the constraining element in the simulations, the optimisation problem equates

to maximisation of carbon conversion efficiency. This objective function is plausible for the seeds of a crop plant as yield in terms of carbon is an important characteristic selected for during the breeding process when other elements are likely not to be limiting from the application of fertilisers. In contrast, in simulations of developing oilseed rape embryos under heterotrophic conditions, the sum of substrate uptake rates was minimised with a fixed biomass synthesis rate and maintenance ATP cost as constraints (Hay and Schwender, 2011a,b). Given that the amounts of nitrogen, sulphur and phosphate taken up are directly influenced by biomass synthesis, this again equates to maximisation of carbon conversion efficiency. However, when the same system was simulated in photoheterotrophic condition, the carbon conversion efficiency was fixed as an additional constraint and the optimisation problem becomes minimisation of light uptake. This objective function may not be as biologically relevant because the metabolism of plant cells is less likely to be limited by light. Similarly in the simulations of phototrophic metabolism of *Arabidopsis* (de Oliveira Dal'Molin et al., 2010a) and three C_4 plants (de Oliveira Dal'Molin et al., 2010b), photon uptake was minimised while achieving a pre-defined growth rate. This may not reflect the selective pressure on the evolution of metabolism in photosynthetic plant cells as other factors such as the uptake rates of CO_2 , nitrogen and other elements are likely to be important.

Instead of maximising molar yield, the objective function of minimisation of total flux was used to simulate the metabolism in *Arabidopsis* heterotrophic cell culture (Poolman et al., 2009; Williams et al., 2010). Minimisation of total flux assumes that the metabolic system is evolved to minimise the costs of

synthesis of enzymes and membrane transporters to carry out its metabolic function(s). A prerequisite of applying flux minimisation is that the outputs of the metabolic system must be constrained. If the inputs are unconstrained, flux minimisation inherently predicts a low but not necessarily the minimum value of input flux. The major difference between maximising molar yield and minimising total flux is that metabolic efficiency is maximised by the former whereas the efficiency in terms of machinery costs in achieving a certain metabolic function is maximised by the latter. The machinery costs can be related to the nitrogen cost invested in enzymes. In effect, minimisation of total flux can be related to maximising nitrogen use efficiency. In the case of modelling the metabolism of Arabidopsis cells, minimisation of total flux is likely to be more appropriate for a wild plant evolved to deal with limited nitrogen availability as opposed to crop plants which are bred in nitrogen-rich environments. Moreover, when simulating cells or tissues that are not limited by substrate availability, leaves in particular where photon availability is not likely to be limiting, flux minimisation is a more plausible objective function than maximising molar yield. In fact, the fitness of an organism in a range of scenarios could involve a trade-off between multiple factors, in which case multi-objective approaches such as Pareto optimisation will be required (Schuetz et al., 2012). The characterisation of appropriate objective functions for plant metabolic models will be essential to improve the accuracy of flux predictions and also to improve our understanding of the evolutionary driving forces which shape the metabolism of plant cells. In chapter 4, the application of different objection functions in combination with a variety of

constraints was explored to look at its effect on the predictive accuracy of an Arabidopsis genome-scale metabolic model.

1.6 Multiple optimal flux solutions

Despite the optimisation step in FBA, it is rarely the case that a unique optimal flux solution can be identified in large-scale metabolic models using a linear objective function. More often than not, multiple alternative flux solutions exist that achieve the optimal value for a defined objective function due to the existence of alternative metabolic routes. Plant metabolic models are especially prone to alternative optimal flux solutions as many enzymes exist in multiple subcellular compartments. In addition, there is a large number of metabolite transporters across subcellular compartments which further increase the connectivity of the metabolic network. The size of the optimal flux solution space, in which multiple optimal flux solutions lie, could also depend on the nature of the objective functions (Chapter 4).

A method for characterising the optimal flux solution space is flux variability analysis (FVA) which calculates the allowable range of flux values that a reaction can carry while achieving the optimal objective value (Mahadevan and Schilling, 2003). Reactions that can carry variable flux values form alternative metabolic routes in achieving the defined metabolic function set by the model constraints and objective function. Alternative flux solutions can be interpreted as the flexibility and robustness of the metabolic system. Using an extended formulation of FVA, the flexibility of the metabolic system of *E. coli* was shown to decrease with the increase in growth rate, suggesting that there is a trade-off between growth rate and flexibility (Chen et al., 2009). In a

large-scale metabolic model of developing oilseed rape embryos with the objective function of minimising nutrient uptake, FVA was applied to categorise reactions based on flux variability types (Hay and Schwender, 2011b). It was found that 10-11% of all reactions are substitutable and 13% are variable, which indicate the presence of metabolic redundancy in the system. Besides looking at variability types within each condition simulated, 57 out of 572 reactions in the model showed changes in variability types across four nutritional conditions, suggesting their role in metabolic adjustments in response to changes in nutritional conditions.

While the variability in fluxes can reveal some information about the metabolic system, there are scenarios where the ability to obtain a single flux solution is preferable. Examples of such scenarios include comparisons with ^{13}C -MFA flux maps and/or multiple flux solutions and correlation analysis with high-throughput data such as gene expression and protein levels. In addition, many FBA studies report a single optimal flux solution as the result of their analyses despite the existence of alternative optimal flux solutions. The choice of the flux solution from the set of alternative optimal solutions is dependent on the software and/or algorithm used to solve the linear programming problem. Different approaches have been applied to identify a single flux solution from an optimal flux solution space. In a model of barley seeds, a two-step optimisation approach was applied, first with a linear objective function of maximising growth followed by a quadratic objective function of minimising the squared sum of fluxes (Grafahrend-Belau et al., 2009). The nature of a quadratic objective function enables a single unique flux solution to be identified. However, not all software packages and solvers

have the capability of solving quadratic programming problems. An algorithm based on linear programming was developed for identifying a single flux solution which is fully reproducible irrespective of the software and linear programming solver used (Smallbone and Simeonidis, 2009). This algorithm identifies the geometric centre of the optimal flux space as an extension of minimising the sum of fluxes. In chapter 6, a novel modelling approach was developed based on minimisation of sum of fluxes using random weighting factors, which allowed the generation of a single averaged flux solution as well as the investigation of variability in fluxes.

1.7 Genome-scale metabolic models

The construction of structural metabolic models at genome-scale has become an increasingly common practice for many organisms due to both the increase in availability of genome sequences and annotations and the advances in development of software and algorithms for model construction. Genome-scale metabolic models are constructed by extracting the metabolic reactions catalysed by enzymes encoded in the genomes of the organisms of interest. Detailed processes in the construction of genome-scale metabolic models are described in Chapter 3. Since 1999, more than 80 genome-scale metabolic models have been published spanning all three domains of life (Kim et al., 2012). To date, genome-scale models of five plants species have been published including Arabidopsis (Poolman et al, 2009; de Oliveira Dal'Molin et al., 2010a; Radrich et al., 2010; Saha et al., 2011; Chung et al., 2012; Mintz-Oron et al., 2012), maize (de Oliveira Dal'Molin et al., 2010b; Saha et al., 2011), sorghum (de Oliveira Dal'Molin et al., 2010b), sugarcane

(de Oliveira Dal'Molin et al., 2010b) and rice (Poolman et al., 2013). In this study, an updated Arabidopsis genome-scale metabolic model was constructed with improved sub-cellular compartmentation and model curation (Chapter 3).

Once a genome-scale metabolic model is established, there are a wide range of applications which can be grouped in five main categories (Oberhardt et al., 2009): 1) interpretation of high-throughput data such as microarray and metabolomics data, 2) guidance of metabolic engineering by identifying possible enzyme targets, 3) generation of experimentally testable hypotheses, 4) investigation of multi-species or multi-tissue relationships, 5) discovery of network properties such as metabolite connectivity and pathway redundancy. While many of these applications refer to microbial models, the published plants models have also been applied towards various ends. A genome-scale metabolic model of Arabidopsis was coupled with transcriptomics data to reveal metabolic pathways involved in light and temperature acclimation (Töpfer et al., 2013). Using tissue-specific models of Arabidopsis generated based on protein expression data, 71 knock-out target enzymes were predicted for increasing vitamin E accumulation (Mintz-Oron et al., 2012). In another study on metabolic engineering, an Arabidopsis genome-scale model was utilised to characterise the optimal metabolic flux modes which were subsequently incorporated into models of microbial hosts for improving terpenoid yield (Chung et al., 2012). These two studies generated model predictions which could be investigated experimentally. With respect to the relationship between plant species, a comparison between metabolic models of maize and Arabidopsis revealed 445 metabolic reactions in both primary

and secondary metabolism which are only in maize but not Arabidopsis (Saha et al., 2011). In the models of C₄ plants, the interactions between the mesophyll and the bundle sheath cells were modelled in all three subtypes of C₄ photosynthesis (de Oliveira Dal'Molin et al., 2010b). In terms of comparisons between plants and other organisms, the topological properties of an Arabidopsis model, including the degree distribution, average clustering coefficient, average path length etc., were analysed and were found to be similar to metabolic networks of other organisms (Radrich et al., 2010). With the increasing number of sequenced genomes of plant species, one can envision the possibility of systematic comparisons between the genome-scale metabolic networks of plant species. An area which is yet to be explored in the field of plant genome-scale modelling is the metabolic interactions between plants and other organisms. Genome-scale metabolic models have been utilised to model interactions between two microbial species (Stolyar et al., 2007; Zhuang et al., 2011) and the host-pathogen interactions between human alveolar macrophage and *M. tuberculosis* (Bordbar et al., 2010). The application of genome-scale metabolic modelling in investigating the metabolic interactions between plants and pathogens and/or soil microbiome could become a reality in the near future with the increasing availability of genomic data for model construction and the continuous development of modelling approaches.

1.8 Modelling metabolism of specific tissues and cell types

A major challenge in applying genome-scale metabolic models to simulate metabolic fluxes in multi-cellular organisms is the division of labour which means each cell type carries out distinct metabolic functions. However, a genome-scale model contains the full set of metabolic enzymes for carrying out the metabolic functions in all cell types. There are two complementary approaches in using genome-scale models to predict metabolic fluxes in specific cell types and tissues. The first approach involves the construction of cell type- or tissue-specific models, each containing a subset of reactions in a generic genome-scale model. This has been done extensively for the human genome-scale model based on a variety of high-throughput data. For instance, a liver-specific model was generated based on an algorithm which integrates literature-based knowledge, transcriptomic, proteomic, metabolomic and phenotypic data (Jerby et al., 2010). For a more extensive coverage of different cell types and tissues, models of 69 human cell types were constructed using protein abundance as the major source of information (Agren et al., 2012). Information on gene expression data and network topology was integrated to generate models of 126 human tissues (Wang, Y. et al., 2012). In regard to plant genome-scale metabolic models, 10 tissue-specific models of Arabidopsis were constructed based on protein expression data (Mintz-Oron et al., 2011). Assuming that the cell type- or tissue-specific models accurately reflect the structure of the metabolic networks *in vivo*, these models are more likely to give accurate flux predictions compared to the generic model as the generic model contains alternative metabolic routes

which are absent in cell type- or tissue-specific models. Moreover, the availability of cell type- or tissue specific models allows the metabolic capabilities of each cell type or tissue can be explored.

While cell type- and tissue-specific models have been generated for human and Arabidopsis, this is not applicable for organisms in which little or no cell type- or tissue-specific information is available. In which case, cell type- and tissue-specific metabolism can be modelled using a generic genome-scale model based on cell type- or tissue-specific inputs and/or outputs constraints. These constraints define the metabolic function(s) of the specific cell type or tissue. By assuming that the metabolic system of each cell type and tissue is regulated such that its metabolic function(s) is/are carried out in an optimal manner, optimisation with an appropriate objective function can be applied to predict the metabolic fluxes in the cell type or tissue of interest. This has been applied to model the metabolism in Arabidopsis heterotrophic cell culture (Poolman et al, 2009; Williams et al., 2010), photosynthetic and heterotrophic metabolism of a generic Arabidopsis cell (de Oliveira Dal'Molin et al., 2010a) and a developing leaf cell of rice (Poolman et al., 2013). In models of C_4 plants, a handful of cell type-specific constraints for the mesophyll and the bundle sheath cells were applied to model C_4 photosynthesis (de Oliveira Dal'Molin et al., 2010b; Saha et al., 2011). Using a generic Arabidopsis genome-scale metabolic model, metabolic fluxes were predicted in heterotrophic cell culture (Chapter 4) and in C_3 and CAM leaves (Chapters 5 and 6). Although flux predictions from a generic model may be less accurate than from cell type- or tissue-specific models due to the presence of alternative metabolic routes, the discrepancies between model

predictions and experimental evidence could provide insights into the choice of metabolic routes in specific cell types and tissues. An example of such discrepancy is discussed in Chapter 5 which relates to starch degradation in C_3 leaf.

Although the metabolism of each cell type or tissue can be predicted in isolation by defining its inputs and outputs, these exchanges are connected to the inputs and outputs of other cell types or tissues. A number of studies on modelling of metabolic interactions between cell types and tissues in multi-cellular organisms have been published in the past three years. Metabolic interactions between different cell types in the human brain were modelled which provided insights into the differential effects of Alzheimer's disease on different cell types and regions of the brain (Lewis et al., 2010). The metabolic functions of three human cell types, liver cells, fat cells and skeletal muscle cells, were integrated to simulate three physiologically important metabolic states in the human body: the Alanine cycle, Cori cycle, and absorptive state (Bordbar et al., 2011). In models of C_4 plants, the interactions between mesophyll cell and bundle sheath cell were modelled as part of the C_4 cycle (de Oliveira Dal'Molin et al., 2010b; Saha et al., 2011). A multi-organ model of barley plant metabolism with 971 reactions was constructed by integrating organ-specific models of leaf, stem and seed and a simplified root model (Grafahrend-Belau et al., 2013). The integration of the static multi-organ model with a dynamic whole-plant multi-scale functional plant model through dynamic FBA allowed the investigation of metabolic behaviour of source and sink organs of the barley plant during the generative phase which revealed a sink-to-source shift of the barley stem as a result of

the senescence-related decrease in leaf source capacity. This approach can potentially be applied with genome-scale metabolic models and extended to model different stages of plant development.

1.9 Summary

A computational model that accurately reflects the behaviour of the system allows rapid and large-scale experimentation to be performed which accelerates the discovery of scientific knowledge and the subsequent technological advances. A model of plant metabolism would enable the identification of targets for manipulation to achieve desired metabolic phenotypes such as increased yield of desired metabolic products. In view of this, a heavily-curated Arabidopsis genome-scale metabolic model was constructed with extensive subcellular compartmentation (Chapter 3). Using this model, the effect of modelling parameters, such as the inclusion of cellular maintenance costs and the use of objective functions, on the accuracy of the model predictions was investigated in the system of heterotrophic cell culture (Chapter 4). A modelling approach was developed to model leaf metabolism over a day-night cycle which was applied to predict metabolic fluxes in C_3 and CAM leaves (Chapter 5). Finally, the model was used to investigate the metabolic response of a C_3 leaf with varying environmental and physiological conditions (Chapter 6).

2 Modelling methods

2.1 Modelling software package – ScrumPy

The construction of a genome-scale metabolic model and the subsequent structural analyses of the model by flux balance analysis and related techniques in this study were carried out using ScrumPy, a Python-based software package (Poolman, 2006). ScrumPy contains a number of modules for performing various types of model analysis such as elementary modes analysis and linear programming. In addition, the functionalities of ScrumPy are easily extendable to by importing codes written for performing more complex and/or specific analyses. This can be done by the command “import codefile” where “codefile.py” is the name of the file containing the codes. In this study, additional codes were written for the construction of an Arabidopsis genome-scale metabolic model and subsequent constraint-based analyses on the model (Appendix S1).

2.2 Construction of genome-scale metabolic model using ScrumPy

A genome-scale metabolic model of Arabidopsis was constructed in this study (Chapter 3) using the codes in “BuildAra.py” (Appendix S1). Once BuildAra.py was imported into ScrumPy, a compartmented Arabidopsis genome-scale metabolic model can be constructed by calling the function “BuildAra.BuildModel()”. This calls several functions which are discussed in detail in the following subsections. A prerequisite is that the files in the AraCyc database (Mueller et al., 2003) are contained in a directory called

“/home/maurice/database/aracyc”. Alternatively, the path in line 31 of ‘BuildAra.py’ can be changed to the path corresponding to the location of the database files. Another prerequisite is the existence of four files, “Unwanted.py”, “Substitutes.py”, “Compartment.py” and “Corrections.spy” (Appendix S1), which contain information for model construction as described in the following subsections and in chapter 3.

2.2.1 Extracting reactions from AraCyc

Information in the raw AraCyc data files was extracted based on the PyoCyc module in ScrumPy. First, a database object was set up using the function “db=BuildAra.SetUpDB()” which calls to function “PyoCyc.Organism()”. A raw model was generated in the form of a stoichiometry matrix of all reactions in “reactions.dat” file of AraCyc by the command “sm=BuildAra.MakeAll()” where “sm” is the stoichiometry matrix.

2.2.2 Incorporating changes into the raw model

Multiple changes were incorporated into the raw model including metabolite and reaction names, reaction stoichiometry and the addition and removal of metabolites and reactions from the model. “BuildAra.SubstMets(sm)” removes the character “[” from metabolite names. Metabolites in the list called “Mets” in “Unwanted.py” file and all reactions involving these metabolites were removed from the raw model using the command “BuildAra.RemoveBadMets(sm)”. These are mainly metabolites which are ill-defined and problematic (see chapter 3 for details). Reactions in the list “Reacs” in “Unwanted.py” were removed by calling “BuildAra.RemoveBadReacs(sm)”. Metabolites in the keys of a dictionary called “MetNames” in “Substitutes.py”

were substituted by the corresponding value of the dictionary using the command "BuildAra. SubMetabolites(sm)". Reactions involving the metabolites "NAD(P)", "NAD(P)H", "NAD-P-OR-NOP" and "NADH-P-OR-NOP" were split into pairs of reactions with the command "BuildAra.FixNads(sm)". This was done by removing the original reaction and adding two reactions, NAD-dependent and NADP-dependent reactions. Duplicate isostoichiometric reactions were removed by "BuildAra. Dellsoforms(sm)". Reactions with incorrect stoichiometries were replaced by reactions with manually corrected stoichiometries contained "Corrections.spy" using the command "BuildAra. AddCorrections(sm)".

2.2.3 Defining the direction and reversibility of reactions

The default direction and reversibility of reactions were taken from the AraCyc database using "BuildAra. DefaultReacDirec(sm)". Reactions with manually curated direction and reversibility were recorded in several lists in "Substitutes.py". Reactions manually curated to be reversible, irreversible in the pre-defined direction and irreversible opposite to the pre-defined direction were in lists "Reversibles", "Irreversibles" and "BackReactions". These changes were incorporated into the model using the commands "BuildAra. MakeReversibles(sm)", "BuildAra. MakeIrreversibles(sm)" and "BuildAra. SwapDirections (sm)".

2.2.4 Compartmentation of the model

The information on the manually assigned reaction compartmentation was recorded in the file "Compartments.py" (Appendix S1) in four lists, "Cytosol", "Plastid", "Mitochondria" and "Peroxisome", which contain reactions in the

corresponding compartments. Note that each reaction can be assigned to more than one compartment. Reactions not included in the four lists were assigned to the cytosol. Based on information in “Compartments.py”, the function “BuildAra.MakeCompartments(sm)” was used to compartment the model by adding suffixes, “_c”, “_p”, “_m” and “_x” to names of metabolites and reactions in the cytosol, plastid, mitochondrion and peroxisome respectively.

2.2.5 Model formats

The “BuildAra.BuildModel()” function generates a model file in ScrumPy format. A structural metabolic model in ScrumPy format begins with “Structural()”. The definition of each reaction is described in chapter 3 (see Figure 3.1B for example). One feature of the ScrumPy format is that it is modular which means the model can be divided into multiple models files. The model can be loaded into ScrumPy through a master file which connects the multiple model files. The master file contains a list of external metabolites and a list of model files which together form the whole model. For the Arabidopsis genome-scale model constructed in this study, “Arabidopsis.spy” is the master file which links to three model files, “FromAraCyc.spy”, “Transporters.spy” and “Extras.spy”, containing reactions from AraCyc, intracellular and extracellular transport reactions and other manually added reactions respectively (Appendix S2).

The model in ScrumPy format can be converted in SBML format by first importing “SBMLPlus.py” (Appendix S1). Then the model in ScrumPy format is loaded using the command “m=ScrumPy.Model()”. Finally an SBML model

file can be generated using the command “SBMLPlus.Spy2SBML(m)”, which writes the information in the loaded model in SBML format (Appendix S2). The prerequisite is that libSBML, a programming library for working with SBML files (<http://sbml.org/Software/libSBML>), must be installed on the system.

2.3 Construction of a diel metabolic model

To model leaf metabolism over a day-night cycle, a diel modelling framework was developed by dividing the day-night cycle into two phases, light and dark (chapter 5). The metabolism in each phase was assumed to be at steady-state. First, a diel metabolic model was constructed by duplicating the Arabidopsis genome-scale metabolic model, with the reactions and metabolites in each duplicate labelled “_Light” or “_Dark” before the compartmentation suffix, representing the light phase and the dark phase respectively. This was done by first loading the compartmented Arabidopsis model, ‘m’, and importing the “BuildAra.py” code into ScrumPy, then the command “BuildAra.Tissue(m.smexterns, [“Light”, “Dark”])” was called to generate two model files, “Light.spy” and “Dark.spy” (Appendix S3), which are the two duplicate models described above. In addition, dummy reactions for “transporting” storage metabolites from one phase to another were manually added with suffix “_LightDark” in the file “TransporterLightDark.spy” (Appendix S3). A master file, “AraLightDark.spy”, connects the three files, “Light.spy”, “Dark.spy” and “TransportLightDark.spy”, to form the diel metabolic model (Appendix S3). The diel model was converted into SBML format as describe in section 2.2.5 (Appendix S3).

2.4 Identification of stoichiometric inconsistencies

Stoichiometric inconsistencies of the model were detected using an algorithm based on mixed-integer linear programming (Gevorgyan et al., 2008). The mixed-integer linear programming problem was formulated to identify possible distribution of molecular masses which fulfill the assumption of mass conservation. A list of unconserved metabolites, metabolites which cannot be assigned a positive mass, can be identified based on this algorithm. To implement this in ScrumPy requires the “StoCons” module (Appendix S1). The code for running this algorithm can be imported into ScrumPy by calling the command “from StoCons import StoiCons”. With a stoichiometric model loaded in ScrumPy, named “m”, the transpose of the stoichiometry matrix must first be generated by “m.smexterns.Tranpose()” before computing the list of unconserved metabolites using “StoiCons.UnconservedMets(m.smexterns)”.

2.5 Performing flux balance analysis in ScrumPy

Flux balance analysis is a modelling technique that uses linear programming to calculate a flux solution which maximises or minimises a defined objective function given the steady-state constraint and other pre-defined flux constraints. Mathematically, it can be formulated as follows.

maximise or minimise $\mathbf{Z} = \mathbf{c} \cdot \mathbf{v}$

subject to: $\mathbf{N} \cdot \mathbf{v} = 0$

$\min_i < \mathbf{v}_i < \max_i$

where $\mathbf{Z}$ is the objective function, $\mathbf{c}$ is a vector of weighting factors, $\mathbf{v}$ is a vector of metabolic fluxes, $\mathbf{N}$ is the stoichiometry matrix, $\mathbf{v}_i$ is the i^{th} element of $\mathbf{v}$ and $\min_i$ and $\max_i$ are the minimum and maximum constraints of $\mathbf{v}_i$. The linear programming problem was solved using the Gnu Linear Programming Kit (<http://www.gnu.org/software/glpk>) through the “LP” module of ScrumPy.

To set up a linear programming problem for FBA in ScrumPy, a stoichiometric model, “m”, must first be loaded. A linear programming object, “lp”, can then be created using the pre-installed “LP” module of ScrumPy by the command “lp=ScrumPyLP.lp(m)”, given that the “ScrumPyLP” code in the “LP” module was imported with “from LP import ScrumPyLP”. The functionalities of the “LP” module in ScrumPy were extended with the codes “LPPlus.py” and “LPARA.py” (Appendix S1) to perform the linear programming analyses in this study. “LPPlus.py” contains general functions applicable for any models whereas “LPARA.py” contains Arabidopsis-specific functions and information for the analyses in this study. To use the general extended functions, “LPPlus.py” can be imported to generate a linear programming object by “lp=LPPlus.lp(m)”. To perform the linear programming analyses specific for this study, “LPARA.py” was imported and a linear programming object was created using the command “lp=LPARA.lp(m)”.

Constraints of individual reactions were applied by the command “lp.SetConstraint(reac,lo,hi)” where “reac” is the reaction name and “lo” and “hi” are the lower and upper constraints of the reaction. As well as numerical values, “None” can be entered as the values of “lo” and/or “hi” representing no lower and/or upper constraint. To set a fixed flux value for a reaction, “lp.SetFixedFlux(constraintdic)” can be used instead where “constraintdic” is

a dictionary with keys as reaction names and values as the defined fixed flux values of the reaction. Fluxes of two reactions can be set to a fixed ratio, by setting the two reactions to carry zero flux and adding a new temporary reaction with a stoichiometry corresponding to the combination of the two reactions at the defined fixed ratio. A fixed ratio of “p1”：“p2” of two reaction fluxes, “reac1” and “reac2” can be set by the command “lp.SetReacsFixedRatio({reac1:p1,reac2:p2})”. This function was called by all functions involving setting a fixed ratio or proportion of two fluxes described in the following subsections. To set a weighted sum of multiple reaction fluxes to a fixed value, “x”, a temporary row (dummy metabolite) was added with the entry of the row corresponds to the weightings of the reactions (stoichiometry of the production or consumption of the dummy metabolite) and “x” unit of the dummy metabolite was set to be removed from the system. The command “lp.SetSumReacsConstraint({reac_i:p_i},x)” sets a constraint of $\sum(p_i \times v_i) = x$, where reac_i is the reaction name corresponding to the reaction flux v_i . ScrumPy commands for setting model constraints specific in this study are described in Appendix 1.

An objective function in the form of $\sum(c_i \times v_i)$, where c_i is the weighting of the reaction flux v_i in the objective function, can be set using the command “lp.SetObjective(objective)” where “objective” is an objective function in the form of “{reac_i:c_i}” with “reac_i” being the reaction name corresponding to v_i . “lp.SetObjDirec(objdirec)” sets the direction of the objective where objdirec can take the values of “Min” or “Max” representing minimisation and maximisation respectively. Given that the “LP” module in ScrumPy splits reversible reactions into forward and backward reactions that can only carry

positive flux, the objective function of minimisation of the sum of fluxes, $\sum |\mathbf{v}_i|$, can be set simply by setting the weighting of all forward and backward reactions to one with the command “`lp.SetMinFluxObj()`”. Details on the use of random weighting factors in flux minimisation are described in section 2.6.

Once a linear programming program is set up, it can be solved using the function “`lp.Solve()`” which displays the status of the solution as “Optimal” (an optimal solution is reached), “No feasible” (no feasible solution exists) and “Unbounded!” (the linear programming problem is unbounded). If an optimal solution exists, it can be obtained by “`sol=lp.GetSol()`” where `sol` is a dictionary of the computed optimal flux solution with reaction names as keys and the flux values as values of the dictionary. Alternatively, the computed optimal flux solution, sorted by the flux value, can be displayed in the ScrumPy terminal by “`lp.PrintSol()`” with the options for filtering by reaction names and/or flux values.

2.6 Flux minimisation with random weighting factors

A method based on the use of random weighting factors in the objective function of flux minimisation was developed to explore the effect of variability in the enzyme machinery costs on the resulting flux solutions (chapter 6).

The minimisation of total flux objective function described in section 2.5 assumed each reaction to have equal weighting, i.e minimise $\sum |c_i \times \mathbf{v}_i|$ where the weighting, c_i , was identical (a value of 1 was used in practice) for all reactions and $\mathbf{v}_i$ represents reaction fluxes. In the random weighting factors approach, c_i of each reaction was assigned a random value between 0 and 1

from a uniform distribution and a flux solution was obtained by minimising $\sum |c_i \times v_i|$. Multiple flux solutions, 1000 in this study, were computed by repeatedly solving optimisation problems, each time with a different set of random weighting factors, c_i . An averaged flux solution was calculated by averaging the multiple flux solutions generated with different sets of c_i . The general algorithm can be implemented in ScrumPy by

“solmtx=lp.RandomMinFlux(runs)” where “solmtx” is a matrix of multiple flux solutions, “runs” is the number of flux solutions to be generated and “lp” is a linear programming object generated by either “lp=LPPlus.lp(m)” or “lp=LPAra.lp(m)” as described in section 2.5. From a matrix of multiple flux solutions, “solmtx”, the command “avgsol=solmtx.AverageFlux()” can be applied where “avgsol” is the averaged flux solution. For the plant-specific analysis performed in chapter 6, reactions representing CO₂ and O₂ diffusion across membrane, accumulation of storage compounds, biomass synthesis, generic ATPases and NADPH oxidases for maintenance costs were set to have weighting factors of zero (see the function “RandomMinFluxAra” in “LPAra.py” in Appendix S1 for details of reaction names) as these reactions were included for modelling purposes and have no associated enzyme machinery costs. In modelling leaf metabolism in chapter 6, with a defined set of constraints, 1000 flux solutions were generated with

“solmtx=lp.RandomMinFluxAra(1000)” where “lp” was created by “lp=LPAra.lp(m)”.

2.6.1 Variability of fluxes in multiple solutions from flux minimisation with random weighting factors

From multiple flux solutions, the variability of a reaction across the multiple flux solutions was inferred by inspection of the range of flux values of the reaction over the multiple flux solutions. In addition, variability of fluxes could be inferred from quantitative measures including relative standard deviation (RSD) and relative average deviation (RAD), defined by the equations:

$$\text{RSD} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (v_i - v_{\text{mean}})^2}}{v_{\text{mean}}} \text{ and } \text{RAD} = \frac{\frac{1}{N} \sum_{i=1}^N |v_i - v_{\text{mean}}|}{v_{\text{mean}}}$$

where v_i is the flux value from each flux solution, v_{mean} is the mean flux value over all flux solutions and N is the number of flux solutions. A value of RSD or RAD close to 0 represents low flux variability. The ranges of flux values of all reactions contained in a matrix of multiple flux solutions, “solmtx”, were calculated in ScrumPy by “fr=solmtx.FluxRange()” where “fr” is a dictionary with reaction names as keys and flux ranges as values in the form of ($\mathbf{v}_{\text{min}}$, $\mathbf{v}_{\text{max}}$), where $\mathbf{v}_{\text{min}}$ and $\mathbf{v}_{\text{max}}$ are the minimum and maximum flux values respectively. RSD and RAD of a reaction flux could be calculated using “solmtx.RelStDev(reac)” and “solmtx.RelAverageDev(reac)”, respectively, where “reac” is the name of the reaction.

2.6.2 Quantifying the relatedness between reactions by the Pearson correlation coefficient

The Pearson correlation coefficient between two reaction fluxes across the multiple flux solutions generated from flux minimisation with random weighting factors was used as a measure of relatedness between the two

variable reactions, reactions which can carry variable fluxes, given a defined set of model constraints (chapter 6). For irreversible reactions, a Pearson correlation coefficient of 1 means that the reactions always vary together in the same direction whereas competing reactions, reactions which vary in opposite directions, have a Pearson correlation coefficient of -1. For reversible reactions, the sign of Pearson correlation coefficient depends on the definition of the reaction. Two independent reactions have a Pearson correlation coefficient of 0. Note that the Pearson correlation coefficient can only be calculated for reactions carrying variable flux, where reactions with fixed flux over the multiple flux solutions are by definition correlated with each other. The Pearson correlation coefficient between two reactions in a matrix of multiple flux solutions, “solmtx”, was calculated in ScrumPy by the command “solmtx. FluxCorrCoef(reac1, reac2)” where “reac1” and “reac2” are the names of two reactions of interest.

2.6.3 Calculating flux solutions with varying constraints using flux minimisation with random weighting factors

The responses of leaf metabolism to varying light intensity and nitrogen availability were modelled by varying the corresponding constraints and calculating flux solutions using flux minimisation with random weighting factors (chapter 6). This was implemented in ScrumPy using the command “scanmtx=lp. ConstraintRandomMinFluxScanAra(constraintdic, lo, hi, n_p, iter)” where “scanmtx” is a matrix of averaged flux solutions, “lp” is a linear programming objective created by “lp=LPAra.lp(m)”, “constraintdic” is a constraint dictionary as input of the function “lp.SetFixedFlux(constraintdic)” described in section 2.5, “lo” and “hi” are the lower and upper values of the

varied constraint flux, “n_p” is the number of steps modelled between “lo” and “hi” and “iter” is the number of flux solutions generated in the random weighting factor algorithm for calculating the averaged flux solutions. The function “lp.ConstraintRandomMinFluxScan(constraintdic,lo,hi,n_p,iter)” can be used for a general model where “lp” is a linear programming object created by “lp=LPPlus.lp(m)” with the definitions of the other parameters the same as before.

2.7 Pareto optimisation

Pareto optimality analysis was performed by linear programming using the weighted sum method (Vo et al., 2004). For the set of objective functions of interest, Z_i , 10000 linear programming problems were solved with the objective function being a weighted sum of Z_i , $\sum(c_i \times Z_i)$, where c_i is the set of weighting factors for Z_i . A random set of weighting factors, c_i , where each value was uniformly distributed between 0 and 1, was generated for each of the 10000 linear programming problems solved. A set of vertices (corners) of the Pareto front, “po”, was calculated using the function “po=lp.Pareto(objectives,objdirec,runs)” where “objectives” is a list of objective functions in the form used as input of the function “lp.SetObjective(objective)” described in section 5.2.1, “objdirec” is the direction of the objective (“Min” or “Max”) and “runs” is the number of linear programming problems to be solved (10000 in the case of this study). Given that all the constraints and objectives were linear, the Pareto front was formed by fitting straight lines between adjacent vertices.

2.8 Flux variability analysis

Flux variability analysis (FVA) was implemented to characterise the optimal flux solution space using the method described by Mahadevan and Schilling (2003). First, the optimal objective value for the primary objective function was computed. Then the primary objective function was set to the computed optimal value as an additional constraint to the linear programming problem, i.e. constraining the feasible flux space to be a flux space optimal towards the primary objective function. This was done by setting a weighted sum of reactions (the primary objective function) to be a fixed value (the optimal objective value) described in section 2.5. For each reaction of interest, the reaction flux was maximized and minimized by giving it a positive weighting in the objective function with all other reactions having a weighting of zero. The minimum and maximum values represent the flux range that was compatible with the optimal value for the primary objective function. FVA was performed in ScrumPy using the command “fva=lp.FVA(reaclist)” where “reaclist” is a list of names of reactions of interest and “fva” is a dictionary with reaction names as keys and flux ranges as values in the form of ($\mathbf{v}_{\min}$, $\mathbf{v}_{\max}$), where $\mathbf{v}_{\min}$ and $\mathbf{v}_{\max}$ are the minimum and maximum feasible flux values respectively. “fva=lp.FVA()” gives the flux ranges of all reactions in the model. The flux range of a grouped reaction, an overall reaction which is a weighted sum of multiple reactions, was calculated by minimizing and maximizing the weighted sum of the reactions. This was done in ScrumPy by replacing the reaction names in “reaclist” with dictionaries of the weighted sum of reactions with keys being the reaction name and values being the weightings, e.g. “fva=lp.FVA([{"reac1":3, "reac2":10}])”, where the grouped

reaction is $3\text{“reac1”} + 10\text{“reac2”}$. Reactions with different types of flux ranges can be extracted, e.g. reactions with variable flux range can be extracted with `“fva.Variable()”`. Further details can be found by the command `“dir(fva)”` or in the `“LPPlus.py”` file.

2.9 Comparison between FBA flux solution and ^{13}C -MFA flux map

Flux solutions generated from FBA were quantitatively compared to experimental flux maps estimated from ^{13}C -metabolic flux analysis (MFA) in chapter 4. To assess the match between the FBA flux solution and the ^{13}C -MFA flux map, reactions in the FBA model were mapped to reactions in the ^{13}C -MFA model. In models of ^{13}C -MFA, multiple enzyme-catalysed reaction steps between branch points are reduced to a single step, therefore multiple reactions in the Arabidopsis genome-scale metabolic model must be mapped onto a single reaction in the ^{13}C -MFA model as described in Appendix 2. For each mapped reaction corresponding to the ^{13}C -MFA model, the flux range from the genome-scale model was calculated by maximizing and minimizing the genome-scale model reactions that are mapped to the ^{13}C -MFA model reaction, with the objective function being the stoichiometry of the mapping listed in Appendix 2. The mean values of the 15 fluxes form feasible flux solutions for all 25 combinations of objective function and constraints (chapter 4), and these values were used for comparison with the ^{13}C -MFA flux map (Appendix S4). The quantitative measure of the mismatch between FBA and MFA fluxes used in this study was the weighted Euclidean distance between the two flux solutions calculated as:

$$\sqrt{\sum_{i=1}^{n_{map}} \frac{(v_{MFA} - v_{FBA_{mean}})^2}{(\sigma_{MFA})^2}}$$

where the subscripts *MFA* and *FBA* denote flux values obtained from ¹³C-MFA and FBA respectively, $v_{FBA_{mean}}$ is the mean flux value obtained using FVA, σ_{MFA} is the MFA experimental standard error and n_{map} (= 15 in this study) is the number of reactions mapped between MFA and FBA models. In practice, this was done in ScrumPy by first setting up the linear programming problem for FBA and then running the function “lp.FluxDistDiff()” which calculates and displays the weighted Euclidean distance as well as the mean flux value and flux ranges of the mapped FBA reactions. This function requires data for the MFA flux map contained in “MFADData.mapTomMFAcpmtabs” (Appendix S1).

2.10 Quantitative comparison between flux ranges calculated from the diel model and the single steady-state model

Flux ranges of reactions in the light were calculated using flux variability analysis (section 2.8) for the diel model, $(\mathbf{v}_{min}^{diel}, \mathbf{v}_{max}^{diel})$ and the single steady-state model, $(\mathbf{v}_{min}^{single}, \mathbf{v}_{max}^{single})$. Firstly, reversible reactions which carry fluxes in opposite directions in the two sets of flux ranges were identified using the function “lp.FluxDiffDirec(fd1,fd2)” where “fd1” and “fd2” are the outputs of FVA, described in section 2.8, of the two models. This was done by identifying reactions where either $\mathbf{v}_{min}^{diel} \geq 0$ and $\mathbf{v}_{max}^{single} \leq 0$ or $\mathbf{v}_{min}^{single} \geq 0$ and $\mathbf{v}_{max}^{diel} \leq 0$. Reactions with overlapping flux ranges were identified with the

command “lp. FluxRangeOverlap(fd1,fd2)” where the definitions of “fd1” and “fd2” are the same as before. This was done by first getting the common set of reactions in “fd1” and “fd2”, then eliminate reactions with non-overlapping flux ranges ($v_{\min}^{\text{diel}} > v_{\max}^{\text{single}}$ or $v_{\min}^{\text{single}} > v_{\max}^{\text{diel}}$). For reactions with non-overlapping flux ranges and do not carry fluxes in opposite directions in the two sets of flux ranges, a similarity measure between the two flux ranges were calculated as $v_{\max}^{\text{low}} / v_{\min}^{\text{high}}$, where $v_{\max}^{\text{low}}$ is the absolute maximum value of the flux range with lower absolute flux and $v_{\min}^{\text{high}}$ is the absolute minimum value of the flux range with higher absolute flux. This measure varies from 0 to 1 where a value close to 0 represents low similarity (large difference) between the flux ranges. A value of 0 means the flux range with lower absolute flux is (0,0). The similarity measure was calculated in ScrumPy using the function “lp.FluxRangeDiff(fd1,fd2)” where the definitions of “fd1” and “fd2” are the same as before (see “LPPlus.py” in Appendix S1 for details).

2.11 Calculations of ATP and NADPH requirements in leaves during the day from flux solutions

The ATP and NADPH requirements during the day were calculated based on the flux solutions generated from the diel model. ATP requirement was calculated as the sum of ATP produced by the plastidial ATP synthase and the mitochondrial ATP synthase in the light given by the equation:

ATP requirement in the light

$$= (v(\text{“Plastidial_ATP_Synthase_Light_p”}) + v(\text{“Mitochondrial_ATP_Synthase_Light_m”})) \times 3$$

where $v(\text{"Plastidial_ATP_Synthase_Light_p"})$ and $v(\text{"Mitochondrial_ATP_Synthase_Light_m"})$ are fluxes through the plastidial and the mitochondrial ATP synthases in the light, respectively. Since the stoichiometries of the two reactions were defined as the production of 3 ATP per flux, the ATP requirement was calculated by multiplying the flux value by a factor of 3. NADPH (reductant) requirement was calculated as the difference between reductant produced by photosystem I and the reductant oxidised by various routes of alternative electron flow, including cyclic electron transport, the water-water cycle and the mitochondrial electron transport chain, in the light. The reductant requirement was calculated in terms of NADPH (two electrons) with the following equation:

NADPH requirement in the light

$$= (v(\text{"RXN490-3650_Light_p"}) \div 2) - (v(\text{"Ferredoxin_Plastoquinone_Reductase_Light_p"}) + v(\text{"Mehler_Reaction_Light_p"}) \div 2 + v(\text{"CYTOCHROME-C-OXIDASE-RXN_Light_m"}) \times 2)$$

where $v(\text{"RXN490-3650_Light_p"})$, $v(\text{"Ferredoxin_Plastoquinone_Reductase_Light_p"})$, $v(\text{"Mehler_Reaction_Light_p"})$ and $v(\text{"CYTOCHROME-C-OXIDASE-RXN_Light_m"})$ are fluxes through photosystem I, ferredoxin plastoquinone reductase, Mehler reaction and mitochondrial cytochrome oxidase in the light, respectively. Reaction fluxes were scaled to per two electrons. Reductant dissipated by reductant-dissipating reactions was assumed to be negligible

under the objective function of flux minimisation with unconstrained photon input from the inspection of the flux solutions (Appendix S5).

3 Reconstruction of a high-quality Arabidopsis genome-scale metabolic model

3.1 Introduction

3.1.1 Structural metabolic modelling

Given the complexity of the plant metabolic network with hundreds or thousands of reactions occurring inside a cell, a computational modelling approach is required to investigate the behaviour of the whole metabolic system. There are two main types of metabolic modelling approaches, kinetic modelling and structural modelling. Kinetic modelling aims to investigate the dynamic behaviour and control of the metabolic system such as response to perturbations and flux control by each enzyme. While kinetic modelling is useful in detailed investigations of a selected section of the metabolic network, it requires many enzyme kinetic parameters which are difficult to obtain for many enzymes, hence hindering its use in analysing metabolic interactions over the whole metabolic network. On the other hand, structural modelling only requires knowledge of the stoichiometry of the metabolic reactions which are known for most reactions in the metabolic network, therefore it is well suited for large scale analysis of the metabolic network. Without any kinetic information, structural modelling assumes the metabolic system to be in a pseudo-steady state, so structural modelling can be used to investigate the steady-state characteristics of a metabolic system.

A structural metabolic model consists solely of the stoichiometry of the metabolic reactions without any kinetic parameters. Traditionally, structural metabolic models are built manually reaction by reaction based on existing

knowledge of the metabolic network in the organism of interest. The content and size of hand-built metabolic models vary greatly depending on the particular questions the model was built to answer. For example, elementary modes analysis was applied to a model of chloroplast stroma with 23 reactions to explore the metabolic flux modes in the light and the dark (Poolman et al., 2003), whereas a model developing oilseed rape embryos with 572 reactions and 10 different subcellular compartments was used to predict metabolic fluxes and investigate metabolic redundancy under four nutritional conditions (Hay and Schwender, 2011b).

With the availability of genome sequences and annotations, metabolic models can be built based on the information in the genome and such metabolic models are called genome-scale metabolic models. A genome-scale metabolic model contains the stoichiometry of metabolic reactions catalysed by all metabolic enzymes encoded in the genome of an organism of interest. Genome-scale metabolic models are advanced functional annotations (de Oliveira Dal'Molin and Nielsen, 2013) with a wide range of applications including contextualization of high-throughput data, guidance of metabolic engineering, directing hypothesis-driven discovery, investigation of metabolic interactions and discovering network properties (Oberhardt et al., 2009). One of the advantages of genome-scale metabolic models over hand-built models is that genome-scale models have larger coverage of the metabolic network. In addition, the automated process of model construction based on genomic data reduces the bias in excluding reactions which are not well studied. On the other hand, reactions that do not occur in the organism of interest could potentially be included in the model based on homology,

hence the need for manual curation for the construction of a high-quality model. For the model plant *Arabidopsis*, six genome-scale metabolic models have been published (Poolman et al., 2009; de Oliveira Dal'Molin et al., 2010a; Radrich et al., 2010; Saha et al., 2011; Chung et al., 2012; Mintz-Oron et al., 2012), and each has its shortcomings, especially in terms of subcellular compartmentation. In this chapter, the aim was to reconstruct a high-quality *Arabidopsis* genome-scale metabolic model suitable for modelling any tissue or cell type, with emphasis on subcellular compartmentation and manual curation.

3.1.2 Reconstruction of genome-scale metabolic models

Reconstruction of genome-scale metabolic models has been developed over the past 14 years and has become a common practice for investigating genotype-phenotype relationships. There are an increasing number of online resources for building genome-scale metabolic models including genomic and biochemical databases, such as KEGG ([http:// www.genome.jp/kegg/](http://www.genome.jp/kegg/); Kanehisa et al., 2012), MetaCyc ([http:// http://metacyc.org](http://metacyc.org); Caspi et al., 2008) and the SEED database (<http://www.theseed.org>; Overbeek et al., 2005), and software packages for model reconstruction such as the COBRA Toolbox (Schellenberger et al., 2011), GEMSiRV (Liao et al., 2012) and the RAVEN toolbox (Agren et al., 2013). A list of online resources frequently used in metabolic reconstructions can be found in a published metabolic reconstruction protocol (Thiele and Palsson, 2010).

The first step in building a genome-scale metabolic model is the automated construction of a draft model from genomic and biochemical databases. The

data source for metabolic reconstruction determines the coverage and the quality of the draft model. The five published Arabidopsis models were constructed based on the information in KEGG and/or AraCyc (Mueller et al., 2003), an Arabidopsis-specific version of MetaCyc. In the KEGG database, the annotated Arabidopsis genes for metabolic enzymes are linked to metabolic reactions on a global metabolic map. While the structure of the KEGG database is well suited for comparative studies between organisms, the amount of curation for Arabidopsis related data is less compared to AraCyc. AraCyc is a highly curated Arabidopsis-specific database with large number of experimentally supported enzymes and pathways, together with computationally predicted enzymes and pathways followed by curator review. For the purpose of constructing a high-quality Arabidopsis genome-scale metabolic model, AraCyc was chosen as the data source for building the draft model. Radrich et al., (2010) and Mintz-Oron et al., (2012) extracted data from both KEGG and AraCyc to construct their draft model. However, information from KEGG was not used for model reconstruction in this study because the integration of the two databases can be problematic and error-prone. Also, since the Arabidopsis specific data on KEGG is not as well curated as AraCyc, the inclusion of reactions extracted from KEGG increases the chance of adding non-existent reactions to the model.

After the initial automated reconstruction of a draft model, manual curation is required to correct the errors and inconsistencies which stem from a number of features in the database including non-unique metabolite identifiers, metabolites as both substrate and product in a given reaction and atomically-incorrect reactions (Poolman et al., 2006). To convert a draft model into a

functional model, the boundaries (inputs and outputs) of the metabolic system must be defined. Then the model must be tested for the ability to carry out known metabolic functions of the organism / cell of interest. In most genome-scale metabolic models, biomass synthesis from a defined set of nutrients is the major metabolic function. Arabidopsis being a multi-cellular organism containing both phototrophic and heterotrophic tissues, the genome-scale metabolic model was tested for biomass synthesis in both phototrophic and heterotrophic conditions.

3.1.3 Compartmentation in models of eukaryotic organisms

A major challenge in constructing metabolic models of eukaryotes compared to bacteria is the presence of subcellular compartments. This is particularly relevant for plants as large number of reactions occur in organelles, with metabolic pathways such as the photorespiratory pathway spanning multiple subcellular compartments (Sweetlove and Fernie, 2013). The importance of compartmentation in flux balanced metabolic models was shown by investigating the effect of removing compartmentation in a yeast metabolic model which contained eight compartments (Klitgord and Segrè, 2010). It was shown that the growth rate and flux predictions from a de-compartmentalised model were significantly different from those obtained from a compartmented model, mainly due to energetic problems caused by the removal of gradients across membranes.

Compartmentation of plant metabolic models is challenging as it requires knowledge of the subcellular localisation of metabolic enzymes and the presence and identity of metabolite transporters, both of which are

incomplete even in Arabidopsis (de Oliveira Dal'Molin and Nielsen, 2013). Although there are databases such as SUBA (Heazlewood et al., 2007; Tanz et al., 2013), which contain information of subcellular localisation of proteins based on experimental evidence, they cover only a relatively small part of the metabolic network (Sweetlove and Ratcliffe, 2011). In four of the six published Arabidopsis models, Poolman model (Poolman et al., 2009), AraGEM (de Oliveira Dal'Molin et al., 2010a), *i*RS1597 (Saha et al., 2011) and *i*AT1475 (Chung et al., 2012), the assignment of metabolic reactions to subcellular compartments and the addition of transmembrane metabolite transport reactions were done manually. Note that the model presented in Radrich et al., (2010) did not include any subcellular compartmentation.

An alternative to manual compartmentation of metabolic models was presented which involves a constraint-based algorithm that predicts enzyme subcellular localisation based on a subset of pre-defined (known) enzyme localisations and a parsimony principle of minimising the number of intracellular transport reactions to activate the enzymes (Mintz-Oron et al., 2009). Mintz-Oron et al., (2012) applied this algorithm to build a compartmented Arabidopsis model, using the known localisation of 49% of the reactions taken from the SUBA database as the pre-defined subset of known enzyme localisation. Two criticisms on this approach were raised in a review paper by de Oliveira Dal'Molin and Nielsen, (2013); firstly, the pre-defined subset of enzyme localisation based on the SUBA database could be highly contaminated by false positive predictions with abundant gene products in multiple compartments; secondly, the parsimony principle of minimising the number of intracellular transporters may not be justifiable.

With the goal of producing a high-quality Arabidopsis genome-scale metabolic model in mind, an updated Arabidopsis model was manually compartmented based on a comprehensive literature review and cross-checking with the SUBA database to assign most metabolic reactions involved in biosynthesis and central carbon metabolism to appropriate subcellular compartments. Intracellular transporters were also manually added to the model with priority given to known metabolite transporters supported by experimental evidence.

3.2 Results

3.2.1 Construction of a draft genome-scale metabolic model from AraCyc database

3.2.1.1 Extracting the list of metabolic reactions from AraCyc database into ScrumPy format

The first step in reconstructing a genome-scale metabolic model of Arabidopsis was to extract a list of metabolic reactions from the metabolic database AraCyc (Mueller et al., 2003). The model was reconstructed based on information in AraCyc 9.0 which was released in March 2012 and the database contains 3320 reactions and, coincidentally, 3320 metabolites. For each entry in the “reactions.dat” file of the AraCyc database (Figure 3.1A), the following information was extracted: the unique reaction identifier (“UNIQUE-ID”), reaction direction (“REACTION-DIRECTION”) and the metabolites (“LEFT” and “RIGHT”) with their corresponding stoichiometry (“^COEFFICIENT”). In the “REACTION-DIRECTION” field of each reaction, there are 3 options: “REVERSIBLE” which represents a reversible reaction, “LEFT-TO-RIGHT” which represents an irreversible reaction from

metabolite(s) in the field “LEFT” to metabolite(s) in the field “RIGHT” and “RIGHT-TO-LEFT” which represents an irreversible reaction from metabolite(s) in the field “RIGHT” to metabolite(s) in the field “LEFT”.

Each reaction was converted and outputted in a format compatible with the modelling software, ScrumPy (Poolman, 2006), with the reaction name being the entry in the field “UNIQUE-ID” (Figure 3.1B). A colon denotes the end of reaction name in ScrumPy format. Metabolite(s) in the field “LEFT” were taken as substrate(s) and metabolite(s) in the field “RIGHT” were taken as product(s) by default. For reactions with “RIGHT-TO-LEFT” in the “REACTION-DIRECTION” field, metabolite(s) in the field “RIGHT” were taken as the substrate(s) and metabolite(s) in the field “LEFT” were taken as the product(s). In ScrumPy format, reversible reactions and irreversible reactions are denoted by “<>” and “->” respectively. Unless there is a “^COEFFICIENT” field associated with a metabolite, in which case the entry in the “^COEFFICIENT” field was used as the stoichiometry of the metabolite, the default stoichiometry of each metabolite was one. For a structural metabolic model, the symbol “~” denotes the end of a reaction in ScrumPy format.

3.2.1.2 Resolving problems with ill-defined metabolites in the AraCyc database

Each metabolite involved in one or more reaction(s) in AraCyc is either a “compound” or a “class”. A “compound” is associated with an entry in the “compounds.dat” file in the AraCyc database, where each entry contains the chemical formula of the compound. A “class” is associated with an entry in “classes.dat” file, which generally represents a group of related compounds

and does not have an associated chemical formula (e.g. “Peptides”, “Acceptor”, “a-long-chain-alcohol”, “ALA-tRNAs” etc.).

“Classes” are problematic in stoichiometric modelling because they usually represent a group of metabolites and it is difficult to relate “classes” with other “compounds” in the model. All “classes” were initially included in a list of metabolites to be removed from the model (“Unwanted.py” in Appendix S1). The list of “classes” was gone through manually to identify metabolites which should not be excluded from the model. These included various electron carriers (e.g. ferredoxin, thioredoxin, cytochromes etc.), acyl-carrier-proteins (for fatty-acid metabolism), carbohydrate polymers (e.g. cellulose, starch etc.), amino-acid-specific tRNAs etc.

While metabolites in the “compound” category have associated chemical formulae, some of these “compounds” contain R-groups and hence have the same problem as “classes” described above. Metabolites with chemical formulae containing R-groups were also included in the list of metabolites to be removed from the model (“Unwanted.py” in Appendix S1). All reactions involving metabolites in “Unwanted.py” were removed from the model.

Another problem with the definition of metabolites in the AraCyc database for stoichiometric model construction is that multiple names are used for the same metabolite. For example, “D-glucose-6-phosphate”, “ALPHA-GLC-6-P” and “GLC-6-P” all represent glucose 6-phosphate. For each metabolite with multiple names, one name was arbitrarily selected to be used in the model and the other names of the same metabolite were substituted by the selected

name. A full list of substituted reaction names can be found in “Substitutes.py” in Appendix S1.

There are reactions in the model containing the metabolites “NAD-P-OR-NOP” and “NADH-P-OR-NOP” which represent NAD(P)^+ and NAD(P)H respectively. This means that reaction can either use NAD^+ or NADP^+ as an electron acceptor with the corresponding NADH or NADPH as an electron donor. For the purpose of model construction, each reaction containing “NAD-P-OR-NOP” and “NADH-P-OR-NOP” was split into two reactions, one reaction uses NAD^+ and NADH with “-(NAD)” appended at the end of the reaction name and the other reaction uses NADP^+ and NADPH with “-(NAD)” and “-(NADP)” added to at the end of the reaction name. Finally, the character “|” was removed from all metabolite names for software compatibility.

3.2.1.3 Removal of reactions and corrections of reaction stoichiometry

In structural metabolic modelling, isostoichiometric reactions (reactions with identical stoichiometry) can cause problems such as cycles and alternative routes. Some isostoichiometric reactions were initially present in the database, whereas some were introduced by substitutions of metabolite names described above. For each set of isostoichiometric reactions, the reaction with the most readable reaction name was selected to be used in the model and other isostoichiometric reaction(s) were removed from the model (see “Unwanted.py” in Appendix S1).

For reactions not involving metabolites in the “class” category, the atomic compositions of the substrates and the products were extracted from the

“compounds.dat” file in the AraCyc database and compared to check for atomic imbalances between the substrates and products of each reaction. The stoichiometries of atomically-imbalanced reactions were manually corrected as far as possible. For reactions that could be corrected manually, the original reactions were removed from the model and the corrected reactions (contained in the file “Corrections.spy” in Appendix S1) were added to the model. Except for imbalances in hydrogen and oxygen, as protons and water are not necessarily balanced in AraCyc reactions, atomically-imbalanced reactions that cannot be corrected manually were removed from the model (see “Unwanted.py” in Appendix S1). Reactions involving metabolites in the “class” category were manually checked for correctness of reaction stoichiometry. These include reactions involving starch, cellulose and xylan, with starch and cellulose defined to have stoichiometry equivalent to one hexose subunit and xylan equivalent to one pentose subunit.

In AraCyc, some redox reactions were defined to use NAD(P)^+ and NAD(P)H as the electron acceptor and donor. By default, these reactions were split into two independent reactions as described previously. The coenzyme specificities of some of these reactions were defined in the model based on information in the plant literature. A notable example was 6-phosphogluconate dehydrogenase, which was defined to use NADP^+ but not NAD^+ as the electron acceptor. Also based on information from the primary literature, reactions that are present in the model but not known to be present *in planta* were removed from the model. Notable examples include pyruvate carboxylase and NAD(P)^+ transhydrogenase. The list of removed reactions can be found in “Unwanted.py” in Appendix S1.

3.2.2 Conversion of a draft model into a functional metabolic model

3.2.2.1 Definition of inputs and outputs reactions

A draft model constructed from the information in the AraCyc database contains metabolic reactions which take place inside the cell. In order to run any model simulations, the draft model must first be converted into a functional model by defining the inputs and outputs of the system. Typically, the inputs of a metabolic model are the nutrients required for biomass synthesis. In the Arabidopsis model, these include photon, glucose, sucrose, starch and inorganic ions of various elements (e.g. NO_3^- , NH_4^+ , SO_4^{2-} , Pi, etc). Reactions representing the exchange of CO_2 and O_2 were included in the model. The main outputs of the model are the individual biomass components including starch, cell wall (cellulose and xylan), lipids (palmitate and glycerol), nucleic acids, protein (20 amino-acids) and soluble metabolites (organic acids, amino-acids, sugars and inorganic ions). Reactions for synthesis of various coenzymes (e.g. NAD^+ , coenzyme A) and secondary metabolites (e.g. chlorophyll) were also included in the model. Reactions representing the synthesis of biomass components were denoted with reaction names ending with “_biomass”. In addition to biomass components, reactions were included to represent cellular maintenance costs in terms of ATP and NADPH.

3.2.2.2 Definition of the internal and external metabolites

In a structural metabolic model, metabolites can be separated into two categories, internal metabolites and external metabolites. Internal metabolites are metabolites that are not at the boundary of the metabolic

system (i.e. internal with respect to the metabolic network). In any analysis that assumes metabolic steady-state, such as flux balance analysis and elementary modes analysis, the rate of production of an internal metabolite must equal its rate of consumption (this is the flux-balance constraint). Internal metabolites include all metabolites extracted from AraCyc and other manually defined metabolites which are confined by the flux-balance constraint.

External metabolites define the boundary of the metabolic system and they are not confined by the flux balance constraint (i.e. rate of production does not necessarily equal to the rate of consumption). These include nutrients and metabolites in the external environment, accumulation (sink) of biomass components and dummy metabolites representing ATP and NADPH cellular maintenance costs. In ScrumPy format, metabolites with names beginning with “x_” are defined as external metabolites. For consistency, external metabolites are defined to be on the left hand side of all reactions involving external metabolites and these reactions were all defined as reversible. Specific constraints on directionality of reactions involving external metabolites were set during model analysis. In theory, for the purpose of flux balance analysis, it is not necessary to include external metabolites in the model as they are not used in the formulation of the optimisation problem.

In addition to the external metabolites listed above, proton and water were set as external for the purpose of modelling. This is because there are imbalances of proton and water in the stoichiometry of reactions in AraCyc. Given that the balancing of proton and water were not of interest in any of the analysis in this study, it is justifiable to set proton and water as external to

allow proton and water to enter or escape from the metabolic system as necessary (Poolman et al., 2009).

3.2.2.3 Model testing for metabolic functions

In order for a model to be functional, the model must be able to carry out known metabolic functions of the organism/cell of interest. For the Arabidopsis model, the ability to grow (synthesise biomass components) under heterotrophic and phototrophic conditions was tested. The synthesis of each biomass component was tested individually with NO_3^- and NH_4^+ as the nitrogen source, SO_4^{2-} as the sulphur source and Pi as the phosphorus source. In heterotrophic condition, glucose was used as the carbon and energy source whereas photons were used as the energy source and CO_2 as the carbon source in phototrophic condition. For each biomass component that could not be synthesised, reactions in the corresponding biosynthesis pathway were investigated to check for reaction direction, coenzymes used and by-products produced from the biosynthesis pathway. Inability to recycle coenzymes could be one of the reasons for the failure to produce a biomass component. Reactions for recycling coenzymes were manually checked for their presence in the model and for reaction direction. Another possible reason for the inability to produce biomass component is that the synthesis of the biomass component is coupled with the production of by-product(s) which cannot be metabolised or exported from the model. Reactions for metabolism of these by-products were manually checked. The process was done until all biomass components could be produced in both heterotrophic and phototrophic conditions. Changes in stoichiometry, direction and reversibility

of reactions can be found in “Corrections.spy” and “Substitutes.py” (Appendix S1).

3.2.3 Compartmentation of the Arabidopsis genome-scale metabolic model

3.2.3.1 Assigning reactions to different subcellular compartments

One of the complications in genome-scale metabolic modelling of eukaryotes compared to bacteria is the presence of subcellular compartments. To capture this in the model, metabolic reactions were assigned to five subcellular compartments, cytosol, mitochondrion, plastid, peroxisome and vacuole, mainly based on the information the literature. The SUBA database (Heazlewood et al., 2007; Tanz et al., 2013), a database for subcellular localisation of Arabidopsis proteins, was also used as a source of information for cross-checking with the primary literature. Reactions were assumed to be in the cytosol if there was no information on enzyme localisation.

In modelling terms, subcellular compartments are represented by the separation of metabolite pools and localisation of metabolic reactions. For each reaction and its associated metabolites, the localisation were defined by adding suffixes after the reaction and metabolite names, “_c” for cytosol, “_m” for mitochondrion, “_p” for plastid, “_x” for peroxisome and “_v” for vacuole.

3.2.3.2 Addition of metabolite transport reactions between subcellular compartments

After reactions were assigned to the five subcellular compartments, the metabolic network was split into five disconnected sub-networks with separate pools of metabolites. The model was unable to produce some of the biomass components after this segregation of the metabolic network. In the

cell, metabolites pools are connected by intracellular metabolite transporters which transport metabolites between subcellular compartments. In the model, metabolite transport reactions were added to allow metabolites to be transported from one compartment to another. As a starting point, the metabolite transporters described in Linka and Weber, (2010) were added to the model. An iterative process of manual addition of metabolite transport reactions and testing for synthesis of biomass components was carried out until all biomass components could be produced in heterotrophic and phototrophic conditions. The metabolite transport reactions were defined with the suffixes “_mc”, “_pc”, “_xc” and “_vc” corresponding to mitochondrial, plastidial, peroxisomal and tonoplast transporters.

3.2.4 Definition of proton gradients across membranes, metabolite transport costs, and energetic costs for other cellular processes in the model

3.2.4.1 Definition of proton gradients across membranes

The establishment of proton gradients across membranes is crucial for the functioning of the cell. The uptake of some nutrients is driven by the proton gradient as many transporters for nutrients uptake are coupled with proton co-transport (Bush, 1993). Another important process that requires the establishment of a proton gradient is the synthesis of ATP by ATP synthase in the plasma membrane of prokaryotes, the inner mitochondrial membrane of eukaryotic cells and the thylakoid membrane in chloroplast (Müller and Grüber, 2003; von Ballmoos et al., 2009). Proton gradients are established by pumping protons across a membrane against the electrochemical gradient in an energy-requiring process. At the plasma membrane and the tonoplast,

proton-pumping ATPases generate proton gradients for driving metabolite transport (Gaxiola et al., 2007). There is also a proton-pumping pyrophosphatase on the tonoplast which uses the hydrolysis of pyrophosphate as an energy source (Maeshima, 2000). In the thylakoid membrane and the inner mitochondrial membrane, the transfer of electrons along the electron transport chains is coupled with proton pumping across membranes.

Ideally in a proton- and charged-balanced flux balance model, the generation and dissipation of proton gradients can be modelled simply as the transport of protons from one compartment to another. However, given that the Arabidopsis model is not balanced with respect to protons and that protons were set as an external metabolite (i.e. protons can freely enter and exit the system), it is problematic to define the proton gradient across membrane. To overcome this problem, an additional metabolite, “Pumped-PROTON”, was introduced to represent the protons pumped across each membrane.

“Pumped-PROTON” was defined in four different compartments with suffixes “_c”, “_m”, “_p” and “_v” representing the proton gradients across the plasma membrane, inner mitochondrial membrane, thylakoid membrane and the tonoplast respectively. Reactions producing “Pumped-PROTON” represent the generation of proton gradient whereas reactions consuming “Pumped-PROTON” dissipate the proton gradient. Although the model is not balanced in terms of protons, the generation and dissipation of proton gradients (represented by the metabolite “Pumped-PROTON”) are balanced in the model (i.e. there is no net movement “Pumped-PROTON” across membranes, hence the proton gradients remain constant at steady-state).

3.2.4.2 Definition of metabolite transport costs

The transport of many metabolites across membranes is energy-dependent which means the costs for metabolite transport contribute to the overall energy budget of the cell. In a plant cell, most energy-dependent metabolite transport is driven by proton gradients across membranes. In the model, this is defined by a coupled reaction of metabolite transport and consumption of “Pumped-PROTON” representing the dissipation of proton gradient with metabolite transport. Another type of energy-dependent metabolite transport is driven by the hydrolysis of ATP. For example, the transport of fatty-acid across the peroxisomal membrane is coupled with ATP hydrolysis by an ATP-binding cassette transporter (Theodoulou and Eastmond, 2012). Energy-dependent metabolite transport between four membranes, plasma membrane, tonoplast, mitochondrial membrane and peroxisomal membrane, were define in the model (see Appendices 3 and S6 for details on the energy-use stoichiometries of metabolite transporters).

3.2.4.3 Energetic costs for other cellular processes

In addition to the energetic costs for synthesising individual biomass components and for metabolite transport, there are other growth-dependent energetic costs including the cost for protein polymerisation, processing of protein and nucleic acids and protein translocation. To define protein associated costs in the model, amino-acids in proteins have to be separated from soluble amino-acids. This was done by duplicating the biomass output reactions for amino-acids and “p” and “s” were added to the front of the reaction names representing amino-acids in protein and soluble amino-acids respectively. For clarification, “s” was added to the front of the reaction

names of all biomass output reactions of soluble metabolites. For protein polymerisation cost, the hydrolysis of two molecules of GTP was coupled with the addition of one amino-acid group onto the polypeptide chain from one aminoacyl- tRNA. In terms of the processing of protein, it includes mRNA synthesis, proofreading, assembly and modification with an estimated cost of 0.306 ATP per amino-acid (Sheikh et al., 2005). RNA processing involves modification and discarding segments which costs 0.4 ATP per nucleotide (Sheikh et al., 2005). DNA processing includes unwinding of the double helix, proofreading, discontinuous synthesis, negative supercoiling and methylation with a total cost of 1.372 ATP per nucleotide (Sheikh et al., 2005). Finally, cost for protein translocation was calculated based on an estimate of 1000 ATP per proOmpA translocated by the Sec pathway (Alder and Theg, 2003), which was translated to 2.89 ATP per amino-acid translocated. Fifty percent of the protein was assumed to be translocated into organelles (Mintz-Oron et al., 2012), meaning that 1.445 ATP is required for protein translocation per amino-acid in protein produced in the biomass output reaction (Appendices 4 and S5). All costs stated above were included in the corresponding biomass output reactions.

In addition to growth-dependent processes, there are energetic costs associated with non-growth-dependent processes such as turnover of macromolecules, overcoming membrane leakage and protection against reactive oxygen species. Here, the costs for non-growth-dependent processes were defined as cellular maintenance costs. Two types of cellular maintenance costs were considered, ATP and NADPH. To model ATP maintenance costs, a generic ATPase reaction was included in cytosol with

the assumption that all ATP maintenance cost is in the cytosol. For NADPH maintenance costs, generic NADPH oxidation reactions were added in the cytosol, mitochondrion, plastid and peroxisome.

3.2.5 Identification of errors and inconsistencies in the structure of the model

3.2.5.1 Testing for stoichiometric inconsistency in the model

The reconstruction of a metabolic model from metabolic databases is susceptible to errors inherited from the databases themselves. Incorrect definition of reaction stoichiometry can lead to the violation of mass conservation. An algorithm based on mixed-integer linear programming (Gevorgyan et al., 2008) was used to identify unconserved metabolites, metabolites whose mass is not conserved, in the Arabidopsis genome-scale metabolic model. The list of unconserved metabolites revealed by this algorithm only contains metabolites with no mass (photon and dummy metabolites representing various energetic costs) and metabolites which consist solely of hydrogen and/or oxygen, e.g. proton, water, oxygen, hydrogen peroxide etc. (Table 3.1). This suggests that the model is consistent in terms of elements other than hydrogen and oxygen. To overcome the problem of imbalances in hydrogen and oxygen in the metabolic system, proton and water were set as external as described in section 3.2.1.3.

3.2.5.2 Identification and removal of internal reaction cycles producing ATP and NAD(P)H

While the violation of mass conservation can be identified by testing for stoichiometric inconsistency in the model, errors in terms of thermodynamics

(reaction direction and reversibility) cannot be detected by looking at mass conservation. While futile cycles with a net stoichiometry of hydrolysing ATP to form ADP and Pi can occur *in vivo*, these cycles cannot run in reverse and convert ADP and Pi to form ATP due to thermodynamic constraints. However, such ATP-producing cycles can exist in the model if there are errors in defining the direction and reversibility of reactions. To identify ATP-producing cycles in the model, a generic ATPase reaction was temporarily added to the model in one particular subcellular compartment and a linear programming problem was solved with the generic ATPase reaction set to carry a non-zero flux (e.g. one arbitrary unit) in the direction of ATP consumption and all other input and output reactions to carry zero flux (i.e. forcing the model to produce ATP without any inputs and outputs). If a feasible solution was found, the direction and reversibility of the reactions in the solution were manually curated to remove the thermodynamically unfavourable cycle. The iterative process of identifying ATP-producing cycles and manual curation of reaction direction and reversibility were carried out until no ATP-producing cycles could be identified (i.e. no feasible solution exists for the linear programming problem stated above). The process was done for all five subcellular compartments independently (i.e. testing for ATP-producing cycle in such compartment one at a time).

A similar iterative process was carried out for NADH and NADPH. A generic NADH oxidase or NADPH oxidase was temporarily added to each subcellular compartment independently and it was set to carry a non-zero flux in the direction of NAD(P)H oxidation. Except for oxygen, which is required for NAD(P)H oxidation, all other inputs and output reaction fluxes were set to

zero. With these constraints, the linear programming problem was solved to identify NAD(P)H-producing cycles. Reaction direction and reversibility were manually curated to remove any NAD(P)H-producing cycles.

3.2.5.3 Removal of transhydrogenase cycles transferring reducing power from NADH to NADP⁺

In the model, there are reaction cycles that convert NADH and NADP⁺ to NAD⁺ and NADPH and *vice versa*. These are termed transhydrogenase cycles as they have a net stoichiometry of a transhydrogenase reaction. However, in plant cells, the NADP(H) pool is more reduced than the NAD(H) pool in both the light and the dark (Heineke et al., 1991; Igamberdiev and Gardeström, 2003), hence it is thermodynamically implausible for reducing power to be transferred from NADH to NADP⁺ without any energy input. A proton-translocating transhydrogenase can transfer reducing power from NADH to NADP⁺ using the energy from proton gradients across membranes (Pedersen et al., 2008), but although such a protein exists in most prokaryotes and in animal mitochondria, the Arabidopsis genome does not appear to contain the relevant gene (Arkblad et al., 2001). Therefore transhydrogenase cycles running in the direction of NADPH production were identified using a similar approach to that used to identify NAD(P)H-producing cycles in section 3.2.5.2. Instead of adding an NAD(P)H oxidase reaction, a transhydrogenase reaction in the direction of NADPH consumption was temporarily added in each of the five compartments independently and set to carry a non-zero flux. With all inputs and outputs including oxygen set to zero flux, linear programming was used to identify transhydrogenase cycles in the direction of NADPH production. The reaction

direction and reversibility were manually curated to remove all transhydrogenase cycles in the direction of NADH production. Notable examples include plastidic NADP-dependent glyceraldehyde 3-phosphate dehydrogenase and NADP-dependent malate dehydrogenase which are described in detail in chapter 4.

3.2.6 Model statistics

The Arabidopsis genome-scale metabolic model reconstructed based on AraCyc 9.0 contains 2769 reactions, with 2039 in the cytosol, 292 in the plastid, 97 in the mitochondrion, 36 in the peroxisome, 1 in the vacuole, 192 intracellular transport reactions and 112 other reactions including nutrient uptake, biomass production and maintenance reactions. 1860 reactions are associated with at least one gene-product and a total of 2857 genes are linked to reactions in the model. With the defined set of inputs and outputs, 1081 reactions can carry a non-zero flux whereas 1688 cannot carry a non-zero flux (dead/blocked reactions), computed by flux variability analysis with no model constraints and a primary objective function. In terms of metabolites, the model contains 2739 metabolites with 2146 in the cytosol, 282 in the plastid, 127 in the mitochondrion, 55 in the peroxisome, 18 in the vacuole and 121 external metabolites other than proton and water. With the defined set of inputs and outputs, 872 metabolites are involved in reactions that can carry a non-zero flux and 1867 metabolites are only involved in blocked reactions.

3.3 Discussion

3.3.1 Improvements and limitations of the updated Arabidopsis genome-scale metabolic model

An updated version of the Arabidopsis genome-scale metabolic model was constructed based on information in AraCyc 9.0, with five subcellular compartments and 2769 reactions in total. Apart from the extensive compartmentation, which is discussed in section 3.3.2, there are five major improvements in the updated model compared to the previous AraCyc-based model (Poolman et al., 2009). (1) The updated model was constructed based on a more recent update of the AraCyc database, AraCyc 9.0 compared to AraCyc 4.5 used in the Poolman model. This update resulted in a larger coverage of the metabolic network in the updated model with more than 2000 reactions in the cytosol alone, compared to 1406 reactions in the whole Poolman model. (2) The information on reaction direction and reversibility of reactions in AraCyc were used during the reconstruction of the draft model in this study but not in the Poolman model. The Poolman model assumed all reactions to be irreversible from left to right by default with a subset of reactions changed to reversible to enable biomass synthesis. However, this approach does not allow the definition of irreversible reactions from right to left. A notable error in reaction direction in the Poolman model includes the NADP-dependent glyceraldehyde 3-phosphate dehydrogenase, which is investigated and discussed in chapter 4. (3) Many reactions involving ill-defined metabolites were manually included in the updated model but not in the Poolman model, such as fatty-acids synthesis and aminoacyl-tRNA charging reactions. (4) The biosynthesis of several coenzymes and

secondary metabolites were curated and included in the updated model including NAD^+ , ascorbate, glutathione, coenzyme-A and chlorophyll etc. (5) The updated model can simulate photosynthetic metabolism with the inclusion of light reactions which were absent in the Poolman model.

The major limitation with this updated Arabidopsis model, as with the Poolman model, is that the model is not balanced with respect to hydrogen and oxygen. This was overcome by setting proton and water as external, which means the balancing of protons and water cannot be investigated with the model. The balancing of protons in eukaryotic genome-scale metabolic models is problematic because various subcellular compartments have different pHs. As a result, some compounds are in different protonation states in different subcellular compartments. Although there are elementary and charge balanced genome-scale metabolic models, e.g. two recent Arabidopsis models (Saha et al., 2011; Mintz-Oron et al., 2012), they assumed the constant pH throughout all subcellular compartments, which is not an accurate representation of the state of the cell. Therefore such models cannot be used to accurately predict the intracellular movements of protons. Another limitation with the model developed in this study is the large number of blocked reactions. This is mainly the consequence of the limited sets of inputs and outputs defined in the model. For the purpose of model simulations in this study, which concentrates on primary metabolism, the outputs for most secondary metabolites were not included in the model. However, the model can be used as a tool for investigating specific secondary metabolism after adding the corresponding output reactions and manually curating the biosynthetic reactions for secondary metabolites in the

model. This was done for terpenoid metabolism by manually curating the corresponding reactions in an existing Arabidopsis genome-scale metabolic model, *i*RS1597 (Saha et al., 2011), to produce a new model, *i*AT1475, for investigating terpenoid metabolism (Chung et al., 2012).

3.3.2 Compartmentation in plants genome-scale metabolic models

Reactions in the Arabidopsis genome-scale metabolic model built in this study were manually assigned to five subcellular compartments, cytosol, mitochondrion, plastid, peroxisome and vacuole, with the cytosol being the default compartment. There are 426 metabolic reactions in the four defined organelles. This was a significant improvement from the previous AraCyc-based model, the Poolman model (Poolman et al., 2009), with only two compartments, cytosol and mitochondrion, which was sufficient for modelling metabolism of heterotrophic cell culture. Two published Arabidopsis genome-scale metabolic models based on KEGG, AraGEM (de Oliveira Dal'Molin et al., 2010a) and *i*RS1597 (Saha et al., 2011), contain the same five compartments as the model in this study, but only 172 and 169 organellar reactions respectively. The compartmentation of many biosynthesis-related reactions was incomplete in the two KEGG-based models. For example, most reactions for amino-acid synthesis were assigned to be in the cytosol, while it is well known that many of the reactions are localised in the plastid (Sweetlove and Ratcliffe, 2011). Another published Arabidopsis model curated for terpenoid metabolism, *i*AT1475 (Chung et al., 2012), was built based on *i*RS1597 (Saha et al., 2011), but interestingly *i*AT1475 only contains four compartments with the vacuole removed from the model.

*i*AT1475 shares similar compartmentation information with AraGEM and *i*RS1597, except for reactions in terpenoid metabolism, some of which were manually assigned to the plastid which makes a total of 281 reactions in organelles.

In the most recently published Arabidopsis genome-scale metabolic model (Mintz-Oron et al., 2012), a constraint-based algorithm was used to predict enzyme subcellular localisation based on a parsimony principle of minimising the number of intracellular transport reactions to activate the enzyme (Mintz-Oron et al., 2009), using the enzyme localisation information from the SUBA database (Heazlewood et al., 2007; Tanz et al., 2013) of 49% of the reactions in the model. The resulting model has seven subcellular compartments, cytosol, mitochondrion, plastid, peroxisome, vacuole, endoplasmic reticulum and Golgi, and contains more than 1700 reactions in organelles. Despite minimising the number of intracellular transporters, 772 intracellular transport reactions were included in the model. However, the accuracy of the model compartmentation from this approach is questionable (Seaver et al., 2012) with two major concerns including the high false positive predictions with abundant gene products in multiple compartments in the SUBA database and the applicability of the parsimony principle in minimising the number of intracellular transporters in the evolution of metabolic systems (de Oliveira Dal'Molin and Nielsen, 2013). With 36% of the metabolic reactions localised in two or more compartments and 772 intracellular transporters in the model, one would expect a high level of redundancy in the model. With the construction of highly compartmented models, questions were raised about the functional significance of subcellular compartmentation

and whether flux-balanced model can be used to resolve metabolic fluxes in subcellular compartments (Sweetlove and Ratcliffe, 2011).

For genome-scale metabolic models of other plants species, a rice genome-scale metabolic model was recently published with limited compartmentation, 15 and 23 metabolic reactions in the mitochondrion and the plastid respectively (Poolman et al., 2013). The compartmentation of models of C_4 plants, including maize, sorghum, sugarcane (de Oliveira Dal'Molin et al., 2010b) and another maize model (Saha et al., 2011), were all based on the information in an Arabidopsis genome-scale metabolic model, AraGEM (de Oliveira Dal'Molin et al., 2010a). Since Arabidopsis is the most studied plant species in terms of enzyme localisation, the compartmentation information in Arabidopsis genome-scale metabolic models will serve as the basis for the construction of other plant metabolic models. In this study, the Arabidopsis genome-scale metabolic model was extensively compartmented through manual curation and this information can be used for the construction of metabolic models of other plant species.

3.3.3 Energetic costs for transport and other cellular processes in plant flux-balance models

Energetic costs for biosynthesis-related processes, including metabolite transport, protein and nucleic acids polymerisation and processing and protein translocation, were included in the stoichiometry of the updated Arabidopsis genome-scale metabolic model in this study. The extensive inclusion of energetic costs for growth-associated processes was done in the mouse genome-scale metabolic model (Sheikh et al., 2005), but not in any plants models. For models of microorganisms such as *E. coli* (Feist et al.,

2007; Orth et al., 2011), an experimentally measured growth-associated energetic cost was used instead of extensively accounting for costs of individual processes. By accounting for growth-associated processes in the stoichiometry of the model, the energetic costs for non-growth associated processes, which are defined as maintenance in this study, can be calculated as the remainder of the energy produced by the metabolic system not used in growth-associated processes. To model cellular maintenance costs, generic ATPase and NADPH oxidases were included in the model to represent ATP and NADPH maintenance costs respectively.

Within the published plant flux-balance metabolic models, there is a large variation in terms of the extent to which the models account for energetic costs for cellular processes. None of the published plant genome-scale metabolic models includes metabolite transport costs in the stoichiometry of the reactions. The Poolman model contains a generic ATPase reaction to account for ATP costs for all cellular processes, both growth-dependent and growth-independent (Poolman et al., 2009). Three Arabidopsis genome-scale metabolic models (de Oliveira Dal'Molin et al., 2010a; Saha et al., 2011; Chung et al., 2012) and all C₄ models (de Oliveira Dal'Molin et al., 2010b; Saha et al., 2011) contain a fixed growth-dependent ATP cost in the biomass equation with no growth-independent maintenance costs. In two hand-built plant flux-balance models, barley seed (Grafahrend-Belau et al., 2009) and oilseed rape embryos (Hay and Schwender, 2011b), the energetic costs for metabolite uptake were included in the stoichiometry of the reactions. The oilseed rape embryos model also included costs for fatty-acid transport across the peroxisomal membrane. However, the two hand-built models

treated energetic costs for other processes differently. The barley seed model includes a fixed growth-dependent ATP cost in the biomass reaction and assumed the growth-independent costs as negligible (Grafahrend-Belau et al., 2009). The oilseed rape model includes macromolecule polymerisation costs and protein processing costs in the stoichiometry of reactions, and a generic ATPase reaction was included to model growth-independent maintenance cost (Hay and Schwender, 2011b). By including the growth-dependent energetic costs in the stoichiometry of the reactions as far as possible and defining reactions for growth-independent ATP and NADPH maintenance costs, the model developed in this study has made a large step forward in terms of accounting for energetic costs in plant flux-balance metabolic models.

3.3.4 Summary

The aim of the chapter was to build a high-quality genome-scale metabolic model of Arabidopsis, with emphasis on subcellular compartmentation and accounting of energetic costs. The Arabidopsis genome-scale metabolic model reconstructed in this study was based on an updated version of the AraCyc database, AraCyc 9.0. An iterative process of manual curation and model testing was performed to correct errors in the model as far as possible. The model was manually compartmented into five subcellular compartments, cytosol, mitochondrion, plastid, peroxisome and vacuole. Energetic costs for growth-dependent processes were extensively included in the stoichiometry of the reactions and maintenance reactions were included for modelling costs for growth-independent processes. The final model, with 2769 reactions and

2739 metabolites, can be used to simulate metabolic fluxes in both heterotrophic and phototrophic conditions.

A

```

UNIQUE-ID - ADENYL-KIN-RXN
TYPES - EC-2.7.4
TYPES - Small-Molecule-Reactions|
COMMON-NAME - adenylate kinase
EC-NUMBER - 2.7.4.3
ENZYMATIC-REACTION - ENZRNXQT-18922
ENZYMATIC-REACTION - ENZRNXQT-18900
ENZYMATIC-REACTION - ENZRNXQT-18465
ENZYMATIC-REACTION - ENZRNXQT-18032
ENZYMATIC-REACTION - ENZRNXNT9-2221
ENZYMATIC-REACTION - ENZRNXNT9-1653
ENZYMATIC-REACTION - ENZRNXNT9-1303
ENZYMATIC-REACTION - ENZRNXNT9-45
ENZYMATIC-REACTION - ENZRNXQT-6584
ENZYMATIC-REACTION - ENZRNXQT-6414
ENZYMATIC-REACTION - ENZRNXQT-6320
ENZYMATIC-REACTION - ENZRNXQT-5188
IN-PATHWAY - DENOVO PURINE3-PWY
LEFT - AMP
LEFT - ATP
MEMBER-SORT-FN - NUMBERED-CLASS-SORT-FN
OFFICIAL-EC? - T
REACTION-DIRECTION - LEFT-TO-RIGHT
RIGHT - ADP
^COEFFICIENT - 2
SYNONYMS - Adenylokinase
SYNONYMS - Adenylic kinase
SYNONYMS - Myokinase

```

B

```

"ADENYL-KIN-RXN":
    1/1 "ATP" + 1/1 "AMP" -> 2/1 "ADP"
    ~

```

Figure 3.1 Example of the definition of a reaction in A) the AraCyc database and B) ScrumPy format.

Table 3.1 List of unconserved metabolites in the model.

Metabolite name in model	Common name
CPD-12377_c	hydroxyl radical
DNA-processing-cost_c	dummy metabolite for DNA processing cost
HYDROGEN-MOLECULE_c	hydrogen molecule
HYDROGEN-PEROXIDE_c	hydrogen peroxide
HYDROGEN-PEROXIDE_m	hydrogen peroxide
HYDROGEN-PEROXIDE_p	hydrogen peroxide
HYDROGEN-PEROXIDE_x	hydrogen peroxide
OH_c	hydroxide ion
OXYGEN-MOLECULE_c	oxygen molecule
OXYGEN-MOLECULE_m	oxygen molecule
OXYGEN-MOLECULE_p	oxygen molecule
OXYGEN-MOLECULE_x	oxygen molecule
Photon_p	photon
Protein-polymerisation-cost_c	dummy metabolite for protein polymerisation cost
Protein-processing-cost_c	dummy metabolite for protein processing cost
Protein-translocation-cost_c	dummy metabolite for protein translocation cost
PROTON_c	proton
PROTON_m	proton
PROTON_p	proton
PROTON_x	proton
Pumped-PROTON_c	dummy metabolite for proton gradient across membrane
Pumped-PROTON_m	dummy metabolite for proton gradient across membrane
Pumped-PROTON_p	dummy metabolite for proton gradient across membrane
Pumped-PROTON_v	dummy metabolite for proton gradient across membrane
RNA-processing-cost_c	dummy metabolite for RNA processing cost
SUPER-OXIDE_c	superoxide
SUPER-OXIDE_m	superoxide
SUPER-OXIDE_p	superoxide
SUPER-OXIDE_x	superoxide
WATER_c	water
WATER_m	water
WATER_p	water
WATER_v	water
WATER_x	water
x_ATPWork	dummy metabolite for ATP maintenance cost
x_NADPHWorkc	dummy metabolite for NADPH maintenance cost
x_NADPHWorkm	dummy metabolite for NADPH maintenance cost
x_NADPHWorkp	dummy metabolite for NADPH maintenance cost
x_NADPHWorkx	dummy metabolite for NADPH maintenance cost
x_O2	external oxygen molecule
x_Photon	external photon

4 Predictive accuracy of a flux balance model of heterotrophic Arabidopsis metabolism

4.1 Introduction

4.1.1 Using flux balance analysis to predict metabolic fluxes

One of the main uses of flux balance metabolic models, whether hand-built or genome-scale, is to predict fluxes through the metabolic network. The most common modelling technique used for predicting fluxes from flux balance models is flux balance analysis (FBA). The use of FBA in analysing metabolic models dates back to the 1980s (Papoutsakis, 1984; Fell and Small, 1986). The basic assumption of FBA is that the metabolic system is at a pseudo-steady state where the rate of production of an internal metabolite equals its rate of consumption, i.e. flux balance. In mathematical terms, the flux balance constraint is defined by the equation $\mathbf{N} \cdot \mathbf{v} = 0$, where $\mathbf{N}$ is the stoichiometry matrix of the metabolic model and $\mathbf{v}$ is a vector of reaction fluxes.

Flux balance analysis is a constraint-based modelling technique where model predictions are made based on the applied constraints. As well as the flux balance constraint, additional metabolic constraints are applied such as the experimentally measured nutrient uptake rate and/or rate of biomass synthesis. Together, these constraints define the allowable flux solution space, i.e. the space of allowable reactions fluxes in the metabolic system. To obtain biologically relevant flux distribution(s) from the allowable flux solution space, it is assumed that the metabolic system has evolved to be optimised toward a biological objective, most commonly maximising biomass production in microbial systems. The notion of a biological objective can be captured mathematically with the use of an objective function. With a linear

objective function (and linear constraints), an optimal flux solution can be calculated using linear programming. The resulting optimal flux solution represents the FBA prediction of the flux distribution corresponding to the given objective. Here, linear programming was used to calculate FBA flux solutions from the Arabidopsis genome-scale metabolic model.

4.1.2 Comparisons between FBA predictions and experimentally measured fluxes in plant models

To investigate the predictive accuracy of FBA in plant systems, flux predictions from the previous version of the AraCyc-based Arabidopsis genome-scale metabolic model were compared with fluxes estimated by ^{13}C -metabolic flux analysis (MFA) (Williams et al., 2010). While the relative changes in metabolic fluxes in glycolysis and the TCA cycle with increased temperature and hyperosmotic conditions in heterotrophic cell culture were well predicted, the absolute values of flux predictions did not match well with experimental values. In particular, flux through the oxidative pentose phosphate pathway (OPPP), an important source of NADPH in plant cells under heterotrophic condition (Kruger and von Schaewen, 2003; Schwender et al., 2003), was poorly predicted from the genome-scale model. While MFA estimated significant flux through the OPPP under all three conditions tested, no flux was predicted by the genome-scale model. Inspection of the FBA flux solutions showed that NADPH was mainly produced by NADP-glyceraldehyde dehydrogenase (GAPDH). A similar comparison was made between the flux predictions from a flux balance model of *Brassica napus* developing embryos and fluxes determined by ^{13}C -MFA (Hay and Schwender, 2011a). Overall, there was a large agreement between the flux predictions

and the MFA flux maps. The major difference between the FBA prediction and the MFA fluxes was again the OPPP. While substantial OPPP flux was found by ^{13}C -MFA (Schwender et al., 2003), FBA predicted that a variety of plastidial dehydrogenases were preferred over the OPPP in generating NADPH (Hay and Schwender, 2011b).

Comparisons between FBA and MFA fluxes highlight the problem of plant flux balance models in predicting OPPP flux in heterotrophic condition (Williams et al., 2010; Hay and Schwender, 2011a). This is likely to be caused by the presence of other dehydrogenases which are used to produce NADPH in preference to the OPPP. However, it is unclear whether the dehydrogenases used as an alternative source of NADPH in these models can function in the direction of NADPH production *in vivo*. To look into the problem of predicting OPPP flux, the effect of the reversibility of two NADP-dependent dehydrogenases on FBA flux predictions was investigated.

4.1.3 Cellular maintenance costs and other energetic costs in flux balance models

FBA is a constraints-based modelling approach and one of the main constraints imposed on such models is the requirement to synthesise biomass of appropriate composition for the tissue under consideration. In addition to tracking the flows of carbon and other elements (nitrogen, phosphorus, oxygen, sulphur) through the metabolic network for the synthesis of biomass, FBA also balances the utilisation of energy and reducing power. Herein lies a major problem: the synthesis of biomass only consumes a small proportion of the total energy budget (Masakapalli et al., 2010). Thus, fluxes through the energy transforming pathways - glycolysis,

the oxidative pentose phosphate pathway (OPPP) and the tricarboxylic acid (TCA) cycle - are greatly underestimated by FBA when solely constrained by biomass synthesis (Poolman et al., 2009).

Accurate prediction of these fluxes requires several other energy costs to be taken into account. First, synthesis of biomass depends on both the uptake of substrates into the cell and the transport of metabolites between subcellular compartments. Many of these transport steps are energy-requiring, being driven by electrochemical gradients that are maintained by expenditure of ATP. Secondly, there is the cost of maintaining the instantaneous steady state, irrespective of whether the system is growing, in the face of processes such as turnover of polymers and the inherent leakiness of membranes.

Transport costs are relatively easy to account for in FBA by adding appropriate transporters with defined ATP stoichiometry to the metabolic model, although this is complicated by the subcellular compartmentation of the metabolic network. Maintenance costs, by contrast, are notoriously difficult to quantify (Penning de Vries, 1975; Amthor, 2000; Piques et al., 2009) and accounting for maintenance has recently been identified as a major challenge for FBA of genome-scale models of plant metabolism (Sweetlove et al., 2013). For microbial models maintenance constraints corresponding to growth-associated and non-growth-associated maintenance can be conveniently obtained from chemostat measurements (Thiele and Palsson, 2010). This approach is not generally applicable to plant cells, which are rarely grown in chemostats, let alone for plants themselves, and so an alternative strategy is to introduce a generic ATPase into the model to account for the cost of cell maintenance (Thiele and Palsson, 2010). This

method for assessing maintenance ATP costs has been adopted in FBA investigations of plant metabolism (Poolman et al., 2009; Hay and Schwender, 2011b) and is based on varying the flux of a generic ATPase in the model until the carbon consumed by the model matches that measured experimentally (Poolman et al., 2009). The importance of including maintenance costs is demonstrated by the large effect of altered ATP demand on the flux distribution in central metabolism (Poolman et al., 2009). However, this method assumes that maintenance costs can be described entirely in terms of ATP expenditure, even though maintenance processes such as antioxidant metabolism and the resynthesis of lipid require substantial amounts of reductant.

To address this, a method for accounting for both the ATP and reductant costs of cell maintenance was developed and applied to the Arabidopsis genome-scale metabolic model. The effect of attributing the cell maintenance costs to the consumption of both ATP and reductant on the predictive accuracy of FBA was investigated by comparing FBA flux solutions with experimentally determined fluxes. Moreover, a comparison of Arabidopsis cells grown under control, elevated temperature and hyperosmotic conditions highlights substantial differences in the ATP and reductant costs of cell maintenance. This leads to the suggestion that stress can be considered as any condition that leads to an increase in cell maintenance costs.

4.2 Results

4.2.1 The effect of reversibility of NADP-dependent dehydrogenase on flux predictions

Comparisons between flux predictions from plant flux balance models and MFA flux maps showed that plant flux balance models significantly underestimate flux through the oxidative pentose phosphate pathway (OPPP) under heterotrophic condition (Williams et al., 2010; Hay and Schwender, 2011a). Given that the OPPP is known to be an important source of NADPH in plant cells (Kruger and von Schaewen, 2003; Schwender et al., 2003), the likely cause of failing to predict OPPP flux in plant flux balance models is the presence of other NADP-dependent dehydrogenases which could produce NADPH in preference to the OPPP given a certain objective function. To address this, the effect of the reversibility of NADP-GAPDH and NADP-dependent malate dehydrogenase (MDH) on flux predictions was investigated by setting the two reactions as either reversible or irreversible and calculating the flux ranges of plastidial NADP-dependent dehydrogenases using flux variability analysis (FVA). Biomass in *Arabidopsis* heterotrophic cell culture grown under control condition was used as constraint (Williams et al., 2010) with minimisation of total flux as the primary objective function. Significant flux through the OPPP can be predicted only when both NADP-GAPDH and NADP-MDH were set as irreversible in the direction of NADPH consumption (Table 4.1). When NADP-GAPDH and/or NADP-MDH were set as reversible, the two reactions were predicted to be the major source of NADPH in the plastid. The two reactions can generate NADPH by participating in transhydrogenase cycles that allow the transfer of

reducing power from either plastidial or cytosolic NADH to plastidial NADP with no net changes in the NAD(H) and NADP(H) pool sizes (Figure 4.1). However, thermodynamic considerations suggest that such transhydrogenase cycles are unlikely to act as an alternative source of plastidial NADPH. In plant cells, the NAD(H) pool is more oxidised than the NADP(H) pool in both the light and the dark (Heineke et al., 1991; Igamberdiev and Gardeström, 2003), so it is thermodynamically implausible for cytosolic NADH to drive the provision of plastidic NADPH without an energy input. A proton-translocating transhydrogenase can achieve this (Pedersen et al., 2008) but although such a protein exists in most prokaryotes and in animal mitochondria, the Arabidopsis genome does not appear to contain the relevant gene (Arkblad et al., 2001). So for subsequent modelling, the plastidial NADP-dependent dehydrogenases were set to operate in the direction of NADPH consumption.

4.2.2 Estimating maintenance costs

FBA predicted that biomass synthesis with associated metabolite transport costs by the Arabidopsis heterotrophic cell culture under control conditions required a glucose consumption rate of 84.5% of the experimentally determined value. It can be assumed that the remaining glucose was consumed to meet maintenance energy demands through the provision of ATP and NADPH. To account for these maintenance costs, generic ATPase and NADPH oxidase steps were added to the model. These steps were iteratively assigned different flux values generating a set of model solutions. The glucose consumption rate in the model increased with increasing flux through the generic steps and, for a subset of ATP and NADPH maintenance

values, the total glucose consumption in the model matched the measured value. However, because ATP and NADPH are generated by different metabolic processes that compete for the same hexose phosphate precursor, a maintenance cost trade-off operates, i.e. it is only possible to produce more ATP at the expense of NADPH production and vice versa.

This trade-off was explored by calculating the Pareto front (Vo et al., 2004; Bordbar et al., 2011) of ATP and NADPH maintenance costs by simultaneously maximising for ATP and NADPH production for maintenance with the constraints of experimentally measured biomass synthesis and glucose consumption rates in heterotrophic cell culture under control condition (Figure 4.2). The Pareto optima that defined the Pareto front corresponded to different flux distributions in the central metabolic network and it was observed that the flux ratio between the oxidative steps of the OPPP (generating NADPH) and glycolysis (leading to ATP generation) decreased from 1.37 to 0.38 as the maintenance demand switched from NADPH to ATP (Figure 4.2). This suggests that this flux ratio could be a good indicator of the actual ratio of NADPH to ATP maintenance. The measured value for this flux ratio for the Arabidopsis cell culture under control conditions (Williams et al., 2010) was used to determine the point on the Pareto front that corresponds to the *in vivo* balance of ATP and NADPH maintenance costs (Figure 4.2). From this, it was possible to deduce that maintenance in the Arabidopsis cell culture under control conditions, required 54.5 mmol ATP d⁻¹ L⁻¹ culture, corresponding to 33% of the total ATP produced, and 19.2 mmol NADPH d⁻¹ L⁻¹ culture, corresponding to 50% of the total NADPH produced.

4.2.3 The effect of transport and maintenance costs on flux predictions

The effect of including transport and maintenance costs in the FBA of the Arabidopsis cell culture grown under control conditions was investigated by comparing the predicted flux solutions with the fluxes in central metabolism obtained by ^{13}C -MFA (Figure 4.3; top left). FBA flux solutions were generated using minimisation of total flux as the objective function. Using just the constraint of biomass synthesis, with transport costs excluded from the stoichiometry of the model and no maintenance costs constraints, resulted in a relatively poor prediction of the fluxes through central metabolism where the glycolytic, OPPP and TCA cycle fluxes were substantially underestimated (Figure 4.3; top middle). Including transport costs improved the predicted fluxes through glycolysis and the TCA cycle, and adding NADPH and ATP maintenance costs as well also altered the predicted flux distribution (Figure 4.3; top right). At the highest NADPH maintenance demand corresponding to point 1 on the Pareto front (Figure 4.2), there was a large flux through the OPPP and very little TCA cycle flux (Figure 4.3; bottom left); whereas at the highest ATP maintenance demand corresponding to point 3 on the Pareto front (Figure 4.2), the fluxes through glycolysis and the TCA cycle were higher compared to the other two scenarios, and there was very little OPPP flux (Figure 4.3; bottom right). However when the balance between the NADPH and ATP maintenance costs was set by matching the flux ratio between the oxidative steps of the OPPP and glycolysis to the measured value (Figure 4.2, point 2), the resulting flux distribution (Figure 4.3; bottom

middle) was qualitatively similar to that obtained by ^{13}C -MFA (Figure 4.3; top left).

The quantitative match between the fluxes predicted with the different transport and maintenance estimates and the experimentally determined fluxes for the Arabidopsis cells under control conditions was assessed by calculating the weighted Euclidean distance between them. Since the reactions in the genome-scale metabolic model and the MFA model are not identical, the genome-scale model reactions were mapped onto 15 MFA model reactions in central carbon metabolism (see Appendix 2 for details). With the application of different combinations of constraints, including biomass synthesis, transport costs, glucose consumption, ATP maintenance and NADPH maintenance, the Euclidean distances, weighted by the standard error of MFA flux estimations, were calculated between the fluxes measured by ^{13}C -MFA and the average flux values of flux ranges calculated from flux variability analysis (FVA) (Figure 4.4; see chapter 2 for detailed methods). The quantitative analysis confirmed that when transport and maintenance costs (both ATP and NADPH) were included as constraints, FBA flux predictions were very close to the ^{13}C -MFA flux map.

4.2.4 The effect of objective functions on flux predictions

Because flux solutions in FBA will also be influenced by the objective function, flux solutions were generated using five different objective functions: minimisation of overall flux, maximisation of biomass, minimisation of glucose consumption, maximisation of ATP production and maximisation of NADPH production. In total, FBA flux solutions from 25 combinations of objective

function and maintenance energy / other constraints were computed (Appendix S7). The average flux values of flux ranges of the 15 reactions mapped to the MFA model were shown to form feasible steady-state flux solutions in all 25 combinations of objective function and constraints and these were compared to the fluxes measured by ^{13}C -MFA by calculating the weighted Euclidean distances (Figure 4.4; see chapter 2 for detailed methods).

Irrespective of the objective function used for generating the flux solutions, FBA gave flux predictions that were very close to the ^{13}C -MFA flux map when transport and maintenance costs (both ATP and NADPH) were included as constraints. Generally, ATP maintenance had a quantitatively larger effect than NADPH maintenance, although including both reduced the Euclidean distance further and thus improved the flux prediction. Inclusion of transport costs improved the flux predictions, except when maximisation of ATP was the objective function. In this case, the predicted flux distribution with the biomass and glucose consumption constraints was unaffected by the extra ATP expenditure on transport because the transport constraint had no bearing on the ATP-generating capacity of the network.

Although the use of different objective functions had little impact on the prediction of the 15 compared fluxes, there may be advantages to choosing one objective function over the others. For example, minimisation of flux eliminates unconstrained cycles that can otherwise carry infinite fluxes in solutions obtained with the other objective functions. This was confirmed by a flux variability analysis (FVA): no reactions with an infinite flux range (unconstrained reactions) were found using minimisation of flux as an

objective function, whereas 36 unconstrained metabolic reactions were found for the other four objective functions. Unconstrained reactions are often involved in parallel metabolic pathways in different compartments, e.g. glycolysis in the cytosol and the plastid. Reactions involved in parallel pathways can be resolved into finite flux ranges with the minimisation of flux objective function, but not with any of the other objective functions investigated. For example, it was possible to predict using minimisation of flux the extent to which glycolysis occurred in both the cytosol and the plastid, and that most of the flux through the TCA cycle was mitochondrial and not cytosolic (Figure 4.5).

4.2.5 Predicting fluxes and estimating maintenance costs in heterotrophic cell culture under stress conditions

The method for estimating the maintenance costs described in section 4.2.2 was applied to heterotrophic cell culture grown at elevated temperature or under hyperosmotic conditions, using the experimentally measured condition-specific biomass synthesis and glucose consumption rates (Williams et al., 2010). The estimated maintenance costs were used as constraints to generate FBA flux predictions with minimisation of total flux as the objective function. Comparison of the ^{13}C -MFA flux maps and the FBA predictions shows that using the ATP and NADPH maintenance costs based on the experimentally determined flux ratio between the oxidative steps of the OPPP and glycolysis (Table 4.2) as constraints resulted in excellent agreement between the predicted and experimental flux maps under both elevated temperature (29°C) and hyperosmotic (171 mM mannitol) conditions (Figure 4.6).

Analysis of the predicted maintenance costs (Table 4.2) shows that the proportion of the available ATP used for maintenance was low under hyperosmotic stress relative to the control, and high at elevated temperature. The proportion of the available NADPH that was used for maintenance was also increased at elevated temperature. These predictions were in good agreement with the values calculated from the MFA flux maps (Table 4.2; Appendix S8). The results suggest that the relative contribution of NADPH and ATP to the maintenance cost was strongly dependent on the conditions imposed on the cell culture, with predicted values for the NADPH: ATP maintenance ratio of 0.35 under control conditions, 0.84 under hyperosmotic conditions and 0.19 at elevated temperature. In contrast the relative contribution of NADPH and ATP to the cost of biosynthesis, computed by FBA with biomass synthesis constraint without maintenance costs, was very similar under the three conditions, reflecting the similarity in the biomass composition of the cells (Table 4.2).

4.3 Discussion

4.3.1 Accounting for energy expenditure in a flux balance model

The utility of flux balance models depends on the extent to which realistic flux distributions can be predicted from them. While the biomass constraint has proved effective in predicting fluxes in microbes (Feist & Palsson, 2010), it is less clear whether this constraint provides sufficient bounds for flux prediction in more complex organisms such as plants. Even though it has been previously shown that there is significant agreement in the flux predictions for heterotrophic plant metabolism and those estimated by ^{13}C -MFA (Williams et

al., 2010; Hay and Schwender, 2011a), certain fluxes, such as those in the central pathways of glycolysis, the TCA cycle, and particularly the OPPP, were not well matched.

The main issue is that these central pathways, in addition to providing carbon skeletons for the synthesis of biomass components, are also the main routes for energy transformation. Although synthesis of biomass consumes energy, there are other substantial energy drains in the cell, including the cost of transporting ions, metabolites and macromolecules, and the cost of cell maintenance. Both of these costs are potentially higher in plants than in microbial systems and could therefore have a substantial impact on predicted flux distributions in central metabolism. In fact under control conditions the analysis presented here shows that synthesis of biomass (including metabolite transport) accounts for only 67% of the total ATP budget in a heterotrophic *Arabidopsis* cell culture under control conditions, with the remainder being used for maintenance (Table 4.2).

Transport costs can be included in FBA by specifying the co-transport of a defined number of protons per transported molecule in most cases and the ATP requirement is then defined by the proton-ATP stoichiometry of the ATPases that generate transmembrane proton gradients. Some transport costs have been included in previous flux balance models of plant metabolism (e.g. Hay and Schwender, 2011b) but these have only considered transport across the plasma membrane. As shown in section 4.2.3, adding costs associated with intracellular transport and protein translocation improved the prediction of the glycolytic and TCA cycle fluxes significantly, but it had no impact on the OPPP flux prediction (Figure 4.3).

Thus adding transport costs to the usual biomass constraint is not sufficient to optimise the predictive accuracy of FBA in a plant metabolic network.

Maintenance costs are more difficult to account for in FBA because maintenance covers many ATP- and NADPH-dependent processes that are difficult to quantify. While some aspects of maintenance, such as macromolecule turnover, may be measurable experimentally, the techniques rely on numerous assumptions and there is a substantial range of estimates for such costs (Piques et al., 2009; Scheurwater et al., 2000). Others, such as membrane leak rate, recycling of unwanted metabolites (Seaver et al., 2012) and antioxidant NADPH expenditure have yet to be experimentally quantified and are not rigorously accounted for, even in microbial FBA studies. This major limitation in FBA can be solved by the method described in section 4.2.2, which leads to a substantial improvement in the accuracy of the flux predictions for the pathways of central metabolism in heterotrophic *Arabidopsis* cells (Figure 4.6).

Using a Pareto optimisation, it was found that the balance between ATP and NADPH maintenance had a substantial impact on the distribution of fluxes in central metabolism (Figure 4.3) and this highlighted the importance of setting the correct division of maintenance energy consumption. It was therefore necessary to introduce a constraint on the relative activities of the generic ATPase and NADPH oxidase that were used to account for maintenance costs. The branch point where hexose phosphates enter either glycolysis or the OPPP is likely to be a key determinant of the commitment to ATP and NADPH production in heterotrophic cells grown in glucose medium. Matching the ratio of fluxes at this branch point to the value taken from ^{13}C -MFA

measurements substantially improved the flux prediction and produced the closest overall match to the ^{13}C -MFA flux distribution for central metabolism irrespective of the objective function used to generate the flux predictions.

The additional information of a single flux ratio, the flux ratio of glycolysis to OPPP, allows ATP and NADPH maintenance costs to be accounted for and produces a more realistic prediction of the flux distribution through the central metabolic network. However, it should be noted that in systems such as developing oilseeds where the OPPP is a less significant source of NADPH (Schwender et al., 2003), information on additional flux ratios may be required. In this study, the required flux ratio was taken from ^{13}C -MFA measurements on the same cell suspension culture. In practice, ^{13}C -MFA flux maps may not always be available and the requirement to generate one would negate the benefit of flux prediction by FBA. However, there are a number of simpler labelling experiments that can give access to this flux ratio. For example, the labelled CO_2 generated from positionally labelled ^{14}C -glucose gives an estimate of the OPPP flux relative to other pathways of carbohydrate oxidation (Malone et al., 2006). A more direct measure of the OPPP flux can be obtained by feeding $[1\text{-}^{14}\text{C}]\text{gluconate}$, which enters the OPPP as 6-phosphogluconate and is released as $^{14}\text{CO}_2$ (Garlick et al., 2002; Zhao et al., 2008). Alternatively, mass isotopomers of downstream metabolic products of $[1,2\text{-}^{13}\text{C}]\text{glucose}$ can allow estimation of the OPPP flux (Dusick et al., 2007). While these methods may require some optimisation, in principle at least, these relatively straightforward labelling experiments can allow the crucial flux ratio to be estimated.

4.3.2 The impact of different constraints and objective functions on flux predictions

The impact of different constraints and objective functions on the accuracy of the flux distribution predicted by flux balance analysis was investigated. The most important conclusion is that the comprehensive accounting for energy costs in the system is more important in terms of predicting a realistic flux distribution than the choice of objective function.

A similar systematic evaluation of objective functions and constraints in *E. coli* (Schuetz et al., 2007) gave a rather different conclusion. In *E. coli*, inclusion of ATP and NADPH maintenance costs actually reduced the predictive fidelity of the flux balance analysis, producing a poorer match to MFA-determined fluxes. This curious result may have been caused either by handling the NADPH and ATP costs separately, or by the relatively low maintenance costs in *E. coli*. In an *E. coli* flux balance model, maintenance ATP demand was $7.6 \text{ mmol ATP g DW}^{-1} \text{ h}^{-1}$, which was only 14% of the $55.2 \text{ mmol ATP g DW}^{-1} \text{ h}^{-1}$ generated in the system (Varma and Palsson, 1994). This contrasts with the 33% of the total ATP produced in heterotrophic *Arabidopsis* cells that was used for maintenance under control conditions (Table 4.2). Nevertheless, it is difficult to see why an FBA solution that does not include maintenance costs would generate a closer match to the MFA-determined fluxes than one that includes maintenance, unless the included maintenance values were substantial over-estimates. This is possible as estimates of *E. coli* maintenance costs (in terms of ATP) vary from $1.9 \text{ mmol ATP g DW}^{-1} \text{ h}^{-1}$ in glucose limitation to 16.8 and $30.8 \text{ mmol ATP g DW}^{-1} \text{ h}^{-1}$ in N and S limitation, respectively (Farmer and Jones, 1976). This highlights the

problem of applying experimentally determined maintenance values obtained in one set of conditions to FBA models of metabolism in different conditions. A sensitivity analysis performed on a genome-scale metabolic model of *E. coli* showed that maintenance costs constraints were important for predictions of growth rates and oxygen uptake rates (Feist et al., 2007). This is in line with the results in this chapter in demonstrating the importance of maintenance costs constraints on model predictions.

Although the choice of objective function had little impact on the accuracy with which the 15 compared fluxes were predicted by FBA, it is nevertheless still an important consideration. For example minimisation of flux tends to eliminate futile cycles from the predicted flux solutions and this is not necessarily in agreement with experimental measurements. Some ^{13}C -MFA measurements have suggested that sucrose cycling can consume a surprisingly large fraction of the available ATP (Rontein et al., 2002; Alonso et al., 2007). However, theoretical analysis has identified an assumption in these steady state analyses that has yet to be verified and is quite likely to be unjustifiable (Kruger et al. 2007). In any case biological systems have evolved to operate over a range of conditions that may require competing objectives and this competition has been explored in *E. coli* metabolism using the Pareto optimality concept (Schuetz et al., 2012). It was found that the Pareto surface formed from three objective functions (minimisation of total flux, maximisation of ATP yield and maximisation of biomass yield) best described the fluxes measured from ^{13}C -MFA. Thus, Pareto optimisation is a versatile tool for FBA, allowing the exploration of competing objective

functions (Schuetz et al., 2012) or, as shown in this study, competing constraints.

4.3.3 Impact of hyperosmotic stress and elevated temperature on maintenance costs

As well as greatly improving the predictive accuracy of FBA, matching the flux ratio between the oxidative steps of the OPPP and glycolysis to the experimentally measured value led to estimates of the ATP and NADPH costs of maintenance under control, elevated temperature and hyperosmotic conditions (Table 4.2). This information can be interpreted in terms of the growth characteristics of the cell line under the three conditions (Table 4.3). Increasing the temperature increased the proportion of both ATP and NADPH that was used for maintenance by the cell culture, while at the same time causing a marked reduction in carbon conversion efficiency. It is likely that the increased membrane leakage (Ruter, 1993) and oxidative stress (Schwärzlander et al., 2012) that occur at elevated temperature make a major contribution to the increased ATP and NADPH maintenance costs respectively. This conclusion is consistent with observations based on respiratory rates suggesting that maintenance respiration and not growth respiration increase with increased temperature (Amthor, 2000). In contrast, hyperosmotic conditions caused a small decrease in the proportion of NADPH used for maintenance, and a substantial reduction in the proportion of ATP used for the same purpose, while at the same time reducing the growth rate and increasing the carbon conversion efficiency. The lower proportion of ATP used for maintenance may reflect the slower turnover of cellular components in the slower growing cells, but more generally the

contrast with the markedly increased maintenance costs in the stressed cells at elevated temperature suggests that the cells were not stressed under these hyperosmotic conditions. This conclusion may be compared with a recent analysis in which several environmental stresses, including elevated temperature, were observed to increase the rate of mitochondrial membrane potential pulsing in parallel with a marked change in mitochondrial redox status in the epidermal cells of Arabidopsis roots (Schwarzländer et al., 2012). Strikingly, hyperosmotic conditions did not alter these parameters, again suggesting that these conditions do not disrupt normal cellular activity. Thus it appears that an increase in cell maintenance costs might be a metabolic signature of stress conditions and that the slow growth of the Arabidopsis cells under hyperosmotic conditions should be interpreted as being the result of restricted glucose uptake rather than the result of perturbed metabolic processes.

4.3.4 Summary

The aim of this chapter was to investigate the importance of various modelling parameters, with emphasis on energy-cost constraints, for the predictive accuracy of the Arabidopsis flux-balance model with respect to heterotrophic metabolism. A method for accounting for ATP and NADPH maintenance costs was developed based on Pareto optimisation and matching to an experimental flux ratio. By comparing with fluxes determined by ^{13}C -MFA, accounting for transport and maintenance costs was shown to substantially improve the accuracy of fluxes predicted from a flux-balance model of heterotrophic Arabidopsis cells in culture, irrespective of the objective function used in the analysis. The method for accounting for

maintenance costs was shown to be effective under both control and stress conditions. Of the two stress conditions, only elevated temperature led to a substantial increase in cell maintenance costs but not hyperosmotic conditions, suggesting that the hyperosmotic conditions tested did not impose a metabolic stress in so far as the metabolic network is not forced to devote more resources to cell maintenance.

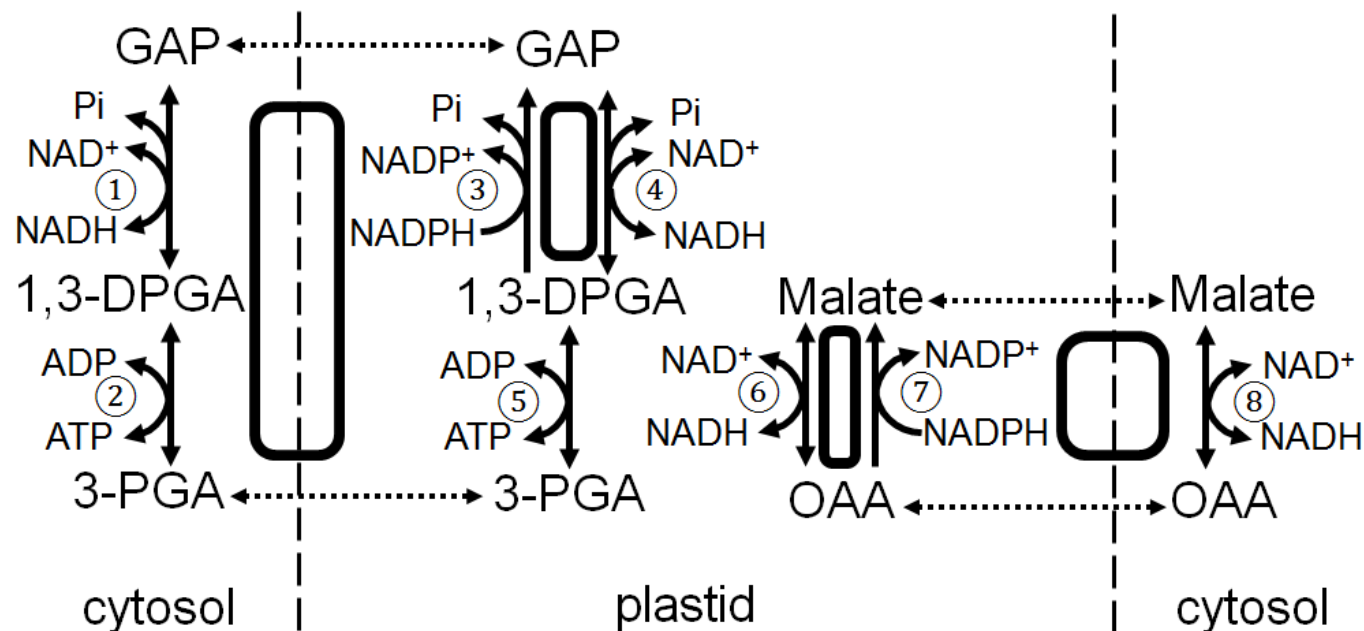


Figure 4.1 Transhydrogenase cycles involving plastidial NADP-GAPDH and NADP-MDH. In the FBA model, NADP-GAPDH and NADP-MDH were set to be irreversible in the direction of NADPH consumption as shown in the diagram. This forces the transhydrogenase cycles (represented by curved rectangles) to go in the direction of NADPH consumption, generating NADH in the plastid or the cytosol at the expense of plastidic NADPH. Setting either or both NADP-GAPDH and NADP-MDH to be reversible, allowed the transhydrogenase cycles to go in the direction of NADPH production, which led to altered flux predictions. Abbreviations: GAP, glyceraldehyde 3-phosphate; 1,3-DPGA, glycerate 1,3-bisphosphate; 3-PGA, glycerate 3-phosphate; OAA, oxaloacetate. Enzymes (circled numbers): (1) cytosolic NAD-dependent GAPDH, (2) cytosolic phosphoglycerate kinase, (3) plastidial NADP-GAPDH, (4) plastidial NAD-dependent GAPDH, (5) plastidial phosphoglycerate kinase, (6) plastidial NAD-dependent MDH, (7) plastidial NADP-MDH, (8) cytosolic NAD-dependent MDH.

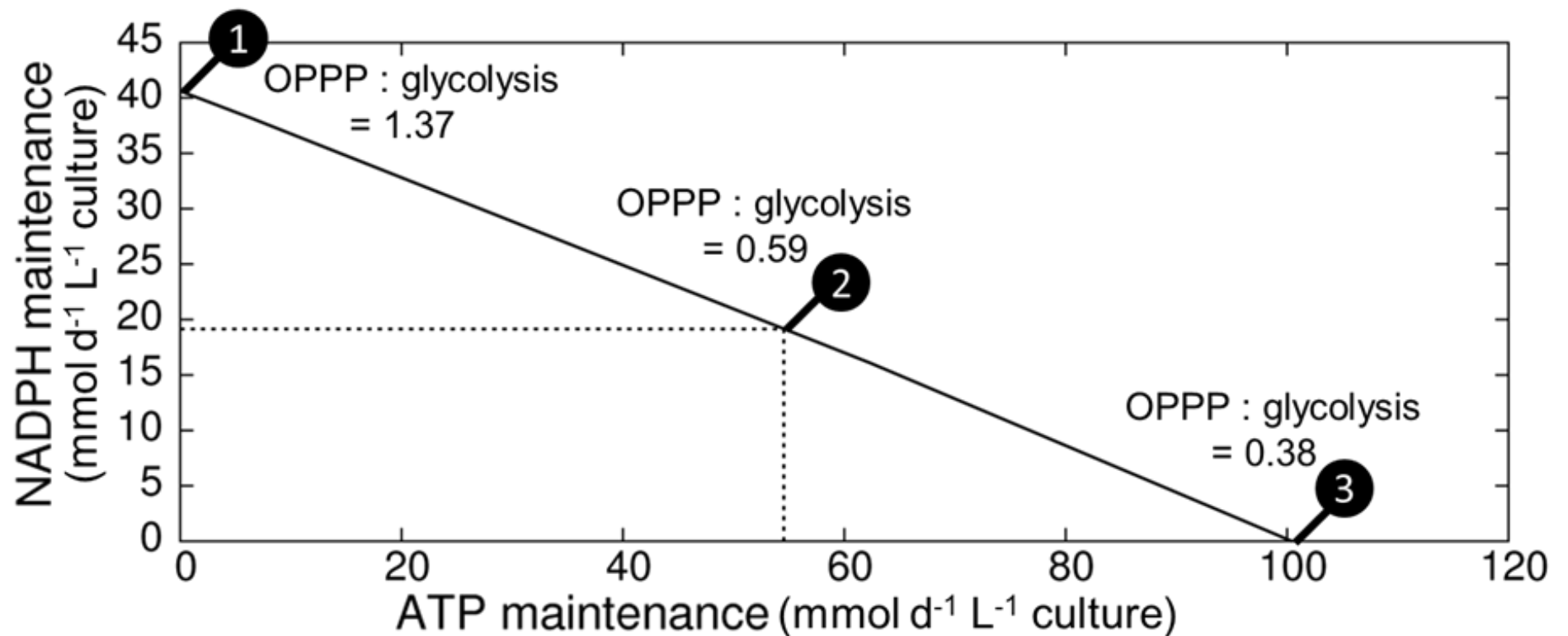


Figure 4.2 Pareto front for the trade-off between ATP and NADPH maintenance in a heterotrophic *Arabidopsis* cell culture under control conditions. Dotted lines represent estimates of the in vivo ATP and NADPH maintenance costs (in mmol d⁻¹ L⁻¹ culture) that correspond to the flux ratio between the oxidative steps of the OPPP and glycolysis determined by ¹³C-MFA (black circle 2).

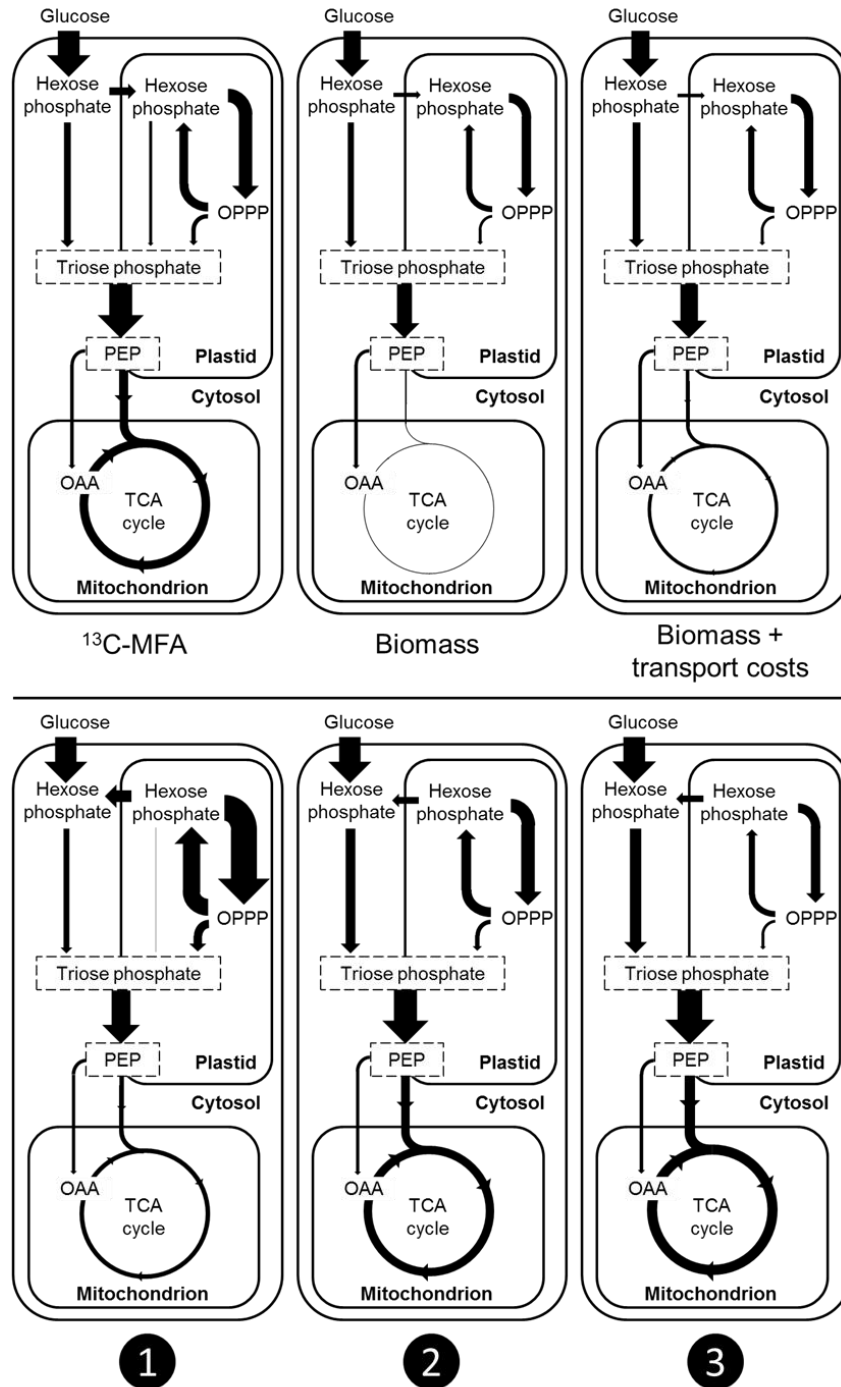


Figure 4.3 Comparison of MFA and FBA flux maps for *Arabidopsis thaliana* cells under control conditions. Flux maps of central carbon metabolism from ^{13}C -MFA (top left); flux prediction with biomass constraint and no transport and maintenance costs (top middle); flux prediction with biomass and transport costs constraints and no maintenance costs (top right); flux predictions with biomass, transport costs and maintenance costs constraints (bottom row). Numbers in black circles correspond to the numbers on the Pareto front indicated in Figure 4.2. Arrow thickness is proportional to flux through the reaction. Average flux through the TCA cycle is shown on each flux map.

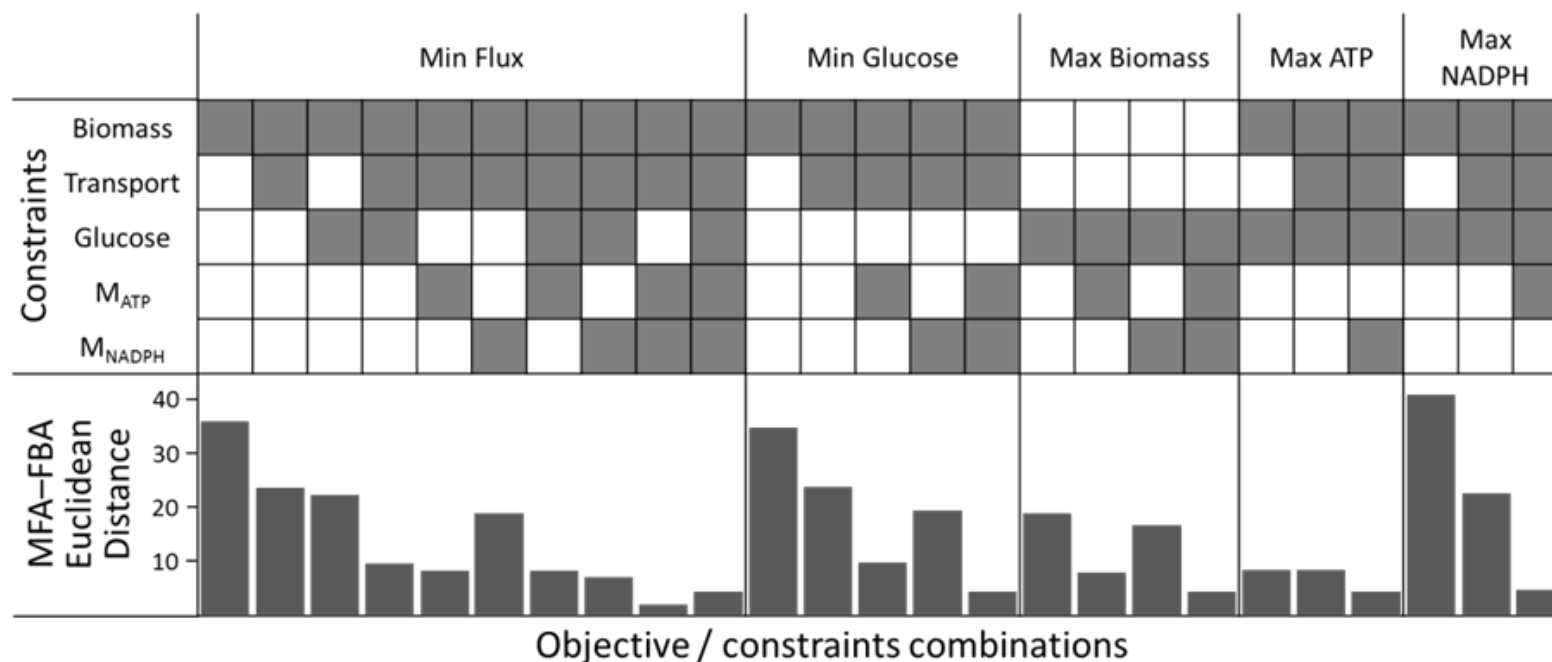


Figure 4.4 Weighted Euclidean distances between flux predictions from 25 different objective function / constraints combinations and the ^{13}C -MFA flux map for the heterotrophic *Arabidopsis* cell culture under control conditions. Each column corresponds to a combination of objective function and one or more constraints where the constraints applied are indicated by the shading. The Euclidean distance is a measure of the closeness of fit between the FBA flux prediction and the ^{13}C -MFA flux map, with a small value being a close fit. Objective functions: minimisation of total flux (Min Flux), minimisation of glucose consumption (Min Glucose), maximisation of biomass production including transport costs (Max Biomass), maximisation of ATP yield (Max ATP), maximisation of NADPH yield (Max NADPH). Constraints: biomass production excluding transport costs (Biomass), transport costs (Transport), glucose consumption rate (Glucose), estimated ATP maintenance cost from Pareto optimisation (M_{ATP}), and estimated NADPH maintenance cost from Pareto optimisation (M_{NADPH}).

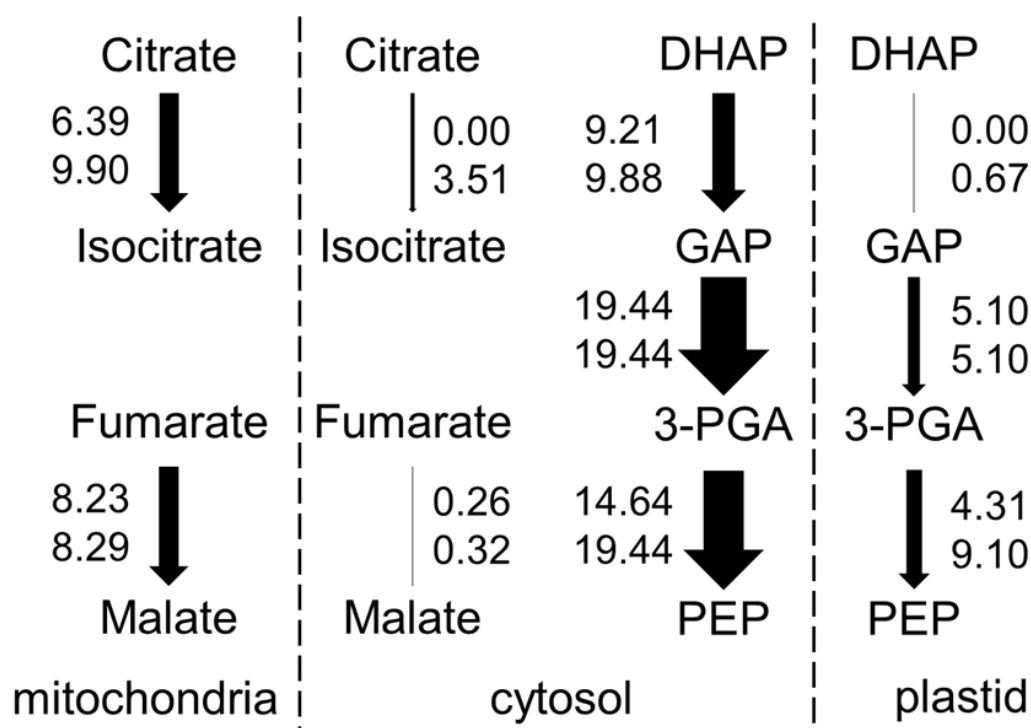


Figure 4.5 Flux ranges for glycolysis, aconitase and fumarase predicted using the objectives of minimisation of total flux (Min Flux) for the heterotrophic *Arabidopsis* cell culture under control conditions.

Thickness of the arrow represents the mean flux value for each reaction. Flux ranges are shown beside the arrows with the top and bottom values being the minimum and maximum flux values respectively. Glycolysis was predicted to operate in both cytosol and plastid whereas fluxes through aconitase and fumarase primarily occur in the mitochondria. Abbreviations: DHAP – dihydroxyacetone phosphate; GAP – glyceraldehyde 3-phosphate; 3-PGA – glycerate 3-phosphate; PEP – phosphoenolpyruvate.

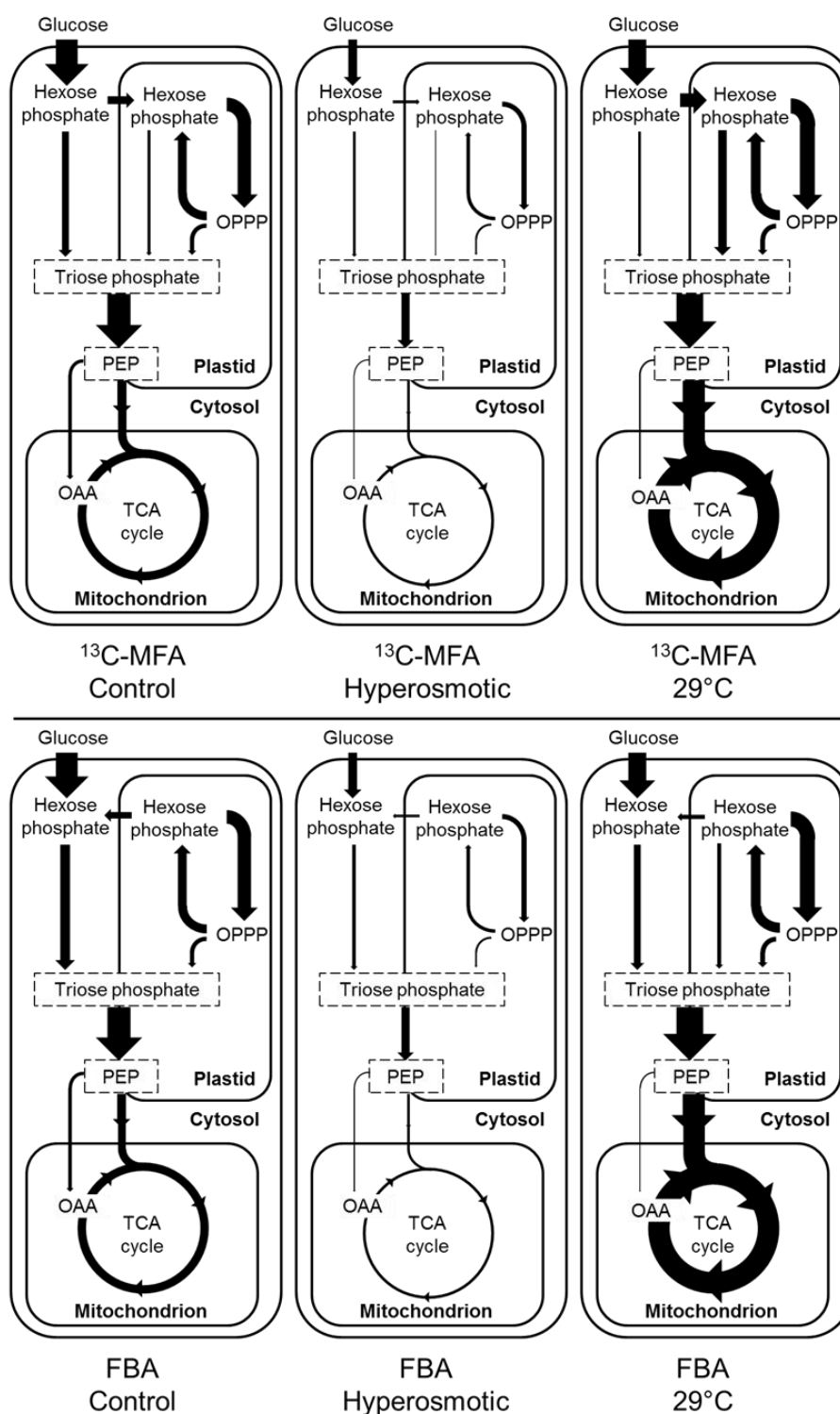


Figure 4.6 Comparison of measured and predicted flux maps for *Arabidopsis thaliana* cells under control conditions, hyperosmotic stress and elevated temperature (29°C). Flux maps were either determined experimentally by ^{13}C -MFA or predicted by FBA using the same biomass constraints, a comprehensive analysis of the transport costs, and the measured value of the flux ratio between the oxidative steps of the OPPP and glycolysis.

Table 4.1 Flux ranges for the reactions catalysed by G6PDH, NADP-GAPDH and NADP-MDH. NADP-GAPDH and/or NADP-MDH were set as reversible or irreversible in the direction of NADPH consumption. Flux ranges (in $\text{mmol d}^{-1} \text{L}^{-1}$ culture) are represented as $(v_{\min}, v_{\max})$, where $v_{\min}$ and $v_{\max}$ are the minimum and maximum feasible flux values compatible with the optimal objective value. G6PDH represents plastidial G6PDH with a positive flux corresponding to NADPH production. NADP-GAPDH and NADP-MDH represent the plastidial isoforms with a negative flux corresponding to NADPH production.

NADP-GAPDH reversibility NADP-MDH reversibility	Irreversible Irreversible	Irreversible Reversible	Reversible Irreversible	Reversible Reversible
G6PDH	(9.65, 9.65)	(0.28, 0.28)	(1.34, 1.34)	(0.74, 0.88)
NADP-GAPDH	(0, 0)	(0, 0)	(-21.67, -21.67)	(-21.80, -21.50)
NADP-MDH	(0, 0)	(-19.64, -18.73)	(0, 0)	(-1.07, -0.78)

Table 4.2 OPPP to glycolysis flux ratio, ATP and NADPH maintenance costs, and ATP and NADPH requirements for biosynthesis in Arabidopsis cell suspension culture under control and stress conditions. Values predicted from FBA and deduced from MFA flux maps are denoted with [FBA] and [MFA] respectively. Maintenance costs computed from FBA expressed as a percentage of total ATP or NADPH produced are shown in parentheses. Details of calculations for maintenance costs from MFA flux maps are shown in Appendix S8. Maintenance costs and coenzyme requirements for biosynthesis are expressed in mmol L⁻¹ culture h⁻¹.

	Control	Hyperosmotic	29°C
OPPP : glycolysis flux ratio [MFA]	0.591	0.637	0.563
ATP maintenance cost [FBA] (as % of total ATP produced)	54.532 (33.3)	7.046 (13.0)	252.608 (79.0)
ATP maintenance cost [MFA]	47.052	4.271	262.741
NADPH maintenance cost [FBA] (as % of total NADPH produced)	19.167 (49.6)	5.881 (41.8)	48.786 (89.7)
NADPH maintenance cost [MFA]	18.530	5.482	48.915
NADPH maintenance cost / ATP maintenance cost [FBA]	0.351	0.835	0.193
NADPH maintenance cost / ATP maintenance cost [MFA]	0.394	1.284	0.186
OPPP : glycolysis flux ratio for biosynthesis [FBA]	0.519	0.539	0.501
ATP requirement for biosynthesis [FBA]	101.236	45.664	28.163
NADPH requirement for biosynthesis [FBA]	19.529	8.239	5.611
NADPH requirement / ATP requirement for biosynthesis [FBA]	0.193	0.180	0.200

Table 4.3 Effect of stress treatments on cell culture growth parameters. Glc consumption rate, final biomass (after 5 days of growth), growth rate (biomass accumulation) and carbon conversion efficiency (CCE) of Arabidopsis cell suspension cultures grown under control conditions (21°C), under hyperosmotic conditions (171 mM mannitol in culture medium) and under high temperature (29°C). Values are the mean of 4 replicates $\pm$ SD.

	GLC consumption (mmol/d/L)	Biomass @ day5 (gDW/L)	Growth (gDW/L/day)	CCE
Control	22.535 $\pm$ 2.169	7.5 $\pm$ 0.5	2.35 $\pm$ 0.31	64.1%
Hyperosmotic	9.103 $\pm$ 0.455	4.5 $\pm$ 0.5	0.90 $\pm$ 1.09	69.1%
High temperature	18.551 $\pm$ 0.943	9.0 $\pm$ 1.1	0.76 $\pm$ 0.67	22.4%

5 Modelling metabolism of C₃ and CAM leaves over a day-night cycle

5.1 Introduction

5.1.1 The use of flux balance analysis in modelling photosynthetic metabolism

Photosynthetic metabolism has been studied in great detail over the past decades due to its importance in agriculture (Murchie et al., 2008) and the global carbon cycle in relation to climate change (Smith and Dukes, 2013). The metabolic pathways and enzymes involved in photosynthesis and related metabolic processes such as the synthesis of sucrose and starch have been well characterised. However, how the metabolic fluxes through central photosynthetic pathways and their integration with other parts of the metabolic network is less well understood. In the last few years, metabolic fluxes under phototrophic conditions have been predicted in several photosynthetic organisms using flux balance analysis (FBA). These include *Synechocystis* (Knoop et al., 2010; Montagud et al., 2010; Nogales et al., 2012; Saha et al., 2012; Knoop et al., 2013), *Chlamydomonas* (Boyle and Morgan, 2009; de Oliveira Dal'Molin, 2011), *Arabidopsis* (de Oliveira Dal'Molin et al., 2010a), rapeseed embryos (Hay and Schwender, 2011b), rice (Poolman et al., 2013), maize (Saha et al., 2011) and three C₄ plants (de Oliveira Dal'Molin et al., 2010b). These models successfully predicted the metabolic routes involved in CO₂ fixation into different biomass components in the presence of a light input. However, one major feature of metabolism of photosynthetic organisms, namely the interactions between light and dark metabolism, was neglected in all but one of the published FBA studies. In

leaves, storage compounds such as starch accumulate in the light to support dark metabolism whereas some metabolites such as nitrate and citrate accumulate in the dark (Scheible et al., 2000). In a recent publication by Knoop et al., (2013), the metabolic fluxes in *Synechocystis* over a day-night cycle were predicted by dividing the diel cycle into 192 steps with glycogen being the only storage compound. Time-courses of metabolic flux predictions over a diel cycle were simulated with the use of a time-dependent biomass objective function. By pre-defining the synthesis and degradation of storage compound as a constraint and/or in the biomass objective function, this approach does not give the model the freedom to explore the interactions between light and dark metabolism. In this chapter, a novel diel modelling framework was developed in which the metabolic fluxes in the light phase and the dark phase were simulated simultaneously in a single optimisation problem, as opposed to the published FBA studies on photosynthetic organisms where light and dark metabolism were simulated independently. With the inclusion of a pre-defined set of storage compounds, which can freely accumulate over the day-night cycle, the diel modelling framework was used to explore the interactions between light and dark metabolism and to predict the metabolic fluxes in a photosynthetic leaf.

5.1.2 C₃ and Crassulacean acid metabolism (CAM) photosynthesis

Most land plants are C₃ plants in which the primary photosynthate is a 3-carbon compound, namely 3-phosphoglycerate (3PGA). C₃ plants fix CO₂ in the light period by the enzyme ribulose 1,5-bisphosphate carboxylase oxygenase (Rubisco) which combines a 5-carbon compound, ribulose 1,5-

bisphosphate (RuBP), and CO_2 to form two molecule of 3PGA. While one-sixth of the 3PGA produced by Rubisco can be used for the synthesis of starch, sucrose and other metabolites, the remaining five-sixths of the 3PGA must be recycled back into RuBP through the reductive pentose phosphate pathway, to provide the substrate for Rubisco. The recycling of 3PGA back into RuBP is an energy- (and electron-) requiring process. Energy in the form of ATP and electrons carried by NADPH are produced by the thylakoid ATP synthase and the photosynthetic linear electron transport, where the ultimate sources of energy and electrons are sunlight and water, respectively. In C_3 photosynthesis, capture of sunlight and net carbon fixation both occur during the light period. Stomata in C_3 leaves are opened during the day to allow CO_2 to diffuse into the mesophyll cells. However, the opening of stomata causes water loss by transpiration, so there is an evolutionary trade-off between the rate of CO_2 diffusion into the mesophyll cells and the rate of water loss.

Some lineages of land plants have evolved to tackle the problem of water loss by closing their stomata during the day when the temperature is high. These are plants which undergo Crassulacean acid metabolism (CAM). The metabolism of CAM photosynthesis over a diel cycle (the CAM cycle) can be separated into four phases as defined in Osmond (1978). In phase I, CO_2 is combined with phosphoenolpyruvate (PEP) to form oxaloacetate by PEP carboxylase during the night. In most CAM plants, oxaloacetate is converted to malate, which is stored in the vacuole as malic acid at night. Phase II corresponds to the early light period when there is a transient peak in CO_2 fixation, believed to be caused by both PEP carboxylase and Rubisco, while the stomata are closing. In phase III, the stored malic acid is released from

the vacuole and decarboxylated to release CO_2 during the main part of light period. There are two subtypes of CAM with distinct decarboxylating enzymes: PEPCK-CAM, which uses malate dehydrogenase and PEP carboxykinase (PEPCK), and ME-CAM, which uses malic enzyme (ME) and pyruvate, Pi dikinase, to convert malate into PEP. With the stomata being closed in CAM plants in the light period during phase III, there is a build-up of internal CO_2 which is reassimilated by Rubisco and the rest of the reductive pentose phosphate pathway as in the CO_2 fixation pathway in C_3 photosynthesis. To complete the CAM cycle, a sufficient amount of carbon must be stored during the day to provide the substrate for carbon fixation (i.e. PEP) at night. Some CAM plants primarily store starch during the day whereas some preferentially store soluble sugars in the vacuole. Lastly, phase IV occurs in the final part of the light period when the nocturnally stored malate is fully decarboxylated before dusk, in which case the stomata open and CO_2 fixation occurs via Rubisco and the reductive pentose phosphate pathway as in C_3 plants.

While the energy required for carbon fixation from sunlight is captured during the day in both C_3 and CAM plants, the fundamental difference between C_3 and CAM photosynthesis is that net carbon fixation occurs during the day in C_3 and during the night in CAM. Due to the temporal separation between the harvesting of sunlight and net carbon fixation, the interactions between light and dark metabolism is especially important in CAM plants. To simulate the complete CAM cycle, the diel modelling framework is necessary as the CAM cycle is strongly dependent on the accumulation and degradation of storage compounds over the diel cycle. In this chapter, the metabolic fluxes in C_3 and

the different subtypes of CAM leaves were predicted using the diel modelling framework. The flux predictions of C_3 and CAM were compared to provide insights into the evolutionary driving force for CAM photosynthesis.

5.2 Results

5.2.1 Diel modelling framework for modelling leaf metabolism

Leaf metabolism undergoes dramatic changes over a day-night cycle in response to the availability of sunlight. Leaf metabolism over a day-night cycle can be broadly divided into two phases, a photosynthetic phase in the light and a heterotrophic phase in the dark. Thus far, FBA studies on photosynthetic organisms were implemented by simulating each phase independently as individual steady-states. This neglects the interactions between the two phases of diel metabolism. In fact, a steady-state corresponding to the light phase without any interaction with the dark phase can be interpreted as the metabolic state when the cell is grown under constant light. Though such a scenario is often found in experimental setups, photosynthetic organisms are usually subjected to a day-night cycle in their natural environment. To capture this feature, a novel modelling framework was developed to simulate the two phases of leaf metabolism over a diel cycle (see section 2.3 for details).

The diel modelling framework assumes each phase to be in a pseudo-steady-state, i.e. no change in metabolic fluxes within each phase. In this study, leaf metabolism over a day-night cycle was simulated as two inter-dependent steady-states, representing a light phase and a dark phase

(Figure 5.1). Photon influx was unconstrained in the light phase but was set to zero in the dark phase. The metabolic interactions between the two phases were captured by the accumulation and breakdown of storage compounds. In the case of a leaf, the storage compounds considered in this study include starch, glucose, fructose, malate, fumarate, citrate and nitrate, which are free to accumulate in both the light and the dark phase (Figure 5.1). Based on experimental measurements of diel fluctuations of amino-acids in leaves (Scheible et al., 2000), the 20 common amino-acids are allowed to accumulate only in the light phase to avoid the cycling of amino-acids between the light phase and dark phase as the main route of carbon storage. The rates of production and consumption of storage compounds were assumed to be constant in each phase with no net production or consumption over a day-night cycle. The two phases were modelled simultaneously in a single optimisation problem analogous to modelling interactions between two micro-organisms (Stolyar et al., 2007; Zhuang et al., 2011) or two tissues/cell-types in multi-cellular organisms (Lewis et al., 2010; Bordbar et al., 2011; de Oliveira Dal'Molin et al., 2010b; Saha et al., 2011).

In this chapter, the metabolism of mature non-growing leaves was modelled by constraining the model with sucrose and amino-acids export to the phloem and cellular maintenance costs (Figure 5.1). Based on measurements on the composition of *Arabidopsis* phloem exudate (Wilkinson and Douglas, 2003), the composition of 18 amino-acids exported into the phloem was constrained (Table 5.1) and the ratio of sucrose to amino-acids export was set to 2.2:1. Nitrate was assumed to be the sole nitrogen source. While it is possible to adjust the light period in the diel modelling framework by scaling the fluxes

with the length of each phase, a 12 hour light-12 hour dark was assumed in this chapter for simplicity. The ratio of total export to the phloem during the light phase and the dark phase was constrained to be 3:1 based on measurements in *Arabidopsis* rosettes (Gibon et al., 2004). The ratio of nitrate import during the day and the night was constrained to be 3:2 taken from experimental studies in various plant species (Delhon et al., 1995; Macduff and Bakken, 2003; Siebrecht et al., 2003). Besides the export of sucrose and amino-acids, cellular maintenance costs were included as constraints by setting a fixed maintenance cost so that half of the carbon stored in the light phase was released as CO₂ in the dark phase, calculated based on the carbon balance from measurements in *Arabidopsis* rosettes (Gibon et al., 2004). The ratio of ATP maintenance cost to NADPH maintenance cost was assumed to be 3:1 taken from the value estimated in chapter 4 for heterotrophic cells under control condition. ATP for maintenance was considered to be required only in the cytosol whereas maintenance NADPH was assumed to be required equally in the cytosol, the plastid and the mitochondrion as in chapter 4. The cellular maintenance cost was assumed to be the same in the light phase and the dark phase. In addition to the above constraints, fluxes through chloroplastic NADPH dehydrogenase and plastoquinol oxidase were set to zero as the contributions of chloroplastic NADPH dehydrogenase (Yamamoto et al., 2011) and plastoquinol oxidase (Josse et al., 2000) to photosynthesis are thought to be very minor. The flux of sucrose synthase was constrained to be irreversible in the direction of sucrose degradation as it is thought that

sucrose synthase is involved in the cleavage of sucrose but not sucrose synthesis in leaves (Nguyen-Quoc et al., 1990).

5.2.2 Comparison of predictions from diel modelling framework and single steady-state model

In the published FBA studies on photosynthetic metabolism, the photosynthetic phase was simulated in single steady-state models independent of the heterotrophic phase. To investigate the difference in predictions from the diel modelling framework and the single steady-state model, comparisons were made between metabolic fluxes of a mature *Arabidopsis* leaf predicted from the two approaches using the *Arabidopsis* genome-scale metabolic model. As *Arabidopsis* is a C₃ plant, the ratio of Rubisco carboxylation to oxygenase was set to 3:1 to simulate photorespiration as in a published FBA study on *Arabidopsis* (de Oliveira Dal'Molin et al., 2010a). Flux through the chloroplastic hexose phosphate transporter was constrained to zero as no hexose phosphate transport activity was detectable despite the presence of gene expression of chloroplastic hexose phosphate transporter in mesophyll cells of *Arabidopsis* leaves (Niewiadomski et al., 2005). The constraints for the diel modelling framework were described in the previous section. The total sucrose and amino-acids export and cellular maintenance costs over a 24-hour period in the diel model were used as constraints in the single steady-state model. The diel model and single steady-state model can be thought of as 12/12 and 24/0 light-dark period respectively, with the same sucrose and amino-acids export and cellular maintenance costs over a 24-hour period.

Flux variability analysis with the primary objective function of minimisation of total flux was applied to compute the flux ranges from the two modelling approaches, the diel model and the single steady-state model. By comparing the flux ranges of reactions in the light, 131 reactions were identified with non-overlapping flux ranges, i.e. carry different fluxes (Appendix S9). Out of the 131 reactions, none carries flux in opposite directions from the two approaches. A quantitative measure of the similarity between the fluxes of a reaction from the two approaches was devised (see chapter 2 for details), which in simple term is the fraction of the smaller flux value over the larger flux value. Reactions with a similarity measure of less than 0.25 were listed in Table 5.2, where the smaller the value of this measure, the larger the difference in fluxes. A similarity measure of zero means that the reaction carries zero flux from one of the modelling approaches. A full list of reactions with non-overlapping flux ranges and their similarity measures can be found in Appendix S9.

Reactions that carry high flux in the diel model and low flux in the single steady-state model are involved in starch synthesis and storage in the vacuole (Table 5.2A). The reason for these reactions carrying high flux only in the diel model is that they are involved in supporting metabolism in the dark, which is absent in the single steady-state model.

The most profound finding from this analysis is the differences in the source of citrate for glutamate and glutamine synthesis in the light. By looking at reactions that carry high flux in the single steady-state model and low flux in the diel model (Table 5.2B), the single steady-state model predicted that the synthesis of citrate was carried out by a metabolic route involving threonine

aldolase (Figure 5.2, left). In this metabolic route, threonine is produced from aspartate which is derived from 3-phosphoglycerate via PEP and oxaloacetate. Threonine aldolase converts threonine into acetaldehyde, which is subsequently converted into acetyl-CoA by acetaldehyde dehydrogenase and acetyl-CoA synthetase, and glycine, which is metabolised via the photorespiration pathway. Acetyl-CoA from acetaldehyde is then used to produce citrate for glutamate synthesis via 2-oxoglutarate. With the diel modelling framework, the model predicted that nocturnally stored citrate is converted into 2-oxoglutarate in the light for glutamate synthesis (Figure 5.2, right). This is consistent with observations from isotopic labelling experiments that the source of citrate in the light in photosynthetic leaves is mainly the citrate pool stored during the night (Tcherkez et al., 2009). This example clearly demonstrated the importance of modelling light and dark metabolism simultaneously in photosynthetic flux balance models.

5.2.3 Predictions of metabolic fluxes in a mature C₃ leaf over a day-night cycle

By applying the diel modelling framework on Arabidopsis genome-scale metabolic model with the constraints described in section 5.2.1 and section 5.2.2 (Figure 5.1), a flux prediction of metabolism in a mature Arabidopsis leaf over a day-night cycle was computed using FBA with the objective function of minimising total flux (Appendix S10). The resulting flux solution is illustrated in a simplified flux map (Figure 5.3). The light and dark phases are represented as the left and right half of the flux map respectively, where storage compounds are illustrated in the middle between the two phases.

The diel modelling framework successfully predicted the accumulation of starch in the light and the breakdown of starch in the dark, illustrated as an arrow pointing from the light phase to the dark phase in Figure 5.3, even when the direction of accumulation was set to be free in the model. While starch was predicted to be the major stored compound, malate was also predicted to accumulate in the light, which feeds into the TCA cycle in the dark. Interestingly the model predicted the accumulations of citrate and nitrate in the dark to support amino-acids synthesis in the light. Citrate is metabolised in the light to provide carbon skeleton for amino-acids synthesis. A proportion of amino-acids synthesised in the light accumulates in the cytosol which is then exported into the phloem in the dark. Nitrogen assimilation was predicted to occur primarily in the light where nitrate imported from the xylem in the dark is stored in the vacuole and released in the light. This is consistent with the reversible phosphorylation of a leaf nitrate reductase (Huber et al., 1992) which rapidly reduces in activity at sudden darkness (Riens and Heldt, 1992).

In the light phase, the model predicted large fluxes through the photosynthetic linear electron transport and the reductive pentose phosphate pathway to support sucrose export and starch accumulation. The classical photorespiratory pathway was also predicted to be active given that the Rubisco carboxylation to oxygenase ratio was set to 3:1. The TCA cycle was predicted to operate in a non-cyclic mode in the light with two separate branches (Figure 5.4). Citrate stored in the dark is converted into 2-oxoglutarate for glutamate and glutamine synthesis. Oxaloacetate produced from PEP carboxylase is converted into malate which accumulates in the

light. A small flux was predicted through fumarase which converts fumarate, produced as a by-product in arginine synthesis, into malate. Other reactions in the TCA cycle were predicted to carry no flux in the light, which is consistent with stable isotopic labelling measurements that there is very low flux through the section of the TCA cycle from 2-oxoglutarate to fumarate in illuminated leaves (Tcherkez et al., 2009). The model predicted that the mitochondrial electron transport chain (ETC) is active in the light, carrying 9.5% of the flux of photosynthetic linear electron transport. The mitochondrial ATP synthase was predicted to contribute to 18% of total ATP synthesis in the light with the rest produced by the chloroplast thylakoid ATP synthase. The two sources of NADH feeding into the mitochondrial ETC are glycine decarboxylase and the malate-oxaloacetate (OAA) shuttle (Figure 5.5) with contributions of 53% and 47% respectively. The ultimate source of reductant in a photosynthetic leaf is from the photosynthetic linear electron transport. While most of the reducing power is used to fix CO₂, a small proportion of reductant is shuttled out of the chloroplast by both the malate-OAA shuttle and the 3-phosphoglycerate (3PGA)-glyceraldehyde 3-phosphate (GAP) shuttle (Figure 5.5). The malate-OAA shuttles transport reductant from the chloroplast into the cytosol, the mitochondria and the peroxisome. The 3PGA-GAP shuttle transports reductant from the chloroplast to the cytosol to produce NADPH by the non-phosphorylating NADP-GAPDH. A proportion of NADPH produced in the cytosol is shuttled into the mitochondria for maintenance processes by the isocitrate-oxoglutarate shuttle (Figure 5.5).

In the dark phase, the largest fluxes are related to starch degradation, sucrose synthesis, glycolysis, the oxidative pentose phosphate pathway

(OPPP), the TCA cycle and the mitochondrial electron transport chain (Figure 5.3). The breakdown of starch was predicted to be carried out by the hydrolytic pathway via maltose which can be transported from the chloroplast to the cytosol. This is the result of constraining the chloroplastic hexose phosphate transporter to carry zero flux which limits the possible metabolic routes for carbon export from the chloroplast. Glycolysis and the OPPP run in both the cytosol and the chloroplast, with 72% of the total GAPDH (enzyme involved in glycolysis) flux and 55% of glucose 6-phosphate dehydrogenase (enzyme involved in the OPPP) flux in the cytosol. The TCA cycle operates in a cyclic mode in the dark (Figure 5.4), consuming pyruvate produced from stored starch via glycolysis. A small proportion of citrate produced by citrate synthase is stored in the dark to provide carbon skeleton for glutamate and glutamine synthesis in the light.

5.2.4 Predictions of metabolic fluxes in a mature CAM leaf over a day-night cycle

From a structural metabolic modelling perspective, the Arabidopsis genome-scale metabolic model contains all the metabolic reactions required to carry out the various subtypes of CAM photosynthesis. Applying the diel modelling framework to the Arabidopsis genome-scale model, the metabolic fluxes in a mature CAM leaf over a 24-hour cycle were simulated. The CAM cycle can be separated into four phases as defined in Osmond, (1978). Phase I and phase III correspond to the dark period and the main part of light period, respectively, whereas phase II and phase IV are the transitional periods at the beginning and end of the light period, respectively. In this study, only

phase I and phase III were simulated for simplicity where phase II and phase IV can be included in an extended diel modelling framework in future studies.

To simulate the metabolic fluxes in a mature CAM leaf, the model was constrained by sucrose and amino-acid export, cellular maintenance costs and other constraints as described in section 5.2.1. The main additional constraint for simulating CAM photosynthesis was that the CO₂ exchange with the environment was set to have zero flux in the light phase to simulate the closure of stomata in CAM plants during the day. While the Rubisco carboxylase to oxygenase ratio was set to 3:1 in simulating C₃ photosynthesis, Rubisco oxygenase activity is likely to be suppressed in phase III of CAM due to the increase in internal partial pressure of CO₂. However, it is unclear to what extent photorespiration occurs in phase III of CAM (Lüttge, 2011), therefore the Rubisco carboxylase to oxygenase ratio was left unconstrained in the light phase and was set to 3:1 in the dark phase when the stomata are opened. The chloroplastic hexose phosphate transporter was constrained to zero flux for modelling Arabidopsis leaves based on the lack of activity in Arabidopsis mesophyll cells (Niewiadomski et al., 2005). However, in the ice plant *Mesembryanthemum crystallinum* L., a facultative CAM plant, chloroplasts of CAM-induced plants mainly export glucose 6-phosphate (G6P) which is hardly exported in chloroplasts of C₃ plants (Neuhaus and Schulte, 1996). The chloroplastic G6P transporter activity in *M. crystallinum* L. was significant in CAM-mode but undetectable in C₃-mode (Kore-eda and Kanai, 1997). Also, there was a 71-fold increase in transcript level for the chloroplastic G6P transporter when ice plants switch from C₃ to CAM (Cushman et al., 2008). Therefore, the chloroplastic G6P

transporter was set to be unconstrained in modelling CAM metabolism. The metabolic fluxes in a generic CAM leaf was simulated by applying the CAM-specific constraints described above. The various subtypes of CAM photosynthesis were simulated by setting the reactions specific to other subtypes to carry zero flux. The constraints applied for modelling the different modes of photosynthesis are summarised in Table 5.3.

Using the Arabidopsis genome-scale metabolic model and the diel modelling framework with CAM-specific constraints, the model successfully predicted metabolic fluxes consistent with the well-known CAM cycle (Appendix S10). In the light phase (phase III of CAM), nocturnally stored malate is decarboxylated by malic enzyme or PEPCK depending on the subtype. In the generic CAM model, 89% of malate decarboxylation was predicted to be carried out by PEPCK, with the remaining 11% carried out by the cytosolic malic enzyme. In addition to the photosynthetic linear electron transport and reductive pentose phosphate pathway, which are expected to carry high fluxes in the light, the gluconeogenesis pathway also carries a high flux in the light to convert PEP, from malate decarboxylation, to starch or soluble sugars for storage as part of the CAM cycle. In the generic CAM model, only starch is stored during the day but not soluble sugars. In the sugar-storing subtypes, the model predicted that fructose is produced in the cytosol by sucrose synthase and subsequently transported into the vacuole by the vacuolar hexose transporter. However, the main function of the vacuolar hexose transporter in sugar-storing CAM plants is thought to be the efflux of hexose during the dark period (Holtum et al., 2005). It is suggested that the vacuolar hexose transporter must be regulated to restrict efflux of vacuolar hexose to

the dark period and avoid futile cycling during the light period, though the mechanism for such regulation remains unknown (Holtum et al., 2005; Antony et al., 2008; Antony and Borland, 2009). When the vacuolar hexose transporter activity was restricted to the dark period, by constraining the transporter to carry zero flux during the light period, the model predicted that sucrose is imported into the vacuole and is broken down into glucose and fructose by the vacuolar invertase, which is in line with the suggestions by Smith and Bryce (1992), McRae et al. (2002) and Holtum et al. (2005). For subsequent modelling, the vacuolar hexose transporter was set to zero flux in the light phase.

The Rubisco carboxylase to oxygenase ratio was unconstrained in modelling CAM and the model predicted no Rubisco oxygenase flux suggesting that Rubisco oxygenase reaction is not necessary for the functioning of a mature CAM leaf under the assumption of minimising total flux. The mitochondrial ETC and ATP synthase also carry high fluxes in the light with the mitochondrial ETC carrying 29% of the flux through the photosynthetic linear electron transport and the mitochondrial ATP synthase contributing to 34% of the total ATP produced in the light in the generic CAM model. Similar values were predicted for the various CAM subtypes. An incomplete TCA cycle was predicted in the light phase with two separate branches, citrate to 2-oxoglutarate and succinate to oxaloacetate (Figure 5.6). Interestingly, while some of the nocturnally stored citrate is used to provide carbon skeleton for glutamate and glutamine synthesis in the light, a proportion of isocitrate from the stored citrate is metabolised by isocitrate lyase in the peroxisome to produce succinate, which feeds into the TCA cycle, and glyoxylate, which

feeds into the photorespiratory pathway (Figure 5.6). It is known that some CAM plants accumulate large amount of citrate during the night (Lüttge, 1988; Borland and Griffiths, 1989). The model prediction hypothesises a previously unknown role of isocitrate lyase in metabolising the nocturnally stored citrate in CAM plants. With respect to nitrogen assimilation, it is predicted to primarily occur in the light phase as in C_3 . The model predicted the use of different electron donors for nitrate reductase in the various subtypes of CAM. Nitrate reductase in PEPCK subtypes uses NADH whereas NADPH is utilised in ME subtypes. This can be explained by the differences in the decarboxylation of malate in the two subtypes. PEPCK subtypes were predicted to convert malate into oxaloacetate by NAD-dependent malate dehydrogenase whereas the cytosolic NADP-dependent malic enzyme was involved in converting malate into pyruvate in ME-subtypes.

In the dark phase, the stored carbon (starch or soluble sugars) is mobilised to produce PEP, the substrate for carboxylation, via glycolysis. PEP is carboxylated by PEP carboxylase to produce oxaloacetate, which is subsequently converted into malate that is stored in the vacuole as malic acid during the night. With respect to starch degradation, the phosphorylating route was predicted in CAM as a result of the presence of the chloroplastic G6P transporter. Starch is metabolised by starch phosphorylase to produce hexose phosphate in the chloroplast. The hexose phosphate is then transported into the cytosol by the chloroplastic G6P transporter, which is in line with the suggested model by Weise et al. (2011). While most of the carbon storage is used to provide the substrate for carboxylation in the CAM cycle, some goes into the OPPP, to produce NADPH for maintenance

processes, and the TCA cycle, to produce NADH for ATP synthesis via the mitochondrial ETC and ATP synthase. As in a mature C₃ leaf, the model predicted a complete TCA cycle in the dark in a mature CAM leaf (Figure 5.6).

5.2.5 Quantitative comparisons between C₃ and various subtypes of CAM

CAM is a CO₂-concentrating mechanism where CO₂ is first fixed into a four carbon compound and stored during the night. During the day when the stomata are closed, the four carbon compound is broken down to produce CO₂, which results in an increase in internal partial pressure of CO₂ in a CAM leaf and hence a potential reduction in Rubisco oxygenase activity compared to a C₃ leaf. The CO₂-concentrating mechanism in CAM is an energy-requiring process where ATP is needed for the decarboxylation of malate to PEP, the production of starch and/or sugars as storage during the day and the transport of storage metabolites into the vacuole. The energetic costs required to support sucrose and amino-acids export and cellular maintenance processes in a mature C₃ leaf and various subtypes of CAM leaves were compared in terms of their photon use (Figure 5.7). The photon use in a mature CAM leaf is similar to that in a C₃ leaf, $\pm 10\%$ of C₃ depending on the CAM subtype. This is consistent with the calculations by Winter and Smith (1996) which suggested that the energetic cost of CO₂ fixation was about 10% higher in CAM plants than in C₃ plants.

Rubisco is the most abundant protein on earth, contributing up to 50% of the soluble protein and 20-30% of the total nitrogen in a C₃ leaf (Feller et al., 2008). The investment in Rubisco, in terms of energetic and nitrogen, is likely to contribute significantly to the evolutionary fitness trade-off in plants. By

comparing the model predictions, the Rubisco carboxylase flux in CAM was found to be similar to that in C_3 (Figure 5.7). This result could be explained by the fact that the model was constrained with the same sucrose and amino-acids export and cellular maintenance costs. In contrast, the Rubisco oxygenase was predicted to carry zero flux in all CAM subtypes given that the Rubisco carboxylase to oxygenase ratio was unconstrained. As a result, the total flux through Rubisco in a mature CAM leaf was predicted to be 21-32% lower than in a C_3 leaf (Figure 5.7). Although CAM requires less Rubisco than C_3 due to repression of photorespiration, the CO_2 -concentrating mechanism requires high fluxes through the CAM cycle. The total flux through the metabolic network is 32-64% higher in CAM than in C_3 (Figure 5.7).

5.3 Discussion

5.3.1 Applications of diel metabolic modelling framework in FBA studies

In this chapter, a diel modelling framework was developed to model light and dark metabolism and their interactions in photosynthetic leaves. The framework allowed the metabolism over a full diel cycle to be modelled simultaneously in a single optimisation problem as opposed to the published FBA studies on photosynthetic metabolism. One of the main advantages of the diel modelling framework over the conventional single steady-state FBA is that the interactions between light and dark metabolism can be modelled by the accumulation and degradation of storage compounds. The diel modelling framework and the single steady-state approach gave different model predictions illustrated by the example of the source of citrate for

glutamate and glutamine synthesis in the light (Figure 5.2). The diel modelling framework predicted the use of nocturnally stored citrate as the source of carbon skeletons for glutamate and glutamine synthesis in the light (Figure 5.2, right), which matches well with the results from isotopic labelling experiments in illuminated *Xanthium strumarium* leaves (Tcherkez et al., 2009). A curious metabolic route involving threonine aldolase was predicted from the single steady-state model (Figure 5.2, left). In another Arabidopsis genome-scale metabolic model (AraGEM), the threonine aldolase reaction is absent and the model predicted that citrate is produced from pyruvate by pyruvate dehydrogenase and citrate synthase (de Oliveira Dal'Molin et al., 2010a). However, this contradicts experimental evidence that the mitochondrial pyruvate dehydrogenase is inhibited transcriptionally and post-transcriptionally in leaves in the light (Tovar-Mendez et al. 2003). The mismatches between the predictions from the single steady-state approach and experimental results are likely to be a result of excluding the interactions with dark metabolism in the model. Another example light and dark metabolic interaction is nitrogen assimilation. Considering the source of amino-acids for export during the night, the model has the freedom to choose between storing amino-acids during the day and fixing nitrogen to produce amino-acids during the night. The model predicted the former with nitrogen assimilation primarily occurs in the light, which is consistent with the known light-dependent regulation of nitrate reductase in leaves (Huber et al., 1992; Riens and Heldt, 1992). The results in this chapter demonstrated the importance of modelling light and dark metabolism simultaneously in photosynthetic flux balance models and highlighted the potential danger in

modelling single metabolic state in isolation without considering its interactions with other metabolic states.

One limitation of the diel modelling framework is that it requires a pre-defined set of storage compounds. In this study, the choice of storage compounds was made based on known biochemistry of C₃ and CAM plants. An alternative to this is to use experimental data, e.g. time-course data over metabolites level, to constrain the accumulation and degradation of metabolites over the diel cycle. The diel cycle can be divided into multiple steady-states similar to the approach used in Knoop et al., (2013). The changes in absolute metabolite levels can be converted into input and output fluxes as constraints for each steady-state. By pre-defining the accumulation and degradation of storage compound(s), a time-series of metabolic flux solutions can be generated over a diel cycle. The choice of modelling approach depends on the aim of the investigation. For example, from an evolutionary perspective, the temporary carbon storage over a day-night cycle must satisfy certain criteria such as chemical stability, elemental composition (only include carbon, hydrogen and oxygen) and the energetic efficiency in synthesis and mobilisation. The diel modelling framework can potentially be used to explore the energetics of different storage compounds.

While the diel modelling framework improves the model prediction in *Arabidopsis*, it is necessary for modelling CAM photosynthesis as the CAM cycle takes place over the full diel cycle. In this study, the diel cycle was divided into two phases, a light phase and a dark phase. This allowed the metabolic fluxes in two of the four phases (phase I and phase III) in CAM photosynthesis to be modelled. The diel modelling framework with two

phases is the temporal analogy of modelling two interacting micro-organisms or two cell-types or tissues in multi-cellular organisms. In theory, the diel modelling framework can be extended to include more phases. The important parameters that set each phase distinct are the phase-specific constraints (e.g. photon influx) and the storage compounds between two consecutive phases. For example in guard cells, K^+ , malate and Cl^- accumulate from the beginning of the light period as the main osmolytes which are replaced by sucrose in the second half of light period (Talbot and Zeiger, 1998). Other than photosynthetic metabolism, this modelling framework can potentially be applied to model any system with periodic environmental and/or physiological changes such as the fast and feast states in liver and the active and rest states in muscle cells with glycogen being the storage compound.

5.3.2 Relating model flux prediction of a C_3 leaf with known biochemistry

Distinct features of light and dark metabolism predicted from the Arabidopsis genome-scale metabolic model using the diel modelling framework match well with the known metabolic modes in a C_3 leaf. The prediction of an incomplete TCA cycle flux in the light with two separate branches, from oxaloacetate to malate and from citrate to 2-oxoglutarate (Figure 5.4), is supported by experimental observations based on stable isotopic labelling measurements (Tcherkez et al., 2009). The citrate to 2-oxoglutarate branch is primarily involved in glutamate and glutamine synthesis. The oxaloacetate to malate branch is involved in producing malate which accumulates during the day. While the main function of the TCA cycle under heterotrophic conditions is to produce NADH for ATP generation by the mitochondrial

electron transport chain (ETC) and mitochondrial ATP synthase, the main source of reducing power and ATP is the photosynthetic ETC and chloroplastic ATP synthase in the photosynthetic condition. To minimise the energy for the re-fixation of CO₂, 2-oxoglutarate dehydrogenase and subsequent reactions in the TCA cycle up to malate were predicted to carry no flux in the light. Similarly, mitochondrial pyruvate dehydrogenase was also predicted to carry no flux in the light which is consistent with the light-dependent inhibition of the mitochondrial pyruvate dehydrogenase in leaves (Tovar-Mendez et al. 2003). The model predicted the accumulation of citrate in the dark as the source of carbon skeletons for glutamate and glutamine synthesis in the light. One possible explanation for preferring the accumulation of citrate in the dark over the synthesis of citrate from primary photosynthate in the light could be that the degradation of starch to produce citrate (via glycolysis, PEP carboxylase, pyruvate dehydrogenase and citrate synthase) provides a carbon-neutral route of transferring ATP and reducing power from the light phase to that dark phase. This conserves the net carbon required to support dark metabolism, where citrate in the dark would otherwise be metabolised through the TCA cycle and the carbon would be released as CO₂. To test this hypothesis, the metabolism of a mature C₃ leaf was simulated with the citrate accumulation constrained to zero flux. The model predicted a 2% increase in CO₂ release in the dark which is consistent with the carbon conservation hypothesis.

The mitochondrial electron transport chain was predicted to carry significant flux in the light which supports the notion that the mitochondrial electron transport chain is important for maintenance of photosynthetic carbon

assimilation (Raghavendra and Padmasree, 2003). However, under the assumption of minimising total flux, which also tends to minimise photon use, cyclic electron transport in the chloroplast electron transport chain and the chloroplast water-water cycle were predicted to carry no flux. An Arabidopsis cyclic electron transport mutant was shown to be only affected under high light condition (Munekage et al., 2008) suggesting that cyclic electron transport may not be important under constant low or moderate light intensity. However, cyclic electron transport could be important for acclimation to low light through state transition (Mullineaux and Emlyn-Jones, 2005). With respect to the water-water cycle, controversies exist as to whether it plays a significant role in photosynthetic metabolism in leaves under sunlight (Heber, 2002). The fluxes through these metabolic routes under varying light intensity will be investigated in chapter 6.

5.3.3 Predicting the metabolism in a CAM leaf using the diel modelling framework

The metabolic fluxes in leaves of various subtype of CAM, including the CAM cycle, were successfully simulated from the Arabidopsis genome-scale metabolic model. To the best of my knowledge, this is the first FBA study of the metabolism of CAM plants. Phase I and phase III of the CAM cycle, corresponding to the dark period and the light period respectively, were simulated using the diel modelling framework. The metabolism of phase II and phase IV can potentially be incorporated into an extended diel modelling framework in future investigations. In addition to C₃ and CAM, the diel modelling framework can be applied to model two related metabolic modes, 'CAM cycling' where net CO₂ fixation occurs during the day with refixation of

respiratory CO₂ during the night and 'CAM idling' where no net carbon fixation occurs with the stomata closed day and night under severe drought conditions (Cushman, 2001).

In this study, the metabolism in a mature CAM leaf was simulated by constraining the model with sucrose and amino-acid export and cellular maintenance costs as in a C₃ leaf. The major constraint for modelling CAM metabolism is that the CO₂ exchange with the external environment was set to zero during the day which represents the closure of stomata. This constraint forces the model to carry out net carbon fixation during the night. While the model has the freedom of choice for the use of carboxylating enzymes, PEP carboxylase and Rubisco, and the nocturnal carbon store, malate, fumarate, citrate, starch and soluble sugars, the conventional CAM cycle was predicted with PEP carboxylase as the carboxylating enzyme and malate as the main temporary storage of fixed carbon in the dark. In addition to malate, a small amount of citrate was predicted to accumulate during the night. In fact, some CAM plants accumulate large amount of citrate during the night (Lüttge, 1988; Borland and Griffiths, 1988). The model predicted that a proportion of the nocturnal stored citrate is metabolised by a metabolic route involving isocitrate lyase (Figure 5.6). Holtum et al. (2005) suggested two possible routes of catabolism of nocturnally stored citrate in CAM plants including the ATP citrate lyase and the TCA cycle. ATP citrate lyase produces oxaloacetate and acetyl-CoA from citrate, but since there is no sink for acetyl-CoA in the output constraints, this route was not chosen by the model. One possible reason for preferring isocitrate lyase over isocitrate dehydrogenase and 2-oxoglutarate dehydrogenase is that the conversion of

citrate to malate by the former route releases less CO₂ per molecule of citrate (glyoxylate produced from isocitrate lyase is metabolised through the photorespiratory pathway, producing 0.5 molecule of 3PGA and 0.5 molecule of CO₂ per molecule of citrate catabolised) than the latter route (2 molecules of CO₂ released per molecule of citrate). Based on this reasoning, ATP-citrate lyase may be preferred if there is a sink for cytosolic acetyl-CoA as it results in no carbon loss from citrate to oxaloacetate. In the absence of measurements of ATP citrate lyase and isocitrate lyase in CAM plants, it is not clear whether these two enzymes are involved in the catabolism of citrate *in vivo*.

In modelling a generic CAM leaf, where the model was free to choose between different decarboxylating enzyme and storage compounds, malate decarboxylation was predicted to be mainly carried out by PEPCK. This can be explained by the observation that the PEPCK-subtypes require fewer photons, which leads to less total flux through the metabolic network, than the ME-subtypes (Figure 5.7). Considering the biochemistry, the conversion of malate into PEP in the PEPCK-subtypes produces one NADH by malate dehydrogenase and consumes one ATP by PEPCK whereas in the ME-subtypes, one NAD(P)H is produced by ME and two ATP equivalents are consumed by pyruvate, Pi dikinase. From an energetics point of view, PEPCK is preferred over ME as the decarboxylating enzyme. However, it is questionable whether the difference in energetics plays any significant role in the evolution of the two subtypes as the availability of energy, in the form of sunlight, is likely not to be the limiting factor in CAM plants. As a consequence of the different coenzyme specificity of the two decarboxylating

enzymes, NAD(H) for PEPCK and NADP(H) for ME, the model predicted the use of different electron donors by nitrate reductase in the two subtypes of CAM. While the reduction of nitrate is attributed to the NADH-specific nitrate reductase in many plant tissues (Beever et al., 1964), NAD(P)H bispecific nitrate reductase has been reported in several plant species (Wells and Hageman, 1974; Dailey et al., 1982). It is likely that a NAD(P)H bispecific nitrate reductase exists in the leaves of ME-CAM species for the reduction of nitrate.

Starch was the preferred storage compound in the light over soluble sugars in the generic CAM model, irrespective of the constraint on the vacuolar hexose transporter. This is explained by the differences in photon use and Rubisco flux required (Figure 5.7). In terms of energetics, starch is broken down into hexose phosphate by starch phosphorylase in the dark, whereas soluble sugars in the form of hexoses stored in the vacuole require ATP to be converted into hexose phosphate. Therefore more carbon must be stored to produce the extra ATP in soluble sugars-storing CAM plants compared to starch-storing CAM plants. This is reflected by the differences in flux through Rubisco carboxylase (Figure 5.7). The carbon fixed by Rubisco in the light must either be exported into the phloem or stored. Given that the carbon exported is constrained to be identical in all simulations, the Rubisco carboxylase flux reflects the amount of carbon stored during the day. When simulating the metabolic fluxes in a soluble sugar-storing CAM leaf, hexose was initially predicted to be produced in the cytosol by sucrose synthase and imported into the vacuole by the vacuolar hexose transport. However, the vacuolar hexose transporter is suggested to be regulated to restrict efflux of

vacuolar hexose to the dark period and avoid futile cycling during the light period (Holtum et al., 2005; Antony et al., 2008; Antony and Borland, 2009). This example showed that the inability to predict regulatory aspect of metabolism is a potential limitation of FBA.

5.3.4 Comparison of model predictions of C₃ and CAM leaves

The major difference in metabolic fluxes between C₃ and CAM leaves is the carbon assimilation pathway with a high flux through the CAM cycle, as a CO₂-concentrating mechanism, on top of the core C₃ pathway in CAM plants. As a result of the CO₂-concentrating mechanism in CAM plants, which is modelled by setting the Rubisco carboxylase to oxygenase ratio to be unconstrained in the light, Rubisco oxygenase flux is suppressed in CAM compared to C₃. In the diel modelling framework, the model has the freedom to choose in which phase the storage compounds accumulate. This can be illustrated with the example of malate which was predicted to accumulate in the light in C₃ whereas in CAM, malate was predicted to be the main temporary storage of fixed carbon in the dark. From the decarboxylation of malate in CAM leaves in the light, large amounts of reductant are produced which leads to a relatively higher flux through the mitochondrial ETC and ATP synthase in CAM than in C₃ (the mitochondrial ETC carries 10% and 27% of the flux through the photosynthetic linear electron transport and the mitochondrial ATP synthase contributes to 16% and 35% of the total ATP produced in the light in C₃ and CAM leaves respectively). Another difference in metabolic fluxes between C₃ and CAM leaves is the pathway of transitory starch degradation. With the chloroplastic hexose phosphate transport flux

set to zero flux in C_3 but unconstrained in CAM, the model predicted the use of the hydrolytic pathway in C_3 , generating maltose from starch, whereas the phosphorolytic pathway was predicted in CAM which produces hexose phosphate in the chloroplast. This is consistent with the model proposed by Weise et al., (2011). However, the hydrolytic pathway could not be predicted in C_3 leaf without the chloroplastic hexose phosphate transport constraint. Instead, the model predicted the use of the phosphorolytic pathway in C_3 leaf when the chloroplastic hexose phosphate transport was unconstrained because it is energetically more efficient than the hydrolytic pathway (Lüttge et al., 1981) as extra ATP is required to convert maltose into hexose phosphate. The reason for C_3 plant in preferring the hydrolytic pathway is unknown and it has been speculated that it could be due to metabolic control (Weise et al., 2011) such as the buffering of carbohydrate availability by heteroglycan (Kartal et al., 2011). While FBA is powerful in predicting fluxes based on energetic efficiency, the examples of hexose storage synthesis in CAM (section 5.3.3) and transitory starch degradation suggest that FBA is unable to give accurate predictions when there is a trade-off between efficiency and regulation as information on metabolic regulation is not included in the model.

CAM is an ecophysiological metabolic adaptation which allows plants to inhabit arid environments and habitats with seasonal or intermittent water availability (Cushman, 2001; Silvera et al., 2010). A key adaption of CAM plants is the closure of stomata during the day which has a net effect of several-fold higher water-use efficiency for CAM plants than for C_3 and C_4 under comparable conditions (Drennan and Nobel, 2000). Together with

decarboxylation of malate during the day in the CAM cycle, another consequence of the stomata closure in CAM plants is an increase in CO₂ concentration around Rubisco which suppresses the Rubisco oxygenase activity and photorespiration. From an energetics point of view, there is a balance between the energetic investment in the CO₂-concentrating mechanism in the form CAM cycle and the energetics benefit from suppressing the energy-consuming process of photorespiration. The model predictions of photon use of C₃ and CAM leaves were found to be similar (Figure 5.7), suggesting that there is little to be gained in terms of energetics in running the CAM cycle. A knock-on consequence of the CO₂-concentrating mechanism and the suppression of rubisco oxygenase flux is that less Rubisco is required to fix a certain amount of carbon. Given that Rubisco contributes up to 50% of the soluble proteins and 20-30% of the total nitrogen in a C₃ leaf (Feller et al., 2008), a reduction of 21-32% of total Rubisco flux in CAM compared to C₃ predicted by the model (Figure 5.7) could be a significant benefit in terms of nitrogen use efficiency. However, similar to the energetics argument, there is an additional enzyme machinery costs for operating the CAM cycle. There are contradicting experimental observations relating to nitrogen efficiency in CAM plants (Lüttge, 2004). From the model predictions, CAM requires 32-64% more total flux through the metabolic network than C₃ (Figure 5.7). Although enzyme machinery cost per metabolic flux for each reaction may vary depending on factors such as the size and the turnover rate of the enzyme, the total network flux can be used as a proxy for the overall enzyme machinery cost by averaging out the variation between reactions. By balancing the benefit from reduced Rubisco

requirement with the increase in enzyme machinery costs in the rest of the metabolic network, the total nitrogen invested in enzyme machinery is likely to be similar in C_3 and in CAM. The modelling results suggest that energetics and nitrogen use efficient are unlikely to drive the evolution of CAM plants under current atmospheric CO_2 concentration.

5.3.5 Summary

The aim of this chapter was to predict metabolic fluxes in photosynthetic leaves in C_3 and CAM plants over a day-night cycle. To explore the interactions between light and dark metabolism, a novel diel modelling framework was developed in which the metabolic fluxes in the light phase and the dark phase were simulated simultaneously in a single optimisation problem. By comparing with flux prediction from single steady state approach, the diel modelling framework was shown to be important in modelling the light and dark metabolism in flux balance models of photosynthetic organisms. Known features of metabolism in C_3 and various subtypes of CAM leaves were accurately predicted from the Arabidopsis genome-scale metabolic model using the diel modelling framework. The model gave novel predictions of the metabolic mode for the catabolism of citrate in the light in CAM plants. The comparison between model predictions of C_3 and CAM leaves suggests that energetics and nitrogen use efficient are unlikely to be the evolutionary driving force for CAM plants under current atmospheric CO_2 concentration.

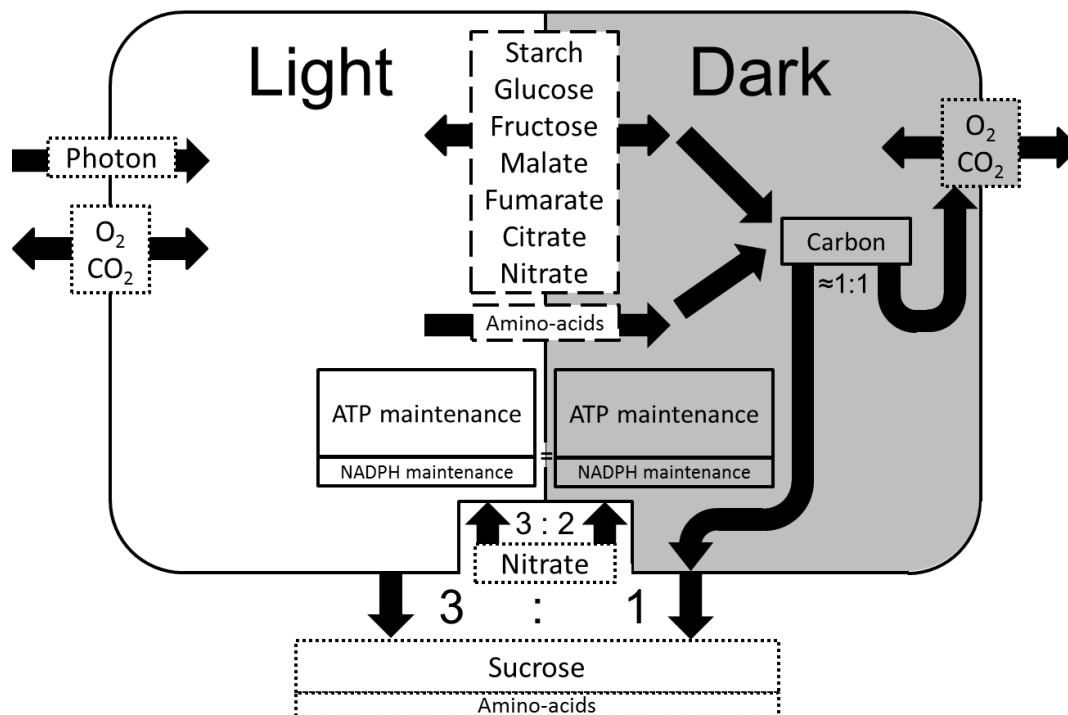


Figure 5.1 Diel metabolic modelling framework illustrating the main constraints for modelling metabolism of a mature leaf. The light phase and the dark phase are represented by the left side (white background) and right side (grey background) of the diagram respectively. Metabolites in between the light phase and the dark phase represent storage compounds. Starch, glucose, fructose, malate, fumarate, citrate and nitrate were allowed to accumulate in the light and in the dark as denoted by the arrow pointing at both the light and dark phases. Twenty common amino-acids were allowed to accumulate in the light but not in the dark as denoted by the arrow pointing from the light phase to the dark phase. Export to the phloem was set to 2.2 sucrose to 1 amino-acid, with 18 amino-acids in fixed proportion as stated in Table 5.1 based on measurements of Arabidopsis phloem exudate composition (Wilkinson and Douglas, 2003). The export rate was set to be three times greater in the light than in the dark (Gibon et al., 2004). Nitrate was set as the sole nitrogen source with the ratio of nitrate uptake from the phloem in the light to that in the dark set to 3 to 2 (Delhon et al., 1995; Macduff and Bakken, 2003; Siebrecht et al., 2003). Cellular maintenance costs in the dark were set to a fixed value where the carbon exported during the night roughly equals the carbon released as CO₂ in the dark phase. Maintenance costs were assumed to be the same in the light as in the dark. The ratio of ATP maintenance cost to NADPH maintenance cost was set to 3 to 1 based on estimates from Arabidopsis heterotrophic cell culture in control conditions (chapter 4). ATP maintenance was assumed to be in the cytosol only and NADPH maintenance was assumed to be equally distributed in the plastid, cytosol and mitochondria.

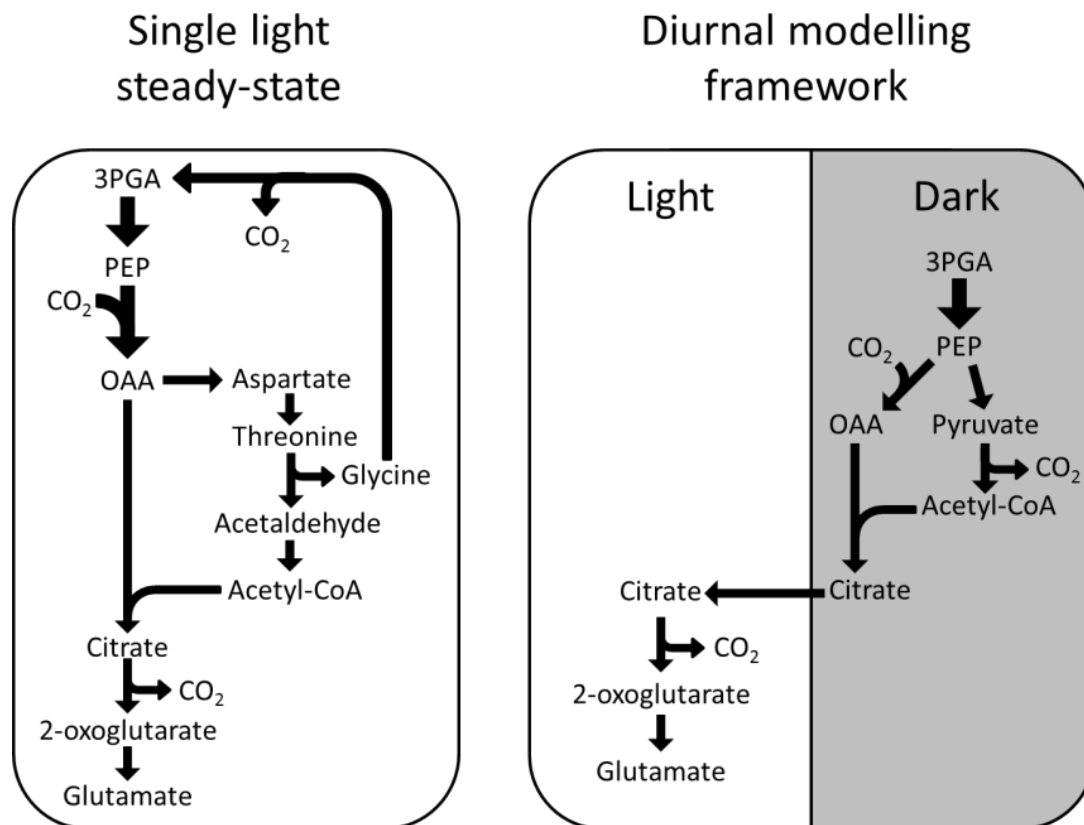


Figure 5.2 Metabolic routes for glutamate biosynthesis from two modelling approaches, single steady-state in the light (left) and the diel modelling framework (right). Threonine aldolase was predicted to be involved in glutamate synthesis in the single steady-state model. Using the diel modelling framework, the model predicted the use of citrate stored in the dark to provide carbon skeleton for glutamate synthesis in the light.

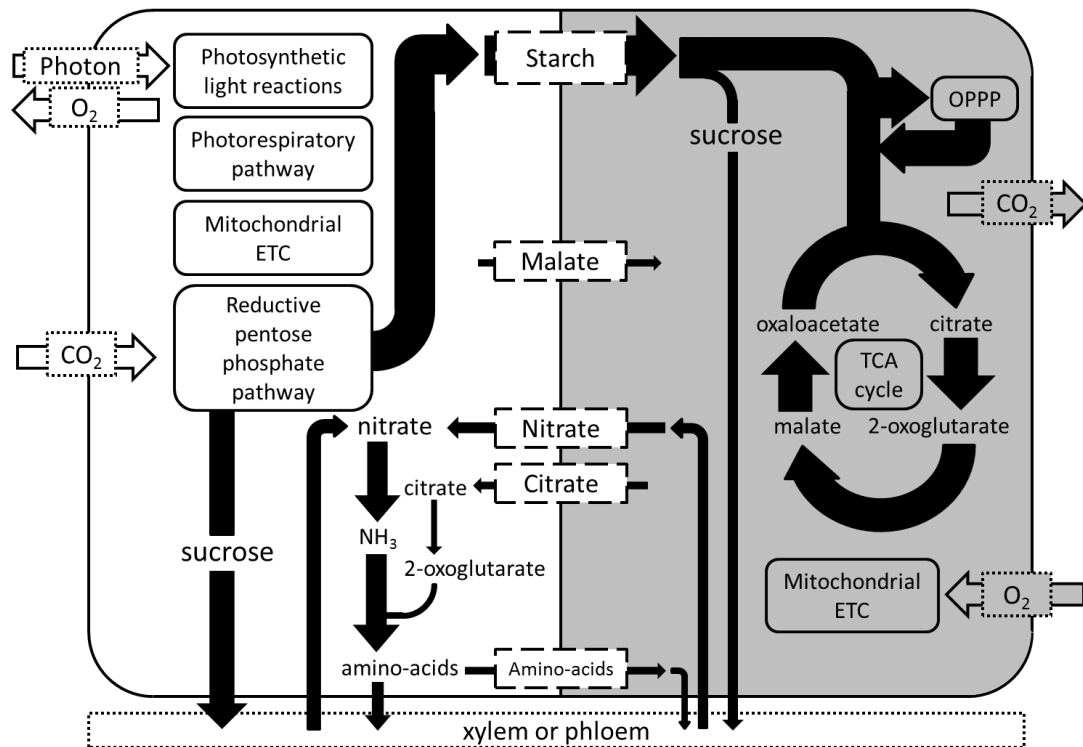


Figure 5.3 Predicted flux map of metabolism of mature C_3 leaf over a day-night cycle. The light phase and the dark phase are represented by the left side (white background) and right side (grey background) of the diagram respectively. Metabolites between the two phases represent storage compounds as described in Figure 5.1. The thickness of the arrow is proportional to the metabolic fluxes through the reactions. Metabolic processes in boxes carry fluxes too large to be represented on the flux map. Abbreviations: ETC, electron transport chain; OPPP, oxidative pentose phosphate pathway; TCA cycle, tricarboxylic acid cycle.

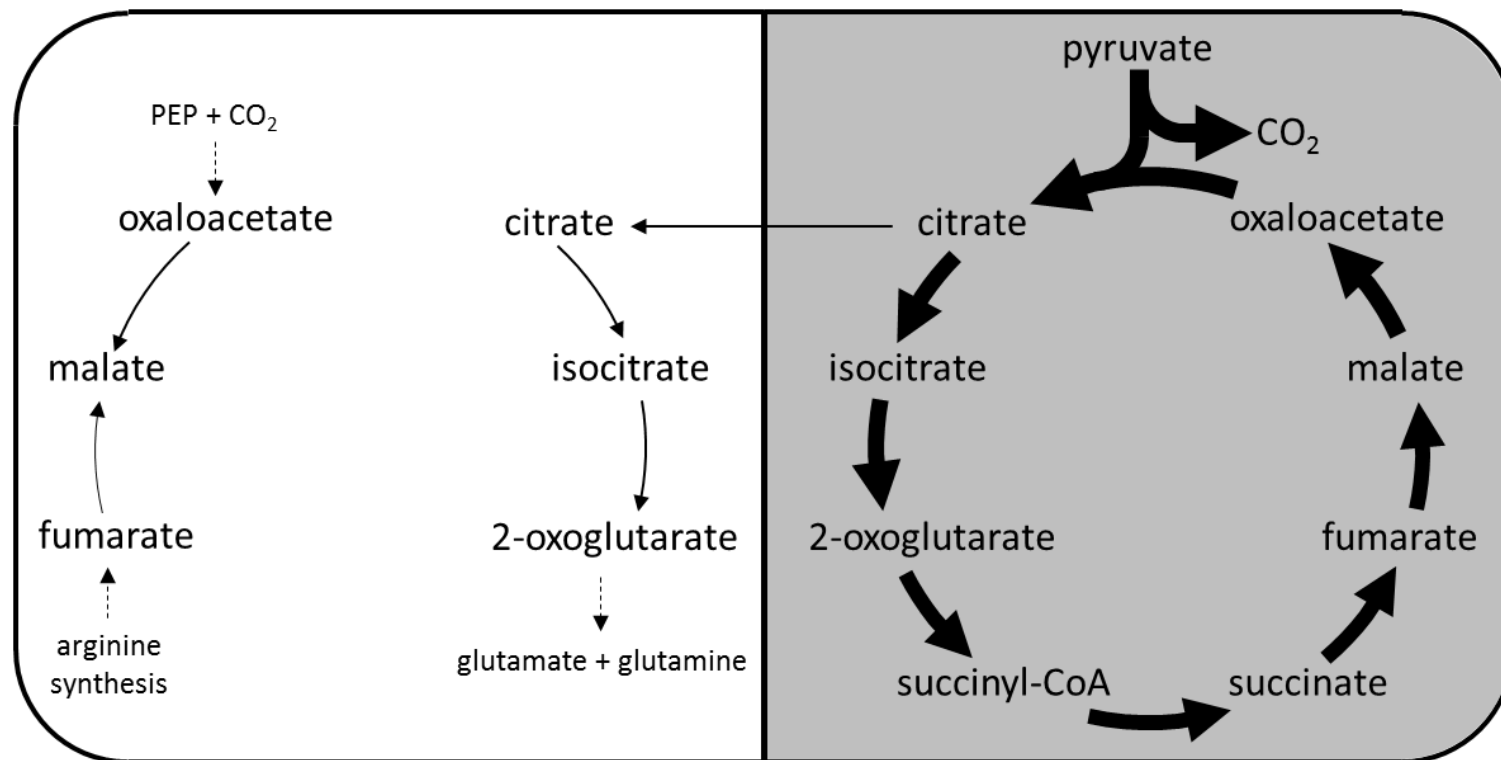


Figure 5.4 Flux predictions through the TCA cycle and pyruvate dehydrogenase in a C_3 leaf in the light phase (left of diagram; white background) and in the dark phase (right of diagram; grey background). A non-cyclic mode with two separate branches, citrate to 2-oxoglutarate and oxaloacetate and fumarate to malate, was predicted to operate in the light. A cyclic mode of the TCA cycle was predicted to operate in the dark, mainly to produce ATP via oxidative phosphorylation. Fluxes illustrated are net conversion between metabolites over all subcellular compartments. The thickness of the arrow is proportional to the metabolic fluxes through the reactions. Citrate stored in the dark is represented by an arrow pointing from the dark phase to the light phase. Malate storage is not illustrated in this diagram.

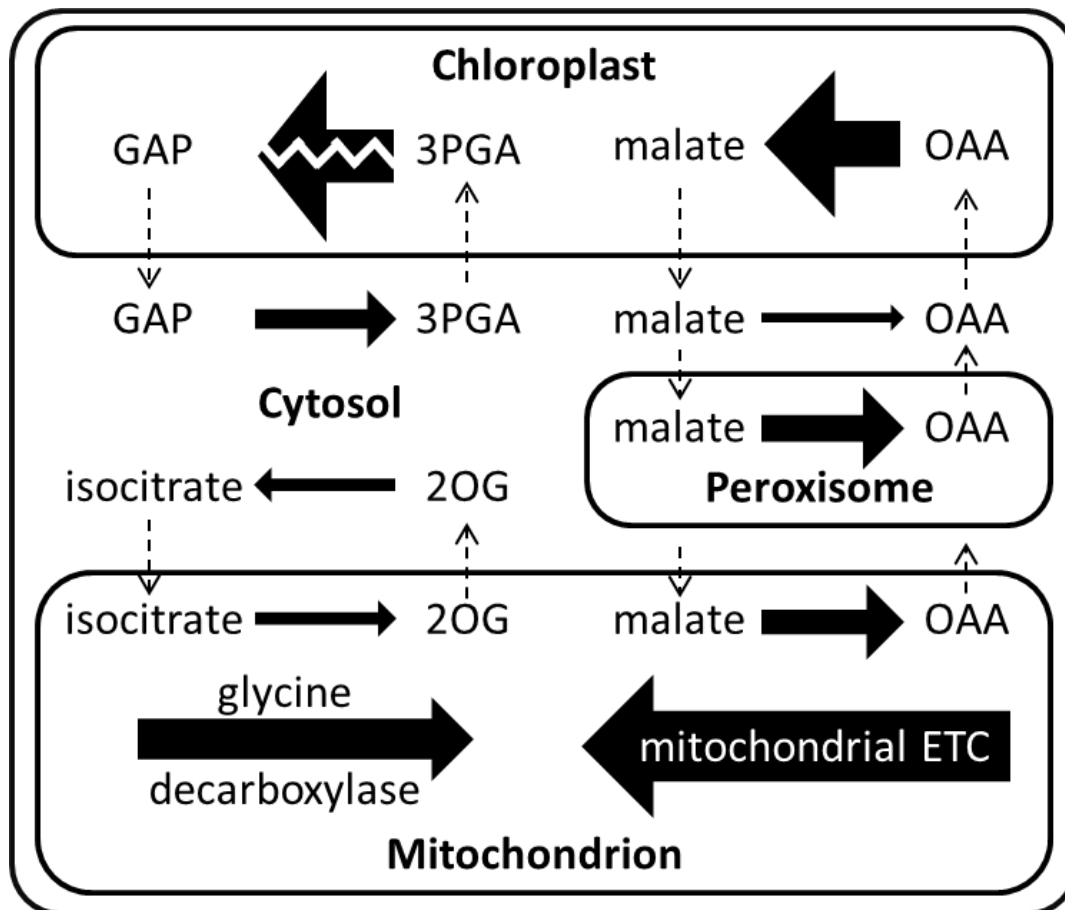


Figure 5.5 Fluxes through reactions related to reductant shuttling between subcellular compartments and the production and consumption of NADH in mitochondria. Left-pointing arrows represent reductant-consuming reactions; right-pointing arrows represent reductant-producing reactions. The thickness of the solid arrow is proportional to the metabolic flux through the reaction, except for the conversion between 3PGA and GAP in the chloroplast which is illustrated by a zig-zag line across the arrow as the flux is too large to be illustrated on the diagram in scale. Reductant shuttles between four subcellular compartments, cytosol, chloroplast, mitochondria and peroxisome, were shown with dashed arrows representing transfer of metabolites between compartments. Abbreviations: GAP, glyceraldehyde 3-phosphate; 3PGA, 3-phosphoglycerate; OAA, oxaloacetate; 2OG, 2-oxoglutarate; ETC, electron transport chain.

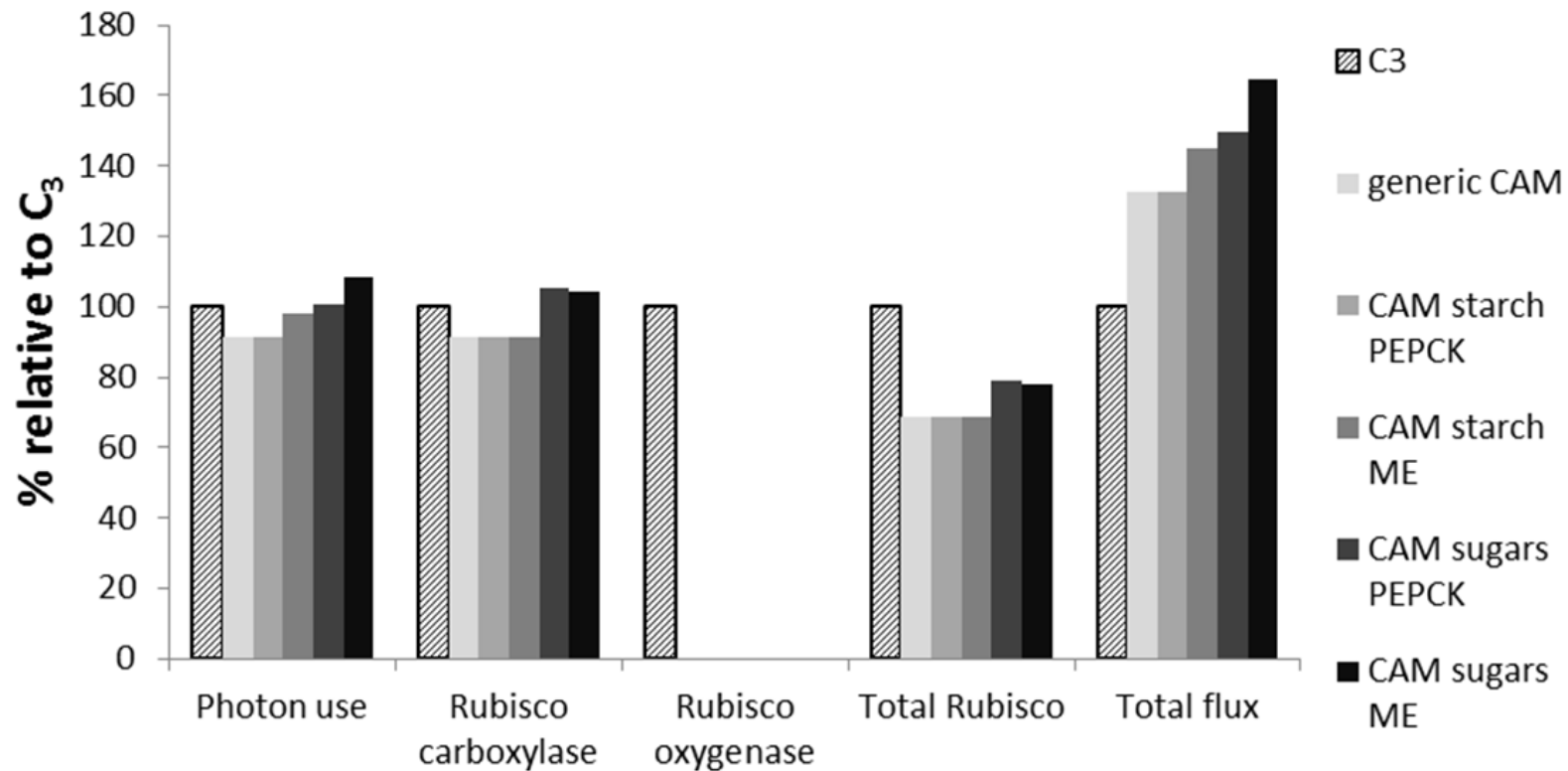


Figure 5.7 Comparison of model predictions between C₃ and various subtypes of CAM. The model predictions of photon use, Rubisco carboxylase flux, Rubisco oxygenase flux, total flux through Rubisco and the total flux in the metabolic model in C₃ and various subtypes of CAM leaves were shown with values scaled as a percentage of prediction of C₃ leaf. Photon use and Rubisco carboxylase flux are similar in C₃ and in various subtypes of CAM. Rubisco oxygenase flux is absent in CAM which results in a lower total Rubisco flux in CAM than in C₃. The total flux through the metabolic network is 32-64% higher in CAM than in C₃.

Table 5.1 Amino-acids composition in the phloem in Arabidopsis.

Values taken from Wilkinson and Douglas (2003) are shown as % of the total amino-acids in the phloem, sorted from the most to the less abundant.

Amino-acid	% of amino-acid in phloem
Glutamine	33.2
Glutamate	11.4
Asparagine	10.1
Alanine	9.7
Aspartate	9.5
Serine	7.7
Threonine	5.6
Valine	2.4
Lysine	1.8
Arginine	1.6
Isoleucine	1.2
Leucine	1.2
Phenylalanine	1.1
Tryptophan	0.9
Glycine	0.7
Methionine	0.7
Tyrosine	0.6
Histidine	0.4

Table 5.2 List of reactions in the light that have dissimilar fluxes in the diel model and the single steady-state model. (A) Reactions with high flux in the diel model and low flux in the single steady-state model. (B) Reactions with high flux in the single steady-state model and low flux in the diel model. The reaction names correspond to those in the diel model (Appendix S3). Reactions with similarity measures less than 0.25 are listed, where a small value represents a large difference in fluxes from the two modelling approaches (see Appendix S9 for a complete list of reactions with non-overlapping flux ranges). Details on the calculations of the similarity measure are given in chapter 2. The metabolic functions of the reactions are listed in the column on the right.

Table 5.2A

Reactions with high flux in the diel model	Similarity measure	Function
GLUC1PADENYLTRANS-RXN_Light_p	0	Starch synthesis
GLYCOGENSYN-RXN_Light_p	0	Starch synthesis
PGLUCISOM-RXN_Light_p	0	Starch synthesis
PHOSPHOGLUCMUT-RXN_Light_p	0	Starch synthesis
CIT_H+_Light_vc	0	Storage in the vacuole
MAL_H+_rev_Light_vc	0	Storage in the vacuole
FUM_H+_rev_Light_vc	0	Storage in the vacuole
NITRATE_Light_vc	0	Storage in the vacuole
H+(PPi)_rev_Light_vc	0	Storage in the vacuole
INORGPYROPHOSPHAT-RXN_Light_p	0.015	Related to starch synthesis

Table 5.2B

Reactions with high flux in the single steady-state model	Similarity measure	Function
INORGPYROPHOSPHAT-RXN_Light_x	0	Related to Glu and Gln synthesis
Pi_Light_xc	0	Related to Glu and Gln synthesis
CITSYN-RXN_Light_x	0	Related to Glu and Gln synthesis
ACETATE--COA-LIGASE-RXN_Light_x	0	Related to Glu and Gln synthesis
AMP_ATP_Light_xc	0	Related to Glu and Gln synthesis
ACET_Light_xc	0	Related to Glu and Gln synthesis
CIT_Light_xc	0	Related to Glu and Gln synthesis
THREONINE-ALDOLASE-RXN_Light_c	0.019	Related to Glu and Gln synthesis
RXN66-3_Light_c	0.019	Related to Glu and Gln synthesis
ATP_ADG_Pi_Light_pc	0.093	Export of ATP from the chloroplast
THR_Light_pc	0.123	Related to Glu and Gln synthesis
THRESYN-RXN_Light_p	0.145	Related to Glu and Gln synthesis
HOMOSERKIN-RXN_Light_p	0.153	Related to Glu and Gln synthesis
HOMOSERDEHYDROG-RXN-(NADP)_Light_p	0.161	Related to Glu and Gln synthesis
ASPARTATEKIN-RXN_Light_p	0.180	Related to Glu and Gln synthesis
ASPARTATE-SEMIALDEHYDE-DEHYDROGENASE-RXN_Light_p	0.180	Related to Glu and Gln synthesis
ATP_AMP_Light_mc	0.202	Related to Glu and Gln synthesis
ADENYL-KIN-RXN_Light_m	0.202	Related to Glu and Gln synthesis
L-ASPARTATE_Light_pc	0.208	Related to Glu and Gln synthesis

Table 5.3 Constraints applied for modelling C₃ and various subtypes of CAM leaves. CO₂ exchange in the light was constrained to zero flux in CAM. Rubisco carboxylase to oxygenase ratio was set to 3:1 in C₃ in both the light and the dark and in CAM in the dark, whereas the ratio was unconstrained in CAM in the light. The chloroplastic glucose 6-phosphate-Pi exchanger was set to zero flux in C₃ and was unconstrained in CAM. The various subtypes of CAM were simulated by setting the reactions specific to other subtypes to carry zero flux.

Constraints for different modes of photosynthesis		C ₃	CAM				
			generic	starch-PEPCK	starch-ME	sugars-PEPCK	sugars-ME
CO ₂ exchange	CO ₂ exchange in light	Free	0				
Photorespiration	Rubisco CO ₂ :O ₂	3:1	Unconstrained in light, 3:1 in dark				
Transporter	chloroplast G6P-Pi	0	Free				
Metabolic reactions	PEPCK	Free	Free	Free	0	Free	0
	ME	Free	Free	0	Free	0	Free
	pyruvate, P _i dikinase	Free	Free	0	Free	0	Free
Storage compounds	Starch	Free	Free	Free	Free	0	0
	Soluble sugars	Free	Free	0	0	Free	Free

6 Modelling photosynthetic metabolism in response to varying environmental and physiological conditions

6.1 Introduction

6.1.1 Importance of alternative electron flow in photosynthetic metabolism

Carbon fixation and other metabolic processes require energy and reducing power in the form of ATP and NAD(P)H. In a photosynthetic leaf, photosynthetic linear electron transport is the primary source of ATP and NADPH, generating ATP and NADPH at a fixed ratio of 1.29 ATP per NADPH (Allen, 2003). However, the fixation of CO₂ by the reductive pentose phosphate pathway requires 1.5 ATP per NADPH (Noctor and Foyer, 1998). In addition to CO₂ fixation, various metabolic processes such as nitrogen assimilation, photorespiration and the synthesis of different biomass components require different ATP:NADPH ratios. The fluxes through these processes vary depending on environmental and physiological conditions. As a result, there is a mismatch between the supply and demand of ATP and NADPH. To balance the supply and demand of ATP and NADPH, plants exploit multiple routes of alternative electron flow (AEF) (Kramer and Evans, 2011). AEFs produce ATP at the expense of reducing power, which has a net effect of increasing the supply of ATP per NADPH. This chapter focused on three major routes of AEF, the mitochondrial electron transport chain (ETC), cyclic electron transport around photosystem I and the water-water cycle. The mitochondrial ETC oxidises NADH to generate a proton gradient across the inner mitochondrial membrane, which can be used to drive the synthesis

of ATP by the mitochondrial ATP synthase. For the mitochondrial ETC to act as AEF, NADPH in the chloroplast is shuttled to the mitochondria by the malate-oxaloacetate (OAA) shuttle. This involves the chloroplastic malate dehydrogenase, which is under redox-control to be active only in the light, and is commonly known as the malate valve (Scheibe, 2004). Within the chloroplast, the two main AEFs are cyclic electron transport and the water-water cycle (Miyake, 2010). In cyclic electron transport, electrons from photosystem I are transferred back to an upstream electron carrier in linear electron transport, namely plastoquinone, rather than to NADPH in linear electron transport. The net effect of cyclic electron transport is the generation of a proton gradient across the thylakoid membrane to drive ATP synthesis. In the water-water cycle, also known as the Mehler-ascorbate peroxidase cycle or pseudocyclic electron transport, electrons from photosystem I are transferred to oxygen to form superoxide in the Mehler reaction (Heber, 2002). The superoxide is subsequently metabolised by superoxide dismutase to form hydrogen peroxide, which is reduced to water by ascorbate peroxidase (Miyake, 2010). Similar to cyclic electron transport, the water-water cycle competes for electrons from photosystem I and produces ATP by generating a proton gradient across the thylakoid membrane.

While these AEFs are well established in plants, quantifying their relative importance in a photosynthetic leaf is experimentally difficult. In addition, changes in environmental and/or physiological conditions alter the energy supply and/or metabolic demand of the system, which may lead to changes in the relative importance of various routes of alternative electron flow. For example, cyclic electron transport seems to be a phenomenon of stress

conditions (Johnson, 2005). Multiple AEFs in *Synechocystis* were characterised using a *Synechocystis* genome-scale metabolic model, and the cooperation between a pair of AEF was found to be important for optimal photosynthetic metabolism (Nogales et al., 2012). While the AEFs were characterised by simulating each AEF one at a time, with constraints on the fluxes of other AEFs, the relative importance of the multiple AEFs was not investigated. One possible reason could be that if the model had the freedom to choose between different AEFs, the standard FBA approach would generate a single flux solution containing only the most efficient AEF in terms of maximising growth with no flux through the other AEFs. In this chapter, a novel modelling approach generating multiple flux solutions based on flux minimisation with random weighting factors was developed and applied to investigate the relative importance of various routes of AEF under different environmental or physiological conditions using the Arabidopsis genome-scale metabolic model (chapter 3) and the diel modelling framework (chapter 5).

6.1.2 Modelling photosynthetic metabolism under different conditions

Plants, being sessile, must regulate their metabolism to adapt to changes in environmental conditions. One of the most variable environmental factors for plants is the availability of light. In plants, a genome-scale metabolic model of rice was used to investigate the metabolic response to light intensity and the metabolic interactions between organelles in a developing leaf cell of rice (Poolman et al., 2013). Besides the variation in quantity of light, metabolism of *Chlamydomonas* growing under various light sources with different light

spectral quality have also been modelled using a genome-scale metabolic model of *Chlamydomonas*, which offers potential routes for genetic engineering and light source design (Chang et al., 2011). Another important environmental factor for plants is water availability. A metabolic/regulatory network model of rice metabolism was developed and utilised to investigate the metabolic changes of seed-derived rice cells to flooded conditions and photorespiring leaf cells to drought conditions (Lakshmanan et al., 2013). With respect to nitrogen, the influence of nitrogen assimilation on the ATP to reductant balance and O_2 and CO_2 exchange during C_3 photosynthesis were investigated based on calculations from a hand-built model (Noctor and Foyer, 1998). Using a flux balance model of developing oilseed rape embryos, the variability of fluxes were investigated with two different nitrogen sources, inorganic or organic, as well as the presence and absence of light (Hay and Schwender, 2011b). Thus far, models were only used to investigate the changes in photosynthetic metabolism with either a single environmental variable or at distinct states of two variables. In this chapter, the continuous metabolic changes in a C_3 leaf with varying light intensity and nitrogen availability were simulated, which allowed the differential responses of various routes of alternative electron flow to the two environmental variables to be investigated.

6.1.3 Physiological constraints in FBA studies of photosynthetic metabolism

Most FBA studies on photosynthetic organisms were modelled as growing cells by including the synthesis of biomass components either as constraints or as an objective function. This is justifiable for modelling microorganisms,

plant cells grown in culture and possibly plant seeds and embryos as their main physiological function is the production of biomass for growth. While most studies assumed a constant biomass composition, the trade-off between oil and protein in the biomass components in developing oilseed rape embryos was investigated and the associated metabolic changes were identified (Schwender and Hay, 2012). However, many cells and tissues in plants, such as leaves and roots, carry out metabolic functions other than growth. To model the metabolism of these cells and tissues, constraints associated with their metabolic functions should be applied. To date, photosynthetic metabolism of plants in published FBA studies has been modelled with the synthesis of biomass components either as a constraint or as an objective function (de Oliveira Dal'Molin et al., 2010a; de Oliveira Dal'Molin et al., 2010b; Saha et al., 2011; Poolman et al., 2013; Lakshmanan et al., 2013). This corresponds to a developing leaf at the stage when the amount of carbon goes into biomass equals the net carbon assimilation with no exchange of carbon with the phloem. In chapter 5, the metabolism of mature C_3 and CAM leaves were modelled by constraining the model with sucrose and amino-acids export to the phloem and cellular maintenance costs. In this chapter, the Arabidopsis genome-scale metabolic model was utilised to model a C_3 leaf at various developmental stages from developing leaf to mature leaf.

6.2 Results

6.2.1 The use of random weighting factors in minimisation of total flux as an objective function

The use of flux minimisation as an objective function in FBA was proposed based on the principle that cells are evolved to minimise the costs for the synthesis of enzymes and membrane transporters to carry out vital functions (e.g. growth and cellular maintenance processes) (Holzhütter, 2004). In previous FBA studies on the Arabidopsis genome-scale metabolic model, fluxes were predicted using flux minimisation assuming equal weighting for each reaction (Poolman et al., 2009; Williams et al., 2010). This assumes the enzyme machinery cost for each reaction is equal per unit flux. However, this assumption is unlikely to be true *in vivo* as enzymes vary largely in terms of their size, number of subunits and catalytic activity. Ideally, each reaction should be weighted by its enzyme machinery costs per unit flux, but such information is not available for the majority of the reactions in a genome-scale metabolic model. To explore the effect of variability in the enzyme machinery costs on the resulting flux solution, a method based on random weighting factors was developed which assigns a uniformly distributed random number between 0 and 1 as the weighting factor for each reaction (see chapter 2 for detailed methods). Multiple flux solutions can be computed by applying flux minimisation multiple times, each with a different set of random weighting factors.

The results from the random weighting factor method can be illustrated by the use of toy models. Toy model 1 has two alternative routes with the same number of reactions (Figure 6.1A). By setting 'W_{in}' with 1 unit of flux,

standard FBA using flux minimisation with uniform weighting factor would give flux predictions which uses either 'R1' or 'R2' depending on the algorithm used to solve the linear programming problem ('R1' in the case of ScrumPy). A flux variability analysis (FVA) gave a flux range of (0,1) for both 'R1' and 'R2' (Table 6.1). With 1000 different sets of random weighting factors, the two routes were predicted with roughly equal probability (496 and 504 times for routes involving 'R1' and 'R2' respectively). By averaging the flux solutions from the multiple runs of FBA with different sets of weighting factors gave an averaged flux solution with 0.496 and 0.504 unit of flux through 'R1' and 'R2' respectively (Table 6.1). A more interesting toy model was investigated with two alternative routes with route 2 involving one reaction more than route 1 (Figure 6.1B). In this case, FBA using flux minimisation with uniform weighting factors always chooses 'R1'. FVA also showed that 'R1' is preferred over 'R2a' and 'R2b' with flux ranges of (1,1) for 'R1' and (0,0) for 'R2a' and 'R2b' (Table 6.1). Using the random weighting approach, 'R2a' and 'R2b' can be chosen when the sum of the weighting factors of 'R2a' and 'R2b' is less than the weighting factor of 'R1'. With 1000 sets of random weighting factors, 'R2a' and 'R2b' were chosen 161 times which means 'R2a' and 'R2b' were predicted to carry 0.161 unit of flux in the averaged flux solution (Table 6.1). While metabolic routes with higher sum of fluxes are always neglected in flux solution generated from the uniform weighting factor approach, the results from the toy models showed that alternative routes with similar sum of fluxes can be captured using the random weighting approach. In addition, the relative fluxes through the alternative metabolic routes can be predicted based on the overall sum of

fluxes. This characteristic of the random weighting approach was exploited to predict the relative contributions of different routes of alternative electron flow (AEF) in photosynthetic metabolism of leaves.

6.2.2 Changes in leaf metabolism under varying light intensity

Plants often experience highly variable light intensities over the course of the day. This leads to a potential mismatch between photon influx and CO₂ uptake which compromises the rate of photosynthesis. The changes in leaf metabolism with varying light intensities were modelled using the Arabidopsis genome-scale metabolic model and the diel modelling framework. Photon influx was varied while applying constraints on sucrose and amino-acids export, cellular maintenance costs and other C₃-specific constraints described in chapter 5. In contrast to a previous FBA study on *Synechocystis* which investigated the response of individual routes of AEF (and the interaction of one particular pair of AEF) to photon influx (Nogales et al., 2012), here all routes of AEF were modelled at the same time using the random weighting factor approach so that the changes in their relative importance, based on model stoichiometry, can be investigated. For each photon influx, an averaged flux solution was calculated from flux minimisation with 1000 different sets of random weighting factors. Reactions representing CO₂ and O₂ diffusion across membrane, accumulation of storage compounds, biomass synthesis, generic ATPases and NADPH oxidases for maintenance costs were set to have a weighting factor of zero in all cases as these reactions were included for modelling purposes and have no associated enzyme machinery costs.

With varying light intensity, the net CO₂ assimilation rate in the light remained more or less constant, with less than 5% deviation from the average value over the range of light intensities investigated (Figure 6.2A; see Appendix S5 for flux solutions). To relate the model predictions to physiological values, the average value of net CO₂ assimilation rate in the light was scaled to an experimentally-determined net CO₂ assimilation rate (11.8 μmol m⁻² s⁻¹) for *Arabidopsis* grown under 200 μmol m⁻² s⁻¹ (Eckardt et al., 1997). The quantum yield calculated from the experimental measurements gave a value of 0.059 CO₂/photon. By varying the light intensity, the model gave a theoretical minimum light intensity of 177.14 μmol m⁻² s⁻¹, which corresponds to a quantum yield of 0.067 CO₂/photon. However, caution must be taken in relating the theoretical photon requirement with experimental measurements as the model assumes 100% of photons are absorbed and converted into chemical energy, which is not true *in planta*. At the minimum photon flux required to support carbon fixation, nitrogen assimilation and cellular maintenance, photosynthetic linear electron transport (Figure 6.2B) and mitochondrial electron transport chain were operating with no other AEF (Figure 6.2C). This is consistent with the flux prediction from chapter 5 where the photon influx was unconstrained and flux minimisation with uniform weighting factor was applied.

Flux through the photosynthetic linear electron transport increased linearly with photon influx (Figure 6.2B). Similar behaviour was predicted for various reductant-dissipating reactions including the chlorophyll cycle, glutathione-ascorbate cycle and xanthophyll cycle (Figure 6.2D) to dissipate the excess reductant generated from the photosynthetic linear electron transport. With

respect to AEF, the model predicted that the relative importance of various routes of AEF changes with varying photon influx. These metabolic changes can be broadly separated into three phases. From a photon influx of around 177 to 190 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (phase 1; not to be confused with phase I in CAM photosynthesis), the model predicted a decrease in flux through the mitochondrial ETC (Figure 6.2C) which leads to a decrease in ATP synthesis by the mitochondrial ATP synthase. This was partly compensated by the increase in the photosynthetic linear electron transport and partly by the increase in cyclic electron transport and the water-water cycle (Figure 6.2C). As photon influx increased from around 190 to 215 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (phase 2), the mitochondrial ETC flux continues to decrease as more ATP is produced by the increasing photosynthetic linear electron transport. Given a relatively constant ATP demand for cellular maintenance and synthesis of sucrose, amino-acids and storage compounds, the continuous increase in ATP synthesis from the photosynthetic linear electron transport with increasing photon influx resulted in a decrease in other ATP generating processes including cyclic electron transport and the water-water cycle (Figure 6.2C). In this phase, the model also predicted an increase in flux through the mitochondrial alternative respiratory pathway which includes the alternative NADH dehydrogenase and the alternative oxidase (Figure 6.2E). These reactions dissipated the reductant generated in the mitochondria by glycine decarboxylase in the photorespiratory pathway. The mitochondrial uncoupling protein was also predicted to have a small increase in flux during phase 2 (Figure 6.2E). Further increase in photon influx from 215 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (phase 3) resulted in little change in the fluxes through the mitochondrial ETC

and the alternative respiratory pathway. As the photosynthetic electron transport continues to increase in flux, ATP is produced in excess and several reactions involved in ATP dissipation start to carry flux and increase in flux with photon influx in phase 3 (Figure 6.2F). These reactions couple with other reactions in the metabolic network to form ATP-dissipating futile cycles. ATP-dissipating reactions and cycles predicted to carry the highest fluxes were identified from the flux solutions (Appendix S5), which include carbamoyl phosphate cycling, formate-tetrahydrofolate (THF) cycling and phosphofructokinase, i.e. cycling of fructose 1,6-bisphosphate. Although the identified futile cycles may not all be in operation *in vivo*, the random weighting factors approach allowed one to explore the possibilities and capabilities of the metabolic network and identify the likely metabolic routes based on flux minimisation.

6.2.3 Changes in leaf metabolism under varying nitrogen availability

Nitrogen is one of the major limiting nutrients for plant growth. The availability of nitrogen to a mature leaf influences the rate of amino-acid export into the phloem as well as the reaction fluxes involved in nitrogen assimilation and related carbon metabolism. Changes in leaf metabolism in response to nitrogen availability were modelled by varying the influx of nitrate, the sole nitrogen source in the leaf model. Net carbon export was constrained to be constant but the ratio of sucrose to amino-acid export was allowed to vary depending on the nitrate influx. At the default sucrose to amino-acid export ratio of 2.2 : 1, the net carbon assimilation rate in the light was scaled to the experimental value of $11.8 \mu\text{mol m}^{-2} \text{s}^{-1}$ (Eckardt et al., 1997) as described in

section 6.2.2. This corresponds to a nitrate influx over a day-night cycle of 20.48 mmol m⁻² day⁻¹ at the default sucrose to amino-acids export ratio. As nitrate influx increases, there was a corresponding increase in photon requirement as nitrate assimilation requires energy and reducing power (Figure 6.3A). One major impact of altering nitrogen assimilation on photosynthetic metabolism is the change in demands of ATP and reducing power, as nitrate assimilation requires more reducing power per ATP than carbon assimilation. This was investigated by calculating the ATP and NADPH demands from flux solutions (see chapter 2 for detailed methods) with varying nitrate influx. The ratio of ATP demand to NADPH demand decreased as nitrate influx increased (Figure 6.3B). In response to the decrease in ATP requirement relative to NADPH, the model predicted a decrease in fluxes through various routes of AEF, which oxidise reductant to generate ATP, as nitrate influx increases (Figure 6.3C).

In a C₃ leaf, the incorporation of ammonia into amino-acids by the glutamine synthetase/glutamate synthase (GS/GOGAT) cycle serves two main functions, *de novo* amino-acids synthesis and the recycling of ammonia released from the photorespiratory pathway. One isoform of GS was recently discovered to be dual targeted to the chloroplasts and mitochondria (Taira et al., 2004). Linka and Weber, (2005) suggested that the mitochondrial GS might be involved in the assimilation of ammonia release by glycine decarboxylase in the mitochondrial matrix and hypothesised a glutamate-glutamine shuttle which transfer glutamate and glutamine between the mitochondria and the chloroplast. These latest findings and hypotheses were incorporated into the Arabidopsis genome-scale metabolic model. The model

flux solutions showed differential responses between the mitochondrial and the chloroplastic GS with varying nitrate influx. While flux through the mitochondrial GS remains relatively constant (with a slight decrease) with nitrate influx, there is a close to linear increase in flux through the chloroplastic GS (Figure 6.3D). The cytosolic isoform of GS was predicted to carry little flux in the leaf. The model predictions suggest that ammonia assimilation for *de novo* amino-acid synthesis is primarily carried out by the chloroplastic GS whereas the mitochondrial GS is primarily involved in reassimilating the ammonia released by glycine decarboxylase in the photorespiratory pathway.

6.2.4 The effect of growth on photosynthetic metabolism

Thus far, only the metabolism of a mature leaf was modelled in this study. To extend the analysis to different stages of leaf development, leaf growth was modelled by including the synthesis of biomass components of an *Arabidopsis* leaf based on the biomass composition used in a published *Arabidopsis* genome-scale metabolic model, AraGEM (de Oliveira Dal'Molin et al., 2010a). The following three growth stages were modelled by setting the corresponding constraints on biomass synthesis and metabolite exchange with the phloem.

1. Developing leaf – biomass synthesis with no carbon import from or export into phloem
2. Maturing leaf – carbon incorporated into biomass equals carbon exported into phloem

3. Mature leaf – all net fixed carbon exported into phloem with no biomass synthesis

Attempts in modelling the metabolism of a sink leaf were unsuccessful due to a mismatch between the defined compositions of phloem and biomass and the lack of model curation on the relevant parts of the metabolic network. For example, histidine was one of the amino-acids defined in the phloem but was missing in the biomass composition, which means in modelling a sink leaf, histidine imported from the phloem must be degraded to provide nitrogen and carbon for the synthesis of other biomass components. Due to the lack of model curation in amino-acids degradation, the model was unable to catabolise histidine imported from the phloem. Given that most plant genome-scale metabolic models were only tested for the synthesis of biomass components, but not their degradation, this example highlighted the limitation in applying these models in modelling the catabolism of biomass components, which is especially relevant for modelling senescence-related processes.

The averaged flux solution from flux minimisation with 1000 different sets of random weighting factors was calculated for each growth stage. To compare between the various growth stages, the flux solutions were normalised to per net carbon fixed over a day-night cycle. The major flux modes in central carbon metabolism remained relatively constant from developing to mature leaf with a large flux through the photosynthetic linear electron transport chain, the reductive pentose phosphate cycle and the photorespiratory pathway in the light (see Appendix S5 for flux solutions). The TCA cycle ran in non-cyclic mode in the light and cyclic mode in the dark in all growth

stages simulated. The major metabolic difference between developing leaf and mature leaf, besides the obvious differences in reaction fluxes involved in biomass synthesis, was the amount of nitrogen assimilation (Figure 6.4A; Table 6.2). The net nitrogen assimilation per carbon fixed was higher in a developing leaf, presumably because a larger proportion of fixed carbon went into protein and amino-acids in a developing leaf compared to a mature leaf. This resulted in a lower ATP demand per NADPH required in a developing leaf compared to a mature leaf (Figure 6.4B; Table 6.2). This change in ratio of ATP demand to NADPH demand was met primarily by the mitochondrial ETC, where the flux ratio of mitochondrial ETC to the photosynthetic linear electron transport increased from developing leaf to mature leaf (Figure 6.4C; Table 6.2). In contrast, the ratio of cyclic electron transport and the water-water cycle to the photosynthetic linear electron transport remained relatively constant throughout leaf development (Figure 6.4C; Table 6.2). In terms of the overall energetics, since the synthesis of different biomass components (e.g. amino-acids, fatty acids etc.) is energetically more costly than sucrose synthesis, the photon requirement per net carbon fixed decreased over the course of leaf development as the proportion of fixed carbon incorporated into biomass components decreases (Figure 6.4D; Table 6.2). Relating back to nitrogen assimilation, the mitochondrial glutamine synthetase (GS) was predicted to carry similar flux over leaf development whereas flux through the chloroplastic GS decreased from developing leaf to mature leaf (Figure 6.4E; Table 6.2). This suggests that the dual targeted GS may serve distinct metabolic functions in the chloroplast and the mitochondria. In line with the results from varying nitrogen availability in section 6.2.3, the mitochondrial

GS flux was predicted to be independent of net nitrogen assimilation, suggesting that it could be responsible for the re-fixation of ammonia from photorespiration. On the other hand, the chloroplastic GS could be responsible for net nitrogen assimilation into amino-acids and other biomass components as its flux (per net carbon fixed) decreased with leaf development from developing leaf to mature leaf.

6.2.5 Exploring the flexibility of leaf metabolism

The analysis in the previous sections used the method of flux minimisation with random weighting factors to generate averaged flux solutions. While an averaged flux solution summarises the set of flux solutions generated from the multiple sets of random weighting factors, information on flux variability can also be extracted by looking at the variability of each reaction fluxes over the set of flux solutions (see chapter 2 for detailed method). This variability in fluxes represents the flexibility of the metabolic system while satisfying the evolutionary constraint of minimising enzyme machinery costs as well as the physiological constraints set by the input and output model constraints. The metabolic flexibility of a mature C₃ leaf was investigated by generating a set of flux solutions using flux minimisation as the objective function with 1000 different sets of random weighting factors (Appendix S11). The model was constrained using sucrose and amino-acid export values and cellular maintenance costs while leaving photon influx unconstrained as described in chapter 5. The net carbon assimilation rate in the light of the averaged flux solution was scaled to the experimental value of 11.8 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (Eckardt et al., 1997) and all flux solutions generated from different sets of random weighting factors were scaled with the same scaling factor.

Most reactions involved in photosynthesis, photorespiration, sucrose synthesis and nitrogen assimilation were found to be well-defined with little flux variability. Interestingly, there was high flux variability in several reactions in the reductive pentose phosphate cycle including the two phosphatases, sedoheptulose biphosphatase and fructose 1,6-bisphosphatase in the chloroplast. A scatter plot of the reaction fluxes of the two phosphatases over the set of 1000 flux solutions showed that the solutions lie in three distinct clusters (Figure 6.5). Other reactions correlated with the two phosphatases were identified by calculating the Pearson correlation coefficients between the fluxes of each reaction and the phosphatases (see chapter 2 for detailed method). Three reactions with absolute values of Pearson correlation coefficients of >0.999 with the two phosphatases were identified including sedoheptulose biphosphate aldolase, fructose 1,6-bisphosphate aldolase and transaldolase (Appendix S12). Inspection of the flux solutions (Appendix S11) showed they can be matched to three distinct flux modes (Figure 6.6). The middle flux mode, with equal flux through the two phosphatases, corresponds to the standard textbook reductive pentose phosphate pathway (Figure 6.6B). This analysis of flux variability showed that the metabolic fluxes of the reductive pentose phosphate pathway could vary between the two extremes, illustrated by the other two flux modes (Figure 6.6A and C), while satisfying the steady-state constraint.

The accumulation of storage carbon during the day was found to have high variability with a tight trade-off between hexose-based compounds (starch and soluble sugars) and four-carbon organic acids (malate and fumarate) with a Pearson correlation coefficient of -0.9992 (Figure 6.7). The high

variability between the storage compounds suggests that the four-carbon organic acids can partially complement the hexose-based compounds as storage during the day. Note that the accumulation of hexose-based compounds has a minimum of $0.417 \mu\text{mol m}^{-2} \text{s}^{-1}$, which presumably is required to support sucrose synthesis and to provide NADPH for maintenance in the chloroplast via the oxidative pentose phosphate pathway (OPPP) in the dark. The accumulation of four-carbon organic acids is highly positively correlated with malic enzyme, pyruvate dehydrogenase and several reactions in the TCA cycle in the dark, with Pearson correlation coefficients >0.95 , where these enzymes may form a probable route for malate and fumarate metabolism in the dark. Conversely, flux through the cytosolic OPPP in the dark is negatively correlated with four-carbon organic acids accumulation, with a Pearson correlation coefficient of -0.947 , suggesting that malic enzyme provided an alternative route for generating NADPH in the cytosol. For reactions in the light, the accumulation of four-carbon organic acids is positively correlated with net carbon dioxide uptake, Rubisco and phosphoenolpyruvate carboxylase, with Pearson correlation coefficients >0.93 , which means a higher carbon assimilation rate is required to accumulate four-carbon organic acids compared to hexose-based compounds. The Pearson correlation coefficients of all reactions with the accumulation of hexose-based compounds and four-carbon organic acids can be found in Appendix S12.

6.3 Discussion

6.3.1 The use of flux minimisation with random weighting factors in analysing FBA models

In this chapter, a novel modelling method of flux minimisation with random weighting factors was developed to investigate the relative importance of metabolic pathways and the flexibility of the metabolic network. The analysis of the toy models showed that the random weighting factors approach can predict relative fluxes through alternative pathways based on model stoichiometry, where such information cannot be obtained using flux minimisation with uniform weighting factor (Table 6.1). This characteristic of the random weighting factors approach was exploited in predicting the relative importance of various routes of AEF under different environmental and physiological conditions. The random weighting factors approach also allowed the identification of multiple possible metabolic routes for the dissipation of excess ATP and NADPH at high light intensity, whereas the uniform weighting factor approach can only identify metabolic routes with the least sum of fluxes per ATP or NADPH dissipated.

Most FBA studies use metabolic efficiency-based objective functions such as minimising nutrient intake or maximising growth, with an underlying assumption that the efficiency in nutrient use is maximised. While this assumption is valid in cells grown under nutrient-limited conditions, such objective functions are not applicable when metabolism is not nutrient-limited. Schuster et al., (2008) argued that maximisation of molar yield is not a universal principle: for example there is experimental support for the operation of fast but low-yielding metabolic pathways such as the Crabtree

effect in *Saccharomyces cerevisiae* in which fermentation is used under aerobic conditions when sufficient glucose is available. The results from this chapter showed that flux minimisation with random weighting factors can be applied to predict metabolic fluxes and explore possible metabolic routes in scenarios with excess nutrient.

In principle, the objective function of flux minimisation is a proxy for minimising the enzyme machinery costs to carry out the metabolic function(s) defined by the model constraints. Given that metabolic fluxes are the variables of the linear programming problem in FBA, the weighting factor of each reaction flux in the objective function of minimising enzyme costs corresponds to the enzyme cost per unit flux. However, this information is difficult to obtain for most reactions in the model as it involves the kinetic properties of the enzymes. A possible method in deriving enzyme cost per unit flux based on protein abundance data is described in chapter 7. In applying the random weighting factors approach, multiple sets of weighting factors were sampled to generate multiple flux solutions, from which the dependence of a reaction flux on the weighting factors can be inferred.

Besides predicting metabolic fluxes, the flexibility of the metabolic system can be investigated using the random weighting factors approach. The variability of fluxes explored using this approach is fundamentally different from the results from flux variability analysis (FVA). FVA calculates the maximum and minimum allowable flux of each reaction while satisfying the optimality of a given objective function (Mahadevan and Schilling, 2003). The random weighting factors approach identifies the variability of fluxes in multiple flux solutions, each generated with a different set of random

weighting factors. This is similar to the method of uniform random sampling of flux space developed by Price et al., (2004), which uniformly samples the entire feasible flux solution space. The key difference between the approach of flux minimisation with random weighting factors and uniform sampling is that the former has a bias towards identifying flux solutions with low sum of fluxes. An additional advantage of flux minimisation is that reaction cycles are eliminated from the flux solutions. In this chapter, the random weighting factors approach was utilised to show the flexibility in the carbon fixation pathway and in the accumulation of carbon storage in the light in a mature C_3 leaf. In addition, the metabolic reactions associated with these processes were identified by calculating the Pearson correlation coefficients between the reaction fluxes over multiple flux solutions generated by the random weighting factors approach.

6.3.2 Flux predictions of alternative electron flow and energy-dissipating reactions

The changes in fluxes through various routes of AEF with light intensity were predicted using flux minimisation with random weighting factors (Figure 6.2C). At low light intensity, the mitochondrial electron transport chain was predicted to be the major route of AEF. This is likely to be due to the high ATP yield per photon for the mitochondrial ETC compared to other AEFs. The mitochondrial ETC was calculated to produce 0.945 ATP per photon whereas cyclic electron transport and the water-water cycle produce 0.45 and 0.32 ATP per photon respectively (Kramer and Evans, 2011). The model prediction supports the notion that the mitochondrial ETC is important in sustaining photosynthetic carbon assimilation (Raghavendra and Padmasree,

2003). Cyclic electron transport was predicted to be active only at high light intensity. The preference of cyclic electron transport over the mitochondrial ETC at high light intensity is presumably because cyclic electron transport provides a shorter metabolic route for ATP generation when the ATP yield is not important. The model prediction is consistent with experimental observations in cyclic electron transport mutant (Munekage et al., 2008) and the notion that cyclic electron transport is important in plant responses to stress (Johnson, 2005). Controversies exist on whether the water-water cycle is a major route of electron flow in leaves under sunlight (Heber, 2002). The model prediction suggests that the water-water cycle plays a relatively minor role in leaves in the light compared to cyclic electron flow. The water-water cycle is not favoured at low light intensity as it has a lower ATP yield compared to the other two AEFs. At high light intensity, cyclic electron transport is preferred over the water-water cycle because fewer reactions are required to run cyclic electron transport than the water-water cycle. With respect to nitrogen assimilation, fluxes through various routes of AEF relative to the photosynthetic linear electron transport decrease with increasing nitrate assimilation (Figure 6.3C) as a result of the decrease in the ratio of ATP demand to NADPH demand (Figure 6.3B). This is in line with the calculations by Noctor and Foyer, (1998) which suggested that nitrate assimilation contributes to the balance of ATP to NADPH for carbon assimilation. The flux predictions of different stages of leaf development showed an increase in flux through the mitochondrial ETC relative to the photosynthetic linear electron transport from developing leaf to mature leaf as the net nitrogen assimilation per carbon decreases (Figure 6.4C; Table 6.2).

The results suggest that nitrate assimilation competes for reducing power with the mitochondrial ETC. By consuming more reductant per ATP than the ratio generated by the photosynthetic linear electron transport, the net effect of nitrate assimilation is an increase in the ratio of ATP supply per NADPH for other metabolic process. In a flux-balance analysis study of AEFs in *Synechocystis*, nitrate assimilation was considered as one of the AEFs (Nogales et al., 2012). In this chapter, it was shown that the modelling approach can be used to investigate the changes in coenzyme demands with varying environmental or physiological conditions and the metabolic changes required to meet the altered demands.

At high light intensity, the model predicted high fluxes through several energy and reductant dissipating reactions. In terms of dissipating reductant in the chloroplast, three metabolic pathways carried high fluxes including the chlorophyll cycle, the glutathione-ascorbate cycle and the xanthophyll cycle (Figure 6.2D). The xanthophyll cycle is well known for its role in photoprotection by dissipating the excess light energy as heat (Demmig-Adams and Adams, 1996). The glutathione-ascorbate cycle is associated with the turnover of ascorbate in both the xanthophyll cycle and the water-water cycle. Interestingly, the chlorophyll cycle was predicted to carry the highest flux out of the three pathways. The major function of the chlorophyll cycle is thought to be the synthesis and catabolism of chlorophyll a and chlorophyll b (Tanaka and Tanaka, 2011). It is not known whether the chlorophyll cycle plays a significant role in energy dissipation *in vivo*. From a structural metabolic modelling perspective, the xanthophyll cycle was less favoured than the chlorophyll cycle because additional reactions in the

ascorbate recycling pathway are required to complete the xanthophyll cycle. The modelling results suggest that the preference of xanthophyll cycle *in vivo* is likely to be due to kinetic and/or regulatory aspects rather than metabolic network structure. In mitochondria, the alternative respiratory pathway was predicted to carry flux at high light intensity to dissipate NADH produced by glycine decarboxylase in the photorespiratory pathway (Figure 6.2E). The mitochondrial ETC was predicted to carry a small flux at high light intensity (Figure 6.2C), where the proton gradient generated can be dissipated by the uncoupling protein (Figure 6.2E). From a study of an uncoupling protein mutant in *Arabidopsis*, it was shown that the uncoupling protein is required for efficient photosynthesis by maintaining the redox poise of the mitochondrial ETC (Sweetlove et al., 2006). The model seemed to underestimate the importance of the uncoupling protein and favour the alternative respiratory pathway presumably due to a lower overall sum of fluxes. However, it was suggested that both mechanisms are likely to operate *in vivo* as the mitochondrial ETC is always in competition with the alternative respiratory pathway (Sweetlove et al., 2006).

6.3.3 Model predictions suggest distinct functions for enzymes in ammonia assimilation

An isoform of glutamine synthetase (GS) was discovered to be dual targeted to the chloroplast and the mitochondria in *Arabidopsis* leaves (Taira et al., 2004). From the analysis on varying nitrate influx and on the different stages of leaf development, the model predicted differential responses for GS in different subcellular compartments (Figure 6.3D; Figure 6.4E; Table 6.2). The results suggest that they carry out distinct metabolic functions where the

chloroplastic GS is primarily involved in ammonia assimilation for *de novo* amino-acid synthesis and the mitochondrial GS is mainly involved in the refixation of ammonia from the photorespiratory pathway. This could be related to the compartmentation of the sources of ammonia in the two processes. The reduction of nitrite to ammonia, by nitrite reductase, for *de novo* amino-acid synthesis occurs in the chloroplast whereas the release of ammonia from the photorespiratory pathway by glycine decarboxylase occurs in the mitochondria. However, the operation of the mitochondrial GS is dependent on the presence of a glutamate-glutamine shuttle, which is hypothesised by Linka and Weber, (2005), to transfer glutamate and glutamine between the mitochondria and the chloroplast. Such shuttling mechanism is yet to be identified *in vivo*.

Another enzyme which is involved in the assimilation of ammonia is glutamate synthase (GOGAT) which is localised in the chloroplast. In plants, there are two distinct forms of GOGAT, Fd-GOGAT and NADH-GOGAT which use reduced ferredoxin and NADH as the electron donors respectively (Suzuki and Knaff, 2005). The flux solutions in this chapter and chapter 5 showed that the use of Fd-GOGAT was preferred over NADH-GOGAT in a photosynthetic leaf (Appendix S10). This is in contrast to the flux solutions in chapter 4, in which NADH-GOGAT was preferred over Fd-GOGAT in heterotrophic cell culture (Appendix S7). The tissue-specific preference in electron donors is reflected by the sources of reducing power in the different tissues. Reduced ferredoxin from photosynthetic linear electron transport is the main source of reducing power in leaves in the light whereas NADH from the catabolism of glucose via glycolysis and pyruvate dehydrogenase in the

plastid is the major source of reductant in heterotrophic cell culture. The examples of GS and GOGAT showed that constraint-based modelling can be utilised to elucidate the distinct metabolic roles of enzymes in different subcellular compartments and the tissue-specific use of distinct forms of enzymes with different coenzyme specificity.

6.3.4 Summary

In this chapter, a novel modelling method using flux minimisation with random weighting factors was developed to predict metabolic fluxes in a C₃ leaf and to investigate the relative importance of various routes of alternative electron flow under different conditions. With varying light intensity, the mitochondrial electron transport chain was predicted to be important at low light intensity whereas cyclic electron flow only carried flux at high light intensity. From the model predictions on varying nitrogen availability and on different stages of leaf development, it was hypothesised that glutamine synthetase in different subcellular compartments may carry out distinct metabolic functions in a photosynthetic leaf. The chloroplastic glutamine synthetase was suggested to be involved in ammonia assimilation for the *de novo* synthesis of amino-acids whereas the refixation of ammonia from the photorespiratory pathway was suggested to be carried out by the mitochondrial glutamine synthetase. In addition to predicting metabolic fluxes, the random weighting factors approach was utilised to investigate the metabolic flexibility of a C₃ leaf in the carbon fixation pathway and carbon storage accumulation in the light.

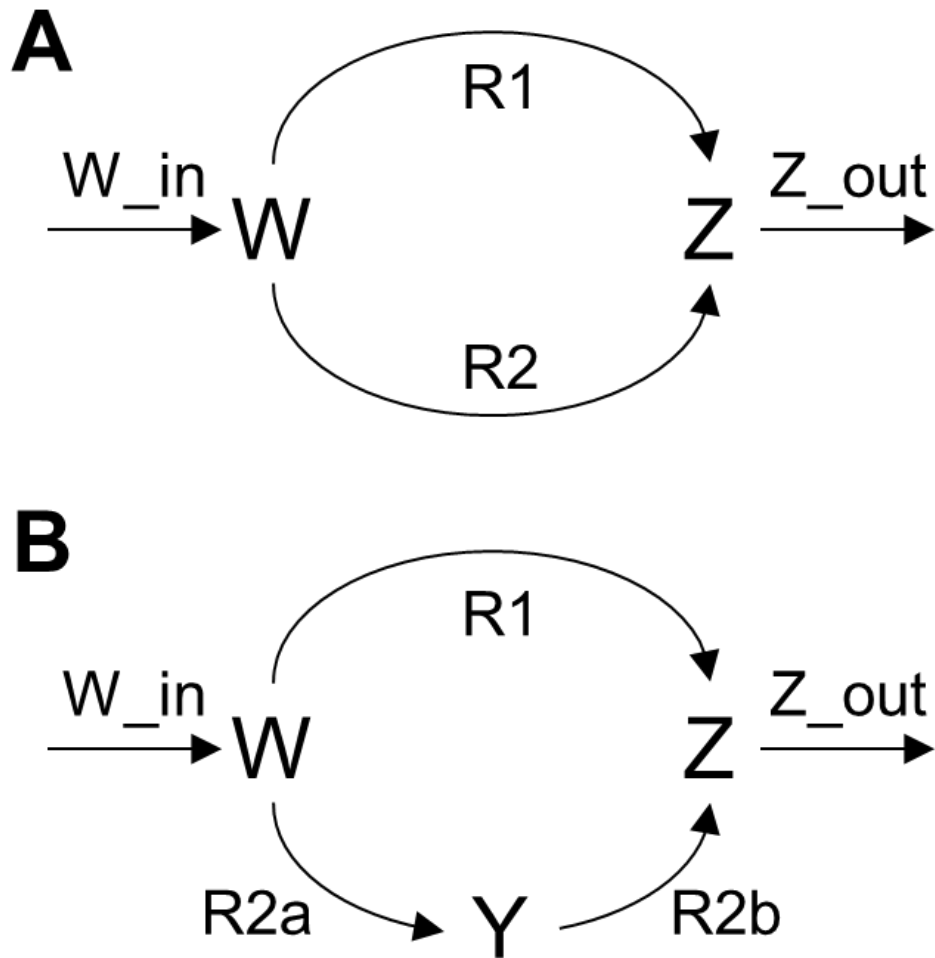


Figure 6.1 Toy models for investigating flux minimisation with random weighting factors. (A) Toy model 1 contains two alternative routes with the same number of reactions and sum of fluxes. (B) Toy model 2 contains two alternative routes where one route contains one reaction more than the other.

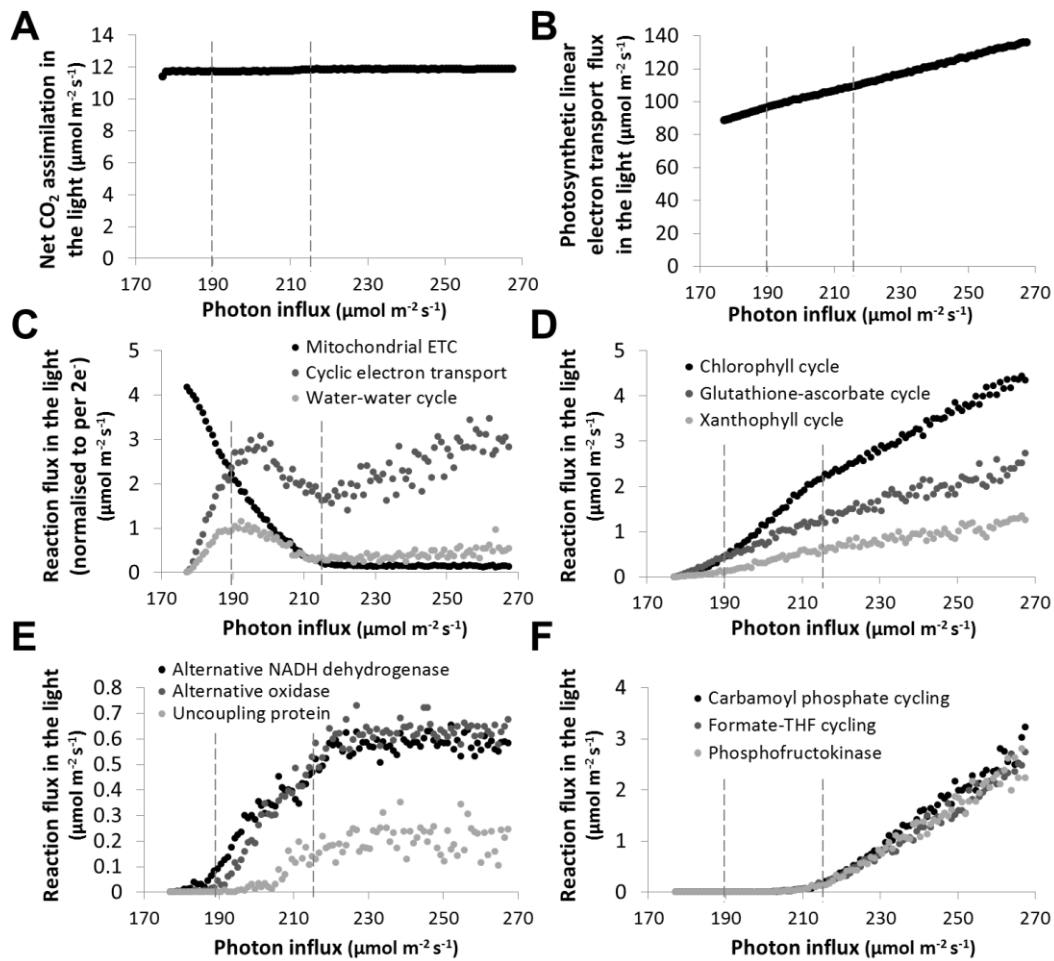


Figure 6.2 Model predictions in the light with varying photon influx. (A) Net CO₂ assimilation. (B) Photosynthetic linear electron transport flux. (C) Fluxes through different routes of alternative electron flow including the mitochondrial electron transport chain (black), cyclic electron transport (dark grey) and the water-water cycle (light grey) normalised to per two electrons. (D) Fluxes through reductant-dissipating metabolic routes including the chlorophyll cycle (represented by the “RXN-7679_Light_p”) (black), the glutathione-ascorbate cycle (represented by “GLUTATHIONE-REDUCT-NADPH-RXN_Light_p”) (dark grey) and the xanthophyll cycle (represented by “RXN-7979_Light_p”) (light grey). (E) Fluxes through reactions in the mitochondrial alternative respiratory pathway, including the alternative NADH dehydrogenase (black) and alternative oxidase (dark grey), and the uncoupling protein (light grey). (F) Fluxes through reactions involved in ATP dissipation including cycling of carbobamyl phosphate (represented by “CARBAMATE-KINASE-RXN_Light_p”) (black), cycling of formate-tetrahydrofolate (THF) (represented by “FORMYLTHFDEFORMYL-RXN_Light_p”) (dark grey) and chloroplastic phosphofructisokinase (cycling of fructose 1,6-bisphosphate) (empty circles). Dashed vertical lines represent boundaries between phases 1, 2 and 3 (from left to right).

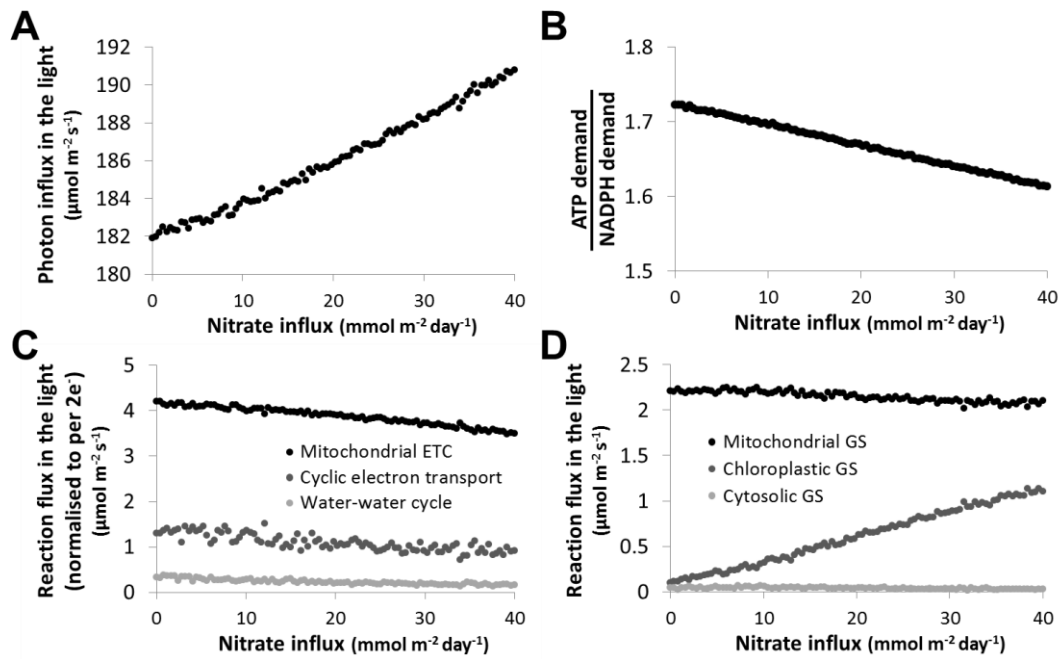


Figure 6.3 Model predictions in the light with varying nitrate influx. (A) Photon influx in the light. (B) The ratio of ATP demand to NADPH demand of the metabolic system. (C) Fluxes through different routes of alternative electron flow including the mitochondrial electron transport chain (black), cyclic electron transport (dark grey) and the water-water cycle (light grey) normalised to per two electrons. (D) Fluxes through glutamine synthetase in the mitochondrion (black), the chloroplast (dark grey) and the cytosol (light grey).

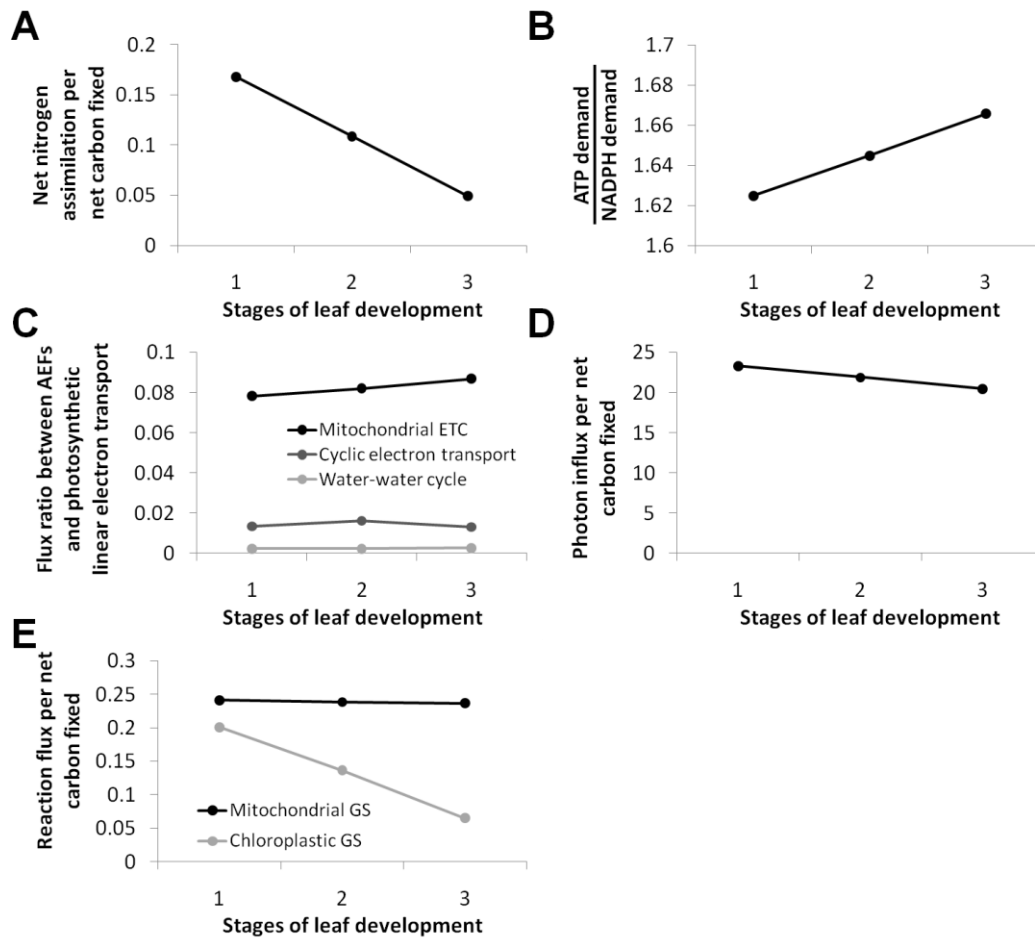


Figure 6.4 Model predictions in the light at different stages of leaf development. Stages of leaf development labelled 1, 2 and 3 represent developing leaf, maturing leaf and mature leaf respectively (see main text and Table 6.2 for growth stage-specific model constraints). (A) Net nitrogen assimilation per net carbon fixed. (B) The ratio of ATP demand to NADPH demand of the metabolic system. (C) Flux ratios between different routes of alternative electron flow (AEF) and the photosynthetic linear electron transport, where the AEFs include the mitochondrial electron transport chain (black), cyclic electron transport (dark grey) and the water-water cycle (light grey) normalised to per two electrons. (D) Photon influx in the light per carbon fixed. (E) Fluxes through glutamine synthetase in the mitochondrion (black) and the chloroplast (dark grey) per net carbon fixed.

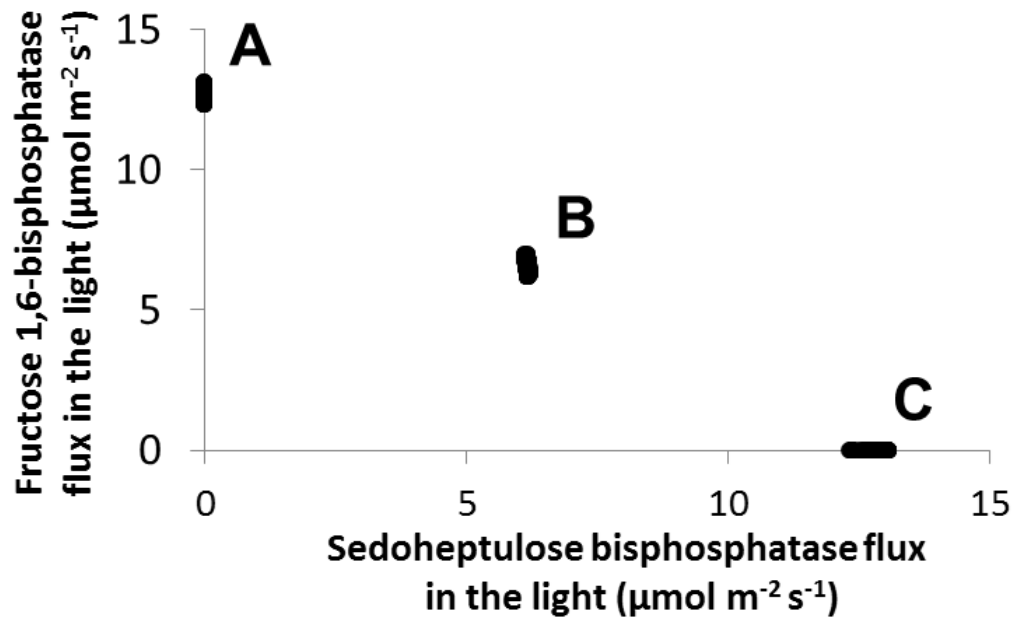


Figure 6.5 Scatter plot of fluxes through sedoheptulose biphosphatase and fructose 1,6-bisphosphatase in the chloroplast in the light. Flux solutions of a mature C_3 leaf were generated using flux minimisation with 1000 different sets of random weighting factors. The flux solutions with respect to the two phosphatases grouped into three distinct clusters, labelled A, B and C, which correspond to the metabolic flux modes in Figure 6.6.

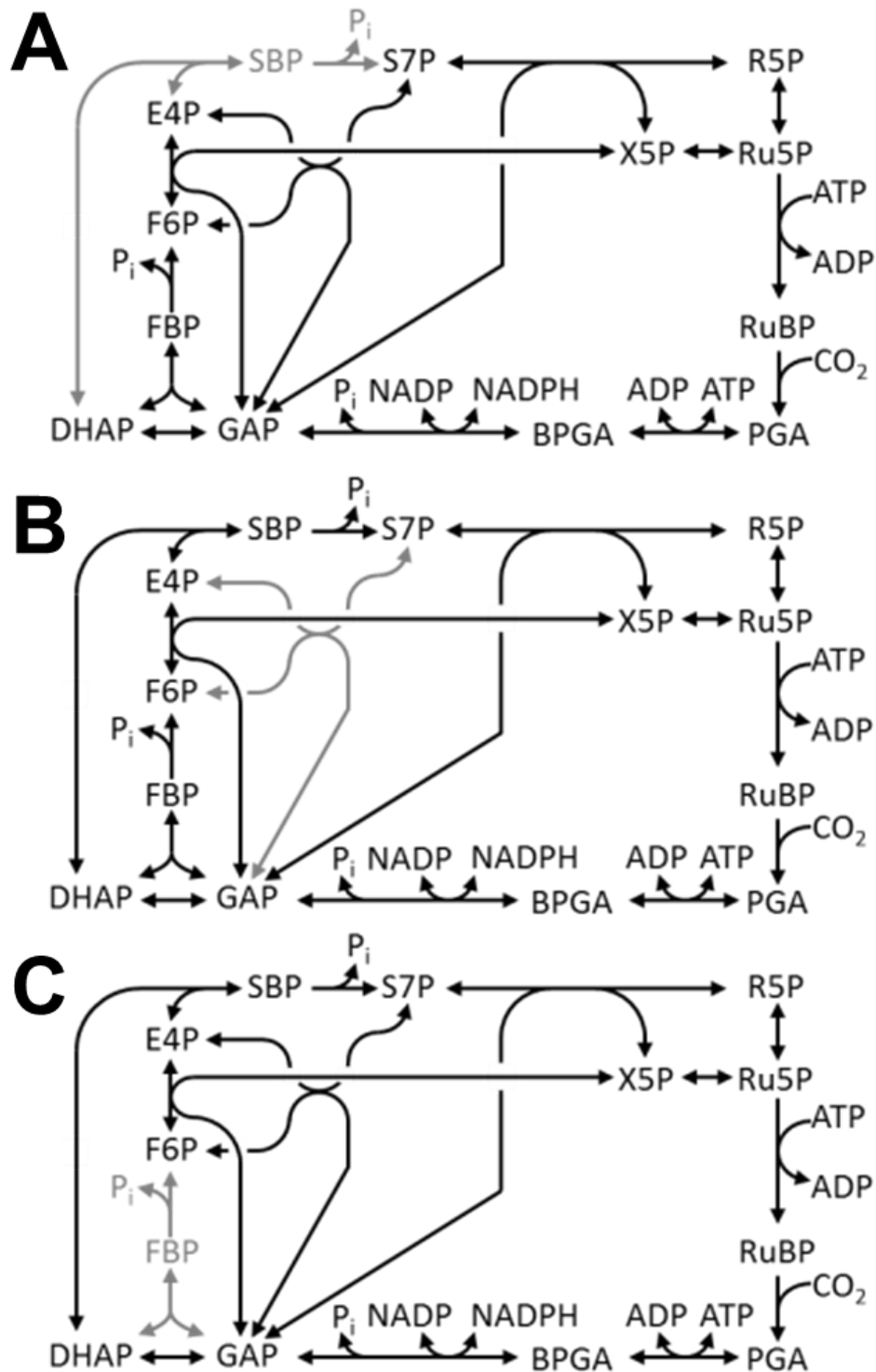


Figure 6.6 Metabolic flux modes of carbon fixation pathway corresponding to flux solutions from flux minimisation with random weighting factors. Three metabolic flux modes (A, B and C) correspond to the flux solutions in Figure 6.5. The middle flux mode (B) matches with the standard textbook reductive pentose phosphate pathway with fluxes through both phosphatases. The other two flux modes (A and C) have no flux through one of the phosphatases.

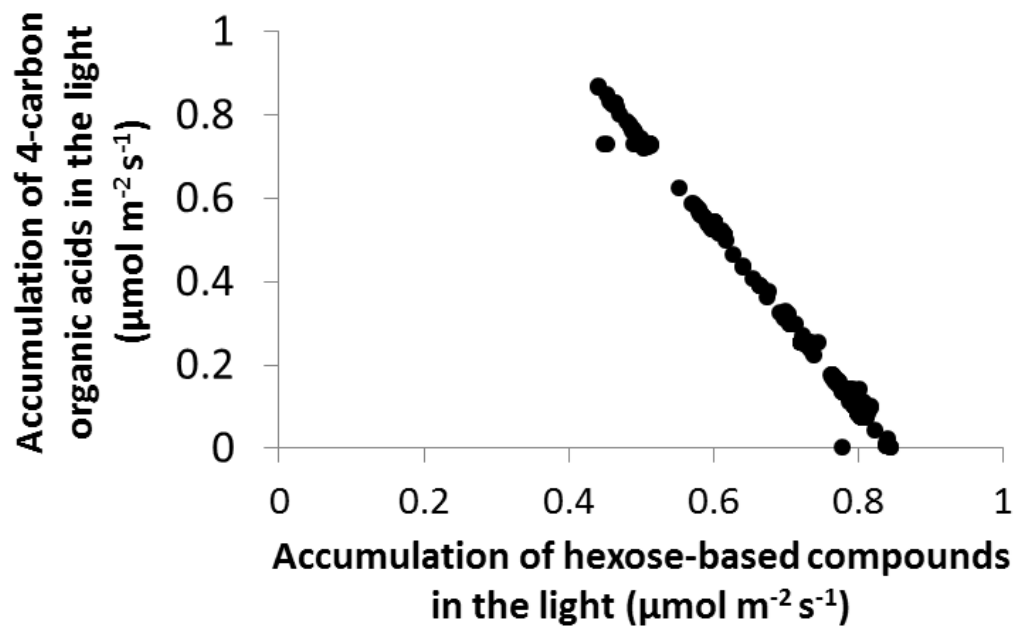


Figure 6.7 Scatter plot of accumulations of hexose-based storage compounds and 4-carbon organic acids in the light. Flux solutions for a mature C_3 leaf were generated using flux minimisation with 1000 different sets of random weighting factors. This plot illustrates the flexibility of the metabolic network in the amount of different storage compounds while supporting the physiological functions of a mature leaf.

Table 6.1 Comparisons of flux predictions of toy models using different modelling approaches. FBA represents the standard flux balance analysis approach using flux minimisation with uniform weighting factor. Flux ranges from flux variability analysis (FVA) were represented as ($v_{\min}$, $v_{\max}$) where $v_{\min}$ and $v_{\max}$ are the minimum and maximum feasible flux values compatible with the optimal objective value. Random weighting factor column contains the averaged flux solution generated from flux minimisation with 1000 different sets of random weighting factor.

		FBA (uniform weighting factor)	FVA	Random weighting factor
Toy model 1	R1	1	(0,1)	0.496
	R2	0	(0,1)	0.504
Toy model 2	R1	1	(1,1)	0.839
	R2a	0	(0,0)	0.161
	R2b	0	(0,0)	0.161

Table 6.2 Model predictions of leaf metabolism at different growth stages. Stages 1, 2 and 3 represent developing leaf, maturing leaf and mature leaf, respectively, as described in section 6.2.4. The different growth stages were modelled by varying the proportion of net fixed carbon which goes into biomass synthesis and phloem export. Net nitrogen assimilation per net carbon fixed decreases with leaf growth with a corresponding increase in the ratio of ATP demand to NADPH demand of the metabolic system (see chapter 2 for details in calculations of coenzyme demands). The ratio of flux through the mitochondrial electron transport chain (ETC) to the photosynthetic linear electron transport (normalised to per two electrons) increased with leaf development, whereas the ratios of cyclic electron transport and the water-water cycle to the photosynthetic linear electron transport (normalised to per two electrons) remained relatively constant. There was a decrease in photon influx per net carbon fixed with leaf development. The mitochondrial glutamine synthetase flux remained relatively constant whereas the chloroplastic glutamine synthetase flux decreased with leaf growth.

Stages of leaf development	1	2	3
Proportion of net fixed carbon goes into biomass synthesis	1	0.5	0
Proportion of net fixed carbon goes into phloem export	0	0.5	1
Net nitrogen assimilation per net carbon fixed	0.168	0.109	0.049
ATP demand : NADPH demand	1.620	1.645	1.666
Mitochondrial ETC : Linear electron transport	0.078	0.082	0.087
Cyclic electron transport : Linear electron transport	0.013	0.016	0.013
Water-water cycle : Linear electron transport	0.002	0.002	0.003
Photon influx per net carbon fixed	23.307	21.894	20.457
Mitochondrial GS flux per net carbon fixed	0.241	0.239	0.237
Chloroplastic GS flux per net carbon fixed	0.201	0.137	0.065

7 General discussion

7.1 Summary of results

In this study, an updated and compartmented genome-scale model of Arabidopsis was constructed (chapter 3) and utilised to predict metabolic fluxes in heterotrophic cell culture (chapter 4) and photosynthesising leaves (chapters 5 and 6).

In chapter 3, a compartmented genome-scale metabolic model of Arabidopsis was constructed with 2769 reactions and 2739 metabolites where energetic costs for metabolite transport, polymerisation of macromolecules and other growth-dependent energy-requiring processes were defined as far as possible. During model construction, an iterative process of model curation was carried out to correct the direction and reversibility of reactions involved in thermodynamically unfavourable reaction cycles that generate ATP or NAD(P)H. A similar process was done to eliminate transhydrogenase cycles transferring reducing power from NADH to NADP⁺. This was shown to be important in terms of model prediction as the definition of the direction and reversibility of a single reaction can have a large impact on the resulting flux solution, illustrated by the example of NADP-dependent dehydrogenases in chapter 4.

With the energetic costs for growth-dependent processes included in the genome-scale metabolic model, cellular maintenance costs in heterotrophic cell culture under a control and two stress conditions were estimated in terms of ATP and reductant (chapter 4). The results showed that elevated temperature led to an increase in maintenance costs whereas hyperosmotic

conditions caused a decrease in maintenance costs, suggesting that the cells were not stressed under hyperosmotic conditions. By comparing between the genome-scale model predictions of metabolic fluxes in heterotrophic cell culture from flux balance analysis (FBA) and experimentally estimated flux maps obtained by ^{13}C -metabolic flux analysis (MFA), it was shown that the inclusion of transport and maintenance costs was important in obtaining accurate flux predictions. Given the application of appropriate model constraints, the five efficiency-based objective functions investigated in this study were shown to give predictions that matched well with experimentally measured fluxes.

In chapter 5, a diel modelling framework was developed to model leaf metabolism over a day-night cycle by considering the interactions between light and dark metabolism. The importance of modelling light and dark metabolism simultaneously in photosynthetic flux balance models was illustrated by the example of glutamate and glutamine synthesis. Based on a stoichiometric model with minimal model constraints and the objective function of minimising total flux, numerous known features of metabolism in a C_3 leaf were predicted such as the nocturnal accumulation of citrate utilised for diurnal glutamate and glutamine synthesis and the operation of an incomplete TCA cycle during the day. Using the Arabidopsis genome-scale metabolic model and the diel modelling framework, the operation of the CAM cycle was predicted as a consequence of constraining CO_2 exchange with the environment to zero during the day to simulate closure of the stomata. The comparisons between model predictions of C_3 and various subtypes of CAM leaves suggested that photon and nitrogen use efficiencies are unlikely

to be the driving forces for the evolution of CAM plants under the current atmospheric CO₂ concentration.

Using the objective function of flux minimisation with random weighting factors, the metabolic fluxes in a C₃ leaf, with emphasis on the relative importance of various routes of alternative electron flow (AEF), were predicted with varying light intensity, nitrogen availability and at different stages of leaf development (chapter 6). The mitochondrial electron transport chain (ETC) was predicted to be the major route of AEF at low light intensity whereas fluxes through the cyclic electron transport and various ATP- and NADPH-dissipating reactions increase at high light intensity. From the model flux predictions, it was shown that constraint-based modelling can be utilised to elucidate the distinct metabolic roles of enzymes in different subcellular compartments and the tissue-specific use of distinct forms of enzymes with different coenzyme specificities, illustrated by the examples of glutamine synthetase and glutamate synthase.

7.2 Scope of FBA in modelling plant metabolism

FBA is a powerful modelling technique in that metabolic fluxes of a large-scale metabolic network can be predicted based solely on reaction stoichiometry, a few constraints and an objective function. The model constraints are mainly the inputs and outputs of the metabolic systems. In the case of plants, the main physiological constraints applied on FBA models are the biomass synthesis for a growing cell or tissue, the metabolic exchanges with xylem and phloem, and exchanges with the environment such as light and gaseous exchanges in photosynthetic organs and the uptake of

inorganic ions from the soil in roots. Given a set of input and output constraints, the results in chapter 4, 5 and 6 showed that FBA was able to accurately predict phototrophic and heterotrophic metabolism using the *Arabidopsis* genome-scale metabolic model with an efficiency-based objective function – flux minimisation. The results in this study together with studies in developing plant seeds and photoautotrophic microorganisms suggest that efficiency-based objective functions may be appropriate for modelling plants and other photosynthetic organisms (Chen and Shachar-Hill, 2012).

The efficiency in solving a linear programming problem to compute an FBA flux solution, typically in milliseconds to seconds, has allowed the investigation of metabolic systems to be performed at a much larger scale than before. The applicability of FBA on large-scale metabolic models, e.g. the diel model in chapters 5 and 6 with more than 5500 reactions, enabled the analysis of systems-wide properties of the whole metabolic network such as the balance between supply and demand of a number of coenzymes. Previous calculations on ATP and NADPH budgets for photosynthesis only took into account a relatively small number (<100) of reactions (Winter and Smith, 1996; Noctor and Foyer, 1998), but with genome-scale metabolic models, a much more complete picture of the overall ATP and NADPH budgets of the metabolic network could be obtained. Results in chapter 6 demonstrated that numerous possible metabolic routes for balancing coenzymes supply and demand could be identified using FBA with a genome-scale metabolic model.

In contrast to MFA, which measures metabolic fluxes at particular metabolic states, the scalability of FBA means that it can be applied to predict continuous metabolic responses to changes in one or more conditions. Assuming that plant metabolic systems are evolved to be plastic in adapting to variation in conditions, FBA is likely to give reasonable predictions in metabolic responses based on an optimisation principle. Such analysis was done in plants genome-scale metabolic models to look at the metabolic changes with different ATP maintenance costs in *Arabidopsis* heterotrophic cell culture (Poolman et al., 2009) and with varying light intensities in a developing leaf cells of rice (Poolman et al., 2013). In chapter 6, the changes in metabolic fluxes in a mature C_3 leaf with varying light intensities and nitrogen availability were predicted by ‘scanning’ across individual constraints. This can be extended to predict the responses to changes in two or more constraints simultaneously by ‘scanning’ across multiple constraints, where the case with two constraints was termed phase plane analysis (Edwards et al., 2001). In an FBA model of barley seed, phase plane analysis was carried out with varying oxygen uptake rate and sucrose uptake rate, where the changes in metabolic fluxes between the five phases of the phase plane, from anoxia to hypoxia to hyperoxia, were predicted (Grafahrend-Belau et al., 2009). This approach allows the interactions between changes in multiple environmental and/or physiological factors to be investigated at the level of metabolic fluxes.

There is synergy between MFA and FBA in that the flux maps characterised by MFA can be used as a reference for curating the FBA model and choosing suitable modelling parameters such as the combination of constraints and

objective functions. This was done in chapter 4 where the effect of combinations of constraints and objective functions on the accuracy of flux prediction was investigated in the system of heterotrophic cell culture. Flux predictions of C₃ leaf metabolism also matched well with results from isotopic labelling experiments (chapter 5). These results give confidence in the genome-scale metabolic model and the modelling framework in performing predictive modelling to explore the flux distributions in multiple cell types and tissues at different conditions. This is especially relevant for scenarios where measuring intracellular fluxes by MFA is experimentally difficult such as in a photosynthetic leaf. Using the diel modelling framework, the Arabidopsis genome-scale metabolic model gave predictions on the possible metabolic route of citrate metabolism in the light in CAM plants (chapter 5) and suggested distinct functions of glutamine synthetase localised in different subcellular compartments (chapter 6).

7.3 Limitations of FBA in modelling plant metabolism

The numerous advantages of FBA that stem from the simplicity of its formulation also come with a number of limitations. The steady-state assumption is generally not valid in plants at the scale of days as plant metabolism is heavily influenced by the diel cycle where large metabolic reconfigurations occur over a day-night cycle. One solution to this problem is to divide a day-night cycle into multiple phases and assume a pseudo-steady-state in each phase. Each phase can be simulated independently by setting time-specific constraints and/or objective function, which was done in

a recent FBA study on *Synechocystis* using a time-dependent biomass objective function (Knoop et al., 2013). The difficulty in applying this approach is that it requires the knowledge of time-dependent biomass compositions and quantitative changes in metabolite levels as constraints and/or objective function. An alternative approach was developed in chapter 5 in which the multiple phases were modelled simultaneously by allowing storage compounds to accumulate in one phase and be metabolised in the next phase. A limitation of this approach is that a set of storage compounds between each phase and the phase-specific physiological constraints must be pre-defined based on prior knowledge.

Another bottleneck in applying FBA to model plant metabolism is the difficulty in obtaining experimentally measured values for some model constraints. For example, chapter 4 demonstrated the use of an experimentally measured flux ratio for estimating ATP and NADPH maintenance costs, but such a method may be not applicable in other tissues such as leaves. The rates of metabolite exchanges between different organs and the xylem and phloem are experimentally difficult to measure and are likely to vary depending on conditions. Another layer of complication is that the compositions of xylem and phloem are likely to vary both temporally and spatially. As a first approximation, the phloem composition was assumed to be constant in most of the analysis in chapter 5 and 6, except for the analysis on nitrogen availability where the ratio of sucrose to amino-acid changed depending on nitrate uptake. While this study explored the use of new modelling techniques in predicting metabolic fluxes in leaves, accurate experimentally determined constraints will be required to obtain quantitatively meaningful predictions.

The optimisation principle in FBA assumes that the operation of the metabolic system satisfies an evolutionary goal that can be captured by the chosen objective function. Most commonly used objective functions in FBA are efficiency-based, therefore metabolic routes that are more efficient are chosen in preference to alternative routes. While this assumption was shown to be applicable for many parts of the metabolic system, e.g. numerous known features of leaf metabolism were predicted based on flux minimisation in chapter 5, there are exceptions to this optimisation principle. An example of this is the failure in predicting the hydrolytic pathway for starch degradation in C_3 plants discussed in chapter 5. The phosphorolytic pathway is energetically more efficient and hence preferred by FBA given an efficiency-based objective function. However, it has been suggested that the hydrolytic pathway may serve as a regulatory mechanism in directing carbon into sucrose synthesis in the cytosol as opposed to being catabolised via glycolysis in the chloroplast (Weise et al., 2011). The solution to this particular case was to set the chloroplastic hexose phosphate transport reaction to carry zero flux based on experimental data. Such behaviour could not be predicted as regulatory information is not included in the standard formulation of FBA. However, the discrepancy between FBA flux solutions and known biochemistry may yield insights into the functional roles of metabolic routes.

Another example of the potential pitfall in applying the optimisation principle is that metabolic fluxes could be constrained by the kinetics of the enzymes, which have not evolved to carry metabolic fluxes that are optimal towards a certain objective function. A notable example is Rubisco which catalyses both

the carboxylation and oxygenation of ribulose 1,5-bisphosphate. While the carboxylase reaction is the key step in CO₂ fixation in plants, the oxygenase reaction is a side-reaction that reduces photosynthetic efficiency. Without any constraint on the Rubisco carboxylase to oxygenase ratio, the oxygenase reaction tends to carry no flux based on an efficiency-based objective function (though curiously the Rubisco oxygenase reaction was predicted to carry some flux in heterotrophic cell culture in a previous *Arabidopsis* model (Poolman et al., 2009) and in a developing rice leaf cell under high light intensity (Poolman et al., 2013)). However, the kinetics of Rubisco in C₃ plants gives a carboxylase to oxygenase ratio of around 3-4 : 1 under the current atmospheric composition. Since FBA models do not contain any information on kinetic parameters and metabolic concentrations, such a flux ratio could not be predicted using FBA. In this study, a constant ratio of carboxylase to oxygenase of 3 : 1 was set in simulations of C₃ leaves based on empirical data. An alternative method is to define and solve kinetic rate equations to calculate the Rubisco carboxylase to oxygenase ratio as a function of CO₂ and O₂ concentrations and apply the flux ratio as an additional constraint on the FBA model (Wang, C. et al., 2012).

A single optimal flux solution is often taken as the result of an FBA simulation, but in fact, the result of an optimisation step in FBA is an optimal flux solution space which contains the set of optimal flux solutions. For large-scale metabolic models, it is likely that multiple optimal solutions exist where some reactions can carry a range of flux values within the optimal flux solution space. This is especially relevant for plant models as metabolic reactions, such as those in glycolysis and parts of the TCA cycle, are present in multiple

subcellular compartments. While the variability in flux values, which can be characterised by flux variability analysis (Mahadevan and Schilling, 2003), presents an obstacle in obtaining a well-defined flux prediction, changes in flux variability type can supplement the information deduced from the changes in uniquely defined fluxes (Sweetlove and Ratcliffe, 2011) demonstrated by an FBA study on oilseed rape embryos (Hay and Schwender, 2011a,b).

Although there are several limitations in applying the standard formulation of FBA in modelling plant metabolism, methods are being developed, including those in this study, to overcome these hurdles to increase the applicability and predictive accuracy of FBA in modelling metabolic fluxes in different plant cells and tissues under multiple conditions.

7.4 Future directions in flux balance modelling of plant metabolism

In chapter 5 and 6, flux balancing analysis with the Arabidopsis genome-scale metabolic model was used to model the net metabolic fluxes in a whole leaf. But the leaf is a complex organ composed of multiple cell types where each cell type has a distinct flux distribution. In addition, there are multiple interactions between different cell types by means of metabolite exchanges. Cell type-specific metabolism and the interactions between two cell types in plants has been modelled for the mesophyll cell and the bundle sheath cells in C_4 plants (de Oliveira Dal'Molin et al., 2010b; Saha et al., 2011) by setting cell type specific constraints, e.g. Rubisco flux was set to zero in the mesophyll cells, and allowing a pre-defined set of metabolites that can be

transported between the cell types. Such analysis can potentially be extended to other cell types such as guard cells and phloem companion cells. This will provide insights into the distinct metabolic roles played by each enzyme in different cell types as well as predictions of metabolite transport between cell types which are difficult to measure experimentally. Besides modelling leaf metabolism, similar approach can be applied to other plant tissues, e.g. three tissue-specific models of oilseed rape seeds were used to investigate the contributions of light energy to the metabolic demands of the three tissues (Borisjuk et al., 2013).

Stepping back from a single organ, a whole plant metabolic model can be constructed by connecting models of individual tissues and organs by metabolite exchanges through the xylem and phloem. A major step forward was taken by Grafahrend-Belau et al. (2013) where three large-scale organ-specific FBA models for leaf, stem and seed, and a simplified root model were integrated into a multi-organ model with the phloem as the exchange component mediating the transport of carbon and nitrogen between the different plant organs. In the same study, a dynamic multi-scale functional plant model was utilised to generate constraints for the multi-organ FBA model throughout the generative phase of barley plants for studying source-to-sink interactions in barley. In theory, this approach can be extended to model the whole course of plant development from germination to seed development, possibly including flowering. To achieve this, further efforts must be put into model curation, not just on the synthesis of biomass components, but also their catabolism as the mobilisation of biomass

components is important in various tissues at different developmental stages as illustrated by the example of the failure to model a sink leaf in chapter 6.

To extend the scope of modelling in the spatial dimension, plant metabolic models can be coupled with models of microorganisms to investigate the metabolic interactions between plants and microbes. For example, genome-scale metabolic models of *Medicago* (a legume) and *Rhizobia* (a nitrogen fixing bacteria) were constructed to investigate the metabolic exchanges between the two organisms (Pfau, T., personal communication). A similar analysis could potentially be performed to investigate metabolic interactions between a host plant and a pathogen, as was done for the host-pathogen interactions between human alveolar macrophage and *M. tuberculosis* (Bordbar et al., 2010).

In the temporal dimension at the scale of hours, a novel modelling framework was developed in this study to model the distinct metabolic behaviour of light and dark metabolism by dividing a diel cycle into two phases (chapter 5). In theory, this modelling framework is extendable to more than two phases given the phase-specific constraints and the definition of storage compounds between two consecutive phases. For example, the metabolism of guard cells could be modelled where K^+ , malate and Cl^- accumulate from the beginning of the light period as the main osmolytes, which are replaced by sucrose in the second half of light period (Talbot and Zeiger, 1998). In chapter 5, phase I and phase III of CAM photosynthesis were modelled which captured the core of the CAM cycle. The extension of the diel modelling framework to include phase II and phase IV could potentially give insights into the functional significance of the two phases in CAM photosynthesis. The

coupling between a multi-cell type model of C₄ mesophyll and bundle sheath cells and the diel modelling framework may provide possible explanations to the experimental observation of preferential starch accumulation in bundle sheath cells in mature C₄ leaves (Weise et al., 2011).

Since the availability of genome-scale metabolic models, efforts have been made to integrate multiple -omics data into the framework of genome-scale metabolic modelling (Oberhardt et al., 2009; Yizhak et al., 2010). With regard to plants, transcriptomics data was coupled to a genome-scale metabolic model of Arabidopsis as bounds of fluxes to reveal metabolic pathways involved in light and temperature acclimation (Töpfer et al., 2013).

Information on protein expression data was used to construct 10 tissue-specific models of Arabidopsis (Mintz-Oron et al., 2011). While the use of transcriptomics and proteomics data in generating cell type- or tissue-specific models based on the presence or absence of gene products is justifiable, taking the quantitative values of such datasets as constraints of metabolic fluxes is questionable as there is no clear quantitative relationship between transcript or protein abundance and metabolic flux. Although protein abundance data cannot be directly related to flux values, it can be used to infer the enzyme machinery costs in carrying the metabolic fluxes. In a reference experimental system where both absolute protein abundance data and flux data, either measured using MFA or predicted from FBA based on physiological constraints, are available, the amount of protein required per unit flux can be calculated as the quotient of the absolute protein abundance and the flux value. This measure could potentially be informative in terms of the kinetic properties of the enzymes, and more importantly with respect to

FBA, it can be used as the weighting factors for flux minimisation as the objective function of minimising the total enzyme machinery costs. By assuming that the enzyme machinery costs per unit flux remains constant, the set of weighting factors can be used for modelling metabolism under different conditions. Despite the lack of correlation between metabolite concentrations and metabolic fluxes, quantitative time-course metabolomics data can be used as FBA model constraints to simulate the metabolic fluxes between two time points, be it at the scale of hours in the case of the diel cycle or at the scale of days in the case of plant development.

One of the major goals in performing genome-scale modelling in plants is to guide metabolic engineering of plants for crop improvement, both quantitatively and qualitatively. With a high quality metabolic model, one can perform *in silico* metabolic engineering to identify possible targets and engineering strategies to achieve a desirable metabolic phenotype. In this study, a well-curated and tested genome-scale metabolic model was constructed and several modelling approaches were developed to improve the accuracy of the model predictions in both heterotrophic and phototrophic conditions. It is hoped that the information in the Arabidopsis model, the modelling approaches developed and the model predictions made in this study will be translated into crop plants.

8 Literature cited

Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I. and Nielsen, J. (2012) Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Computational Biology*, **8**, e1002518.

Agren, R., Shoaie, S., Vongsangnak, W., Nookaew, I. and Nielsen, J. (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Computational Biology*, **9**, e1002980.

Alder, N. N. and Theg, S. M. (2003) Energy use by biological protein transport pathways. *Trends in Biochemical Sciences*, **28**, 442–451.

Allen, J. F. (2003) Cyclic, pseudocyclic and noncyclic photophosphorylation: new links in the chain. *Trends in Plant Science*, **8**, 15–9.

Allen, D. K., Ohlrogge, J. B. and Shachar-Hill, Y. (2009) The role of light in soybean seed filling metabolism. *The Plant Journal*, **58**, 220–34.

Alonso, A. P., Raymond, P., Rolin, D. and Dieuaide-Noubhani, M. (2007) Substrate cycles in the central metabolism of maize root tips under hypoxia. *Phytochemistry*, **68**, 2222–2231.

Amthor, J. S. (2000) The McCree-de Wit-Penning de Vries-Thornley respiration paradigms: 30 years later. *Annals of Botany*, **86**, 1–20.

Antony, E., Taybi, T., Courbot, M., Mugford, S. T., Smith, J. A. C. and Borland, A. M. (2008) Cloning, localization and expression analysis of vacuolar sugar transporters in the CAM plant *Ananas comosus* (pineapple). *Journal of Experimental Botany*, **59**, 1895–908.

Antony, E. and Borland, A. M. (2009) The role and regulation of sugar transporters in plants with Crassulacean Acid Metabolism in CAM Plants. *Progress in Botany*, **70**, 127–143.

Arkblad, E. L., Betsholtz, C., Mandoli, D. and Rydström, J. (2001) Characterization of a nicotinamide nucleotide transhydrogenase gene from the green alga *Acetabularia acetabulum* and comparison of its structure with those of the corresponding genes in mouse and *Caenorhabditis elegans*. *Biochimica et Biophysica Acta*, **1520**, 115–123.

- Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Rouspard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F. I. and Papale, D.** (2010) Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science*, **329**, 834–8.
- Beevers, L., Flesher, D. and Hageman, R. H.** (1964) Studies on the pyridine nucleotide specificity of nitrate reductase in higher plants and its relationship to sulfhydryl level. *Biochimica et Biophysica Acta*, **89**, 453–64.
- Bhullar, N. K. and Gruissem, W.** (2013) Nutritional enhancement of rice for human health: the contribution of biotechnology. *Biotechnology Advances*, **31**, 50–7.
- Bordbar, A., Lewis, N. E., Schellenberger, J., Palsson, B. Ø. and Jamshidi, N.** (2010) Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Molecular Systems Biology*, **6**, 422.
- Bordbar, A., Feist, A. M., Usaite-Black, R., Woodcock, J., Palsson, B. O. and Famili, I.** (2011) A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC Systems Biology*, **5**, 180.
- Borisjuk, L., Neuberger, T., Schwender, J., Heinzl, N., Sunderhaus, S., Fuchs, J., Hay, J. O., Tschiersch, H., Braun, H. P., Denolf, P., Lambert, B., Jakob, P. M. and Rolletschek, H.** (2013) Seed architecture shapes embryo metabolism in oilseed rape. *The Plant Cell*, **25**, 1625–1640.
- Borland, A. M. and Griffiths, H.** (1989) The regulation of citric acid accumulation and carbon recycling during CAM in *Ananas comosus*. *Journal of Experimental Botany*, **40**, 53–60.
- Boyle, N. R. and Morgan, J. A.** (2009) Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC Systems Biology*, **3**, 4.
- Bush, D. R.** (1993) Proton-coupled sugar and amino acid transporters in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, **44**, 513–542.
- Burgard, A. P. and Maranas, C. D.** (2003) Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnology and Bioengineering*, **82**, 670–7.

- Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S. Y., Shearer, A. G., Tissier, C., Walk, T. C., Zhang, P. and Karp, P. D.** (2008) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, **36**, D623–31.
- Chang, R. L. Ghamsari, L., Manichaikul, A., Hom, E. F. Y., Balaji, S., Fu, W., Shen, Y., Hao, T., Palsson, B. Ø., Salehi-Ashtiani, K. and Papin, J. A.** (2011) Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Molecular Systems Biology*, **7**, 518.
- Chen, T., Xie, Z. and Ouyang, Q.** (2009) Expanded flux variability analysis on metabolic network of *Escherichia coli*. *Chinese Science Bulletin*, **54**, 2610–2619.
- Chen, X. and Shachar-Hill, Y.** (2012) Insights into metabolic efficiency from flux analysis. *Journal of Experimental Botany*, **63**, 2343–51.
- Chung, B. K.-S., Lakshmanan, M., Klement, M., Mohanty, B. and Lee, D.-Y.** (2012) Genome-scale in silico modeling and analysis for designing synthetic terpenoid-producing microbial cell factories. *Chemical Engineering Science*, doi:10.1016/j.ces.2012.09.006
- Covert, M., Schilling, C., Famili, I. and Edwards, J.** (2001) Metabolic modeling of microbial strains *in silico*. *Trends in Biochemical Sciences*, **26**, 179–186.
- Cushman, J.** (2001) Crassulacean acid metabolism. A plastic photosynthetic adaptation to arid environments. *Plant Physiology*, **127**, 1439–1448.
- Cushman, J. C., Tillett, R. L., Wood, J. A., Branco, J. M. and Schlauch, K. A.** (2008) Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum*, performing C3 photosynthesis and Crassulacean acid metabolism (CAM). *Journal of Experimental Botany*, **59**, 1875–94.
- Dailey, F. A., Kuo, T. and Warner, R. L.** (1982) Pyridine nucleotide specificity of barley nitrate reductase. *Plant Physiology*, **69**, 1196–9.
- de Oliveira Dal’Molin, C. G., Quek, L.-E., Palfreyman, R. W., Brumbley, S. M. and Nielsen, L. K.** (2010a) AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant Physiology*, **152**, 579–89.
- de Oliveira Dal’Molin, C. G., Quek, L.-E., Palfreyman, R. W., Brumbley, S. M. and Nielsen, L. K.** (2010b) C4GEM, a genome-scale metabolic model to study C4 plant metabolism. *Plant Physiology*, **154**, 1871–85.

- 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.
- de Oliveira Dal'Molin, C. G. and Nielsen, L. K.** (2013) Plant genome-scale metabolic reconstruction and modelling. *Current Opinion in Biotechnology*, **24**, 271–7.
- Delhon, P., Gojon, A. and Tillard, P.** (1995) Diurnal regulation of NO₃– uptake in soybean plants I. Changes in NO₃– influx, efflux, and N utilization in the plant during the day/night cycle. *Journal of Experimental Botany*, **46**, 1585–1594.
- 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.
- Drennan, P. M. and Nobel, P. S.** (2000) Responses of CAM species to increasing atmospheric CO₂ concentrations. *Plant, Cell and Environment*, **23**, 767–781.
- Dusick, J. R., Glenn, T. C., Lee, W. N. P., Vespa, P. M., Kelly, D. F., Lee, S. M., Hovda, D. A. and Martin, N. A.** (2007) Increased pentose phosphate pathway flux after clinical traumatic brain injury: a [1,2-¹³C₂]glucose labeling study in humans. *Journal of Cerebral Blood Flow & Metabolism*, **27**, 1593–1602.
- Eckardt, N. A., Snyder, G. W., Portis, A. R. and Orger, W. L.** (1997) Growth and photosynthesis under high and low irradiance of *Arabidopsis thaliana* antisense mutants with reduced ribulose-1,5-bisphosphate carboxylase/oxygenase activase content. *Plant Physiology*, **113**, 575–86.
- Edwards, J. S., Ramakrishna, R. and Palsson, B. O.** (2002) Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnology and Bioengineering*, **77**, 27–36.
- Farmer, I. S. and Jones, C. W.** (1976) The energetics of *E. coli* during aerobic growth in continuous culture. *European Journal of Biochemistry*, **67**, 115–122.
- Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V. and Palsson, B. Ø.** (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, **3**, 121.

- Feist, A. M. and Palsson, B. Ø.** (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nature Biotechnology*, **26**, 659–67.
- Feist, A. M. and Palsson, B. O.** (2010) The biomass objective function. *Current Opinion in Microbiology*, **13**, 344–349.
- Fell, D. A. and Small, J. R.** (1986) Fat synthesis in adipose tissue. *Biochemical Journal*, **238**, 781–786.
- Fell, D. A.** (1997) *Understanding the Control of Metabolism*. Portland Press, London.
- Feller, U., Anders, I. and Mae, T.** (2008) Rubiscolytics: fate of Rubisco after its enzymatic function in a cell is terminated. *Journal of Experimental Botany*, **59**, 1615–24.
- Garlick, A. P., Moore, C. and Kruger, N. J.** (2002) Monitoring flux through the oxidative pentose phosphate pathway using [1-14C]gluconate. *Planta*, **216**, 265–272.
- Gaxiola, R. A., Palmgren, M. G. and Schumacher, K.** (2007) Plant proton pumps. *FEBS Letters*, **581**, 2204–2214.
- Gevorgyan, A., Poolman, M. G. and Fell, D. A.** (2008) Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics*, **24**, 2245–51.
- Gianchandani, E. P., Oberhardt, M. A., Burgard, A. P., Maranas, C. D. and Papin, J. A.** (2008) Predicting biological system objectives de novo from internal state measurements. *BMC Bioinformatics*, **9**, 43.
- Gibon, Y., Bläsing, O. E., Palacios-Rojas, N., Pankovic, D., Hendriks, J. H. M., Fisahn, J., Höhne, M., Günther, M. and Stitt, M.** (2004) Adjustment of diurnal starch turnover to short days: depletion of sugar during the night leads to a temporary inhibition of carbohydrate utilization, accumulation of sugars and post-translational activation of ADP-glucose pyrophosphorylase in the following light period. *The Plant Journal*, **39**, 847–62.
- Grafahrend-Belau, E., Schreiber, F., Koschützki, D. and Junker, B. H.** (2009) Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism. *Plant Physiology*, **149**, 585–98.
- Grafahrend-Belau, E., Junker, A., Eschenröder, A., Müller, J., Schreiber, F. and Junker B. H.** (2013) Multiscale metabolic modeling: dynamic flux balance analysis on a whole plant scale. *Plant Physiology*, **163**, 637–647.

- Hay, J. and Schwender, J.** (2011a) Computational analysis of storage synthesis in developing *Brassica napus* L. (oilseed rape) embryos: flux variability analysis in relation to ^{13}C metabolic flux analysis. *The Plant Journal*, **67**, 513–25.
- Hay, J. and Schwender, J.** (2011b) Metabolic network reconstruction and flux variability analysis of storage synthesis in developing oilseed rape (*Brassica napus* L.) embryos. *The Plant Journal*, **67**, 526–41.
- Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I. and Millar, A. H.** (2007) SUBA: the Arabidopsis Subcellular Database. *Nucleic Acids Research*, **35**, D213–8.
- Heber, U.** (2002) Irrungen, Wirrungen? The Mehler reaction in relation to cyclic electron transport in C3 plants. *Photosynthesis Research*, **73**, 223–31.
- Heineke, D., Riens, B., Grosse, H., Hoferichter, P., Peter, U., Flügge, U. I. and Heldt, H. W.** (1991) Redox transfer across the inner chloroplast envelope membrane. *Plant Physiology*, **95**, 1131–1137.
- Hibberd, J. M., Sheehy, J. E. and Langdale, J. A.** (2008) Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. *Current Opinion in Plant Biology*, **11**, 228–31.
- Holtum, J. A. M., Smith, J. A. C. and Neuhaus, H. E.** (2005) Intracellular transport and pathways of carbon flow in plants with crassulacean acid metabolism. *Functional Plant Biology*, **32**, 429–449.
- Holzhütter, H.-G.** (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European Journal of Biochemistry*, **271**, 2905–22.
- Huber, J. L., Huber, S. C., Campbell, W. H. and Redinbaugh, M. G.** (1992) Reversible light/dark modulation of spinach leaf nitrate reductase activity involves protein phosphorylation. *Archives of Biochemistry and Biophysics*, **296**, 58–65.
- Igamberdiev, A. and Gardeström, P.** (2003) Regulation of NAD- and NADP-dependent isocitrate dehydrogenases by reduction levels of pyridine nucleotides in mitochondria and cytosol of pea leaves. *Biochimica et Biophysica Acta*, **1606**, 117–125.
- Jerby, L., Shlomi, T. and Ruppin, E.** (2010) Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Molecular Systems Biology*, **6**, 401.
- Johnson, G. N.** (2005) Cyclic electron transport in C3 plants: fact or artefact? *Journal of Experimental Botany*, **56**, 407–16.

- Josse, E. M., Simkin, A. J., Gaffé, J., Labouré, A. M., Kuntz, M. and Carol, P.** (2000) A plastid terminal oxidase associated with carotenoid desaturation during chromoplast differentiation. *Plant Physiology*, **123**, 1427–36.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M.** (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, **40**, D109–14.
- Kartal, Ö., Mahlow, S., Skupin, A. and Ebenhöh, O.** (2011) Carbohydrate-active enzymes exemplify entropic principles in metabolism. *Molecular Systems Biology*, **7**, 542.
- Kauffman, K. J., Prakash, P. and Edwards, J. S.** (2003) Advances in flux balance analysis. *Current Opinion in Biotechnology*, **14**, 491–496.
- Kebeish, R., Niessen, M., Thiruveedhi, K., Bari, R., Hirsch, H.-J., Rosenkranz, R., Stäbler, N., Schönfeld, B., Kreuzaler, F. and Peterhänsel, C.** (2007) Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. *Nature Biotechnology*, **25**, 593–9.
- Kim, T. Y., Sohn, S. B., Kim, Y. Bin, Kim, W. J. and Lee, S. Y.** (2012) Recent advances in reconstruction and applications of genome-scale metabolic models. *Current Opinion in Biotechnology*, **23**, 617–23.
- Klitgord, N. and Segrè, D.** (2010) The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles. *Genome Informatics. International Conference on Genome Informatics*, **22**, 41–55.
- Knoop, H., Zilliges, Y., Lockau, W. and Steuer, R.** (2010) The metabolic network of *Synechocystis* sp. PCC 6803: systemic properties of autotrophic growth. *Plant Physiology*, **154**, 410–422.
- Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W. and Steuer, R.** (2013) Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis* sp. PCC 6803. *PLoS Computational Biology*, **9**, e1003081.
- Kore-eda, S. and Kanai, R.** (1997) Induction of glucose 6-phosphate transport activity in chloroplasts of *Mesembryanthemum crystallinum* by the C3-CAM transition. *Plant and Cell Physiology*, **38**, 895–901.
- Kramer, D. M. and Evans, J. R.** (2011) The importance of energy balance in improving photosynthetic productivity. *Plant Physiology*, **155**, 70–8.
- Kruger, N. J. and von Schaewen, A.** (2003) The oxidative pentose phosphate pathway: structure and organisation. *Current Opinion in Plant Biology*, **6**, 236–246.

- Kruger, N. J., LeLay, P. and Ratcliffe, R. G.** (2007) Vacuolar compartmentation complicates the steady-state analysis of glucose metabolism and forces reappraisal of sucrose cycling in plants. *Phytochemistry*, **68**, 2189-2196.
- Kruger, N. J. and Ratcliffe, R. G.** (2009) Insights into plant metabolic networks from steady-state metabolic flux analysis. *Biochimie*, **91**, 697–702.
- Lakshmanan, M., Zhang, Z., Mohanty, B., Kwon, J.-Y., Choi, H.-Y., Nam, H.-J., Kim, D.-I. and Lee, D.-Y.** (2013) Elucidating rice cell metabolism under flooding and drought stresses using flux-based modeling and analysis. *Plant Physiology*, **162**, 2140–50.
- Lewis, N. E., Schramm, G., Bordbar, A., Schellenberger, J., Andersen, M. P., Cheng, J. K., Patel, N., Yee, A., Lewis, R. A., Eils, R., König, R. and Palsson, B. Ø.** (2010) Large-scale *in silico* modeling of metabolic interactions between cell types in the human brain. *Nature Biotechnology*, **28**, 1279–85.
- Liao, Y.-C., Tsai, M.-H., Chen, F.-C. and Hsiung, C. A.** (2012) GEMSiRV: a software platform for genome-scale metabolic model simulation, reconstruction and visualization. *Bioinformatics*, **28**, 1752–8.
- Lim, E.-K. and Bowles, D.** (2012) Plant production systems for bioactive small molecules. *Current Opinion in Biotechnology*, **23**, 271–7.
- Linka, N. and Weber, A. P. M.** (2005) Shuffling ammonia between mitochondria and plastids during photorespiration. *Trends in Plant Science*, **10**, 461–5.
- Linka, N. and Weber, A. P. M.** (2010) Intracellular metabolite transporters in plants. *Molecular Plant*, **3**, 21–53.
- Lüttge, U., Smith, J. A. C., Marigo, G. and Osmond, C. B.** (1981) Energetics of malate accumulation in the vacuoles of *Kalanchoë tubiflora* cells. *FEBS Letters*, **126**, 81–84.
- Lüttge, U.** (1988) Day-night changes of citric-acid levels in crassulacean acid metabolism: phenomenon and ecophysiological significance. *Plant, Cell and Environment*, **11**, 445–451.
- Lüttge, U.** (2004) Ecophysiology of Crassulacean Acid Metabolism (CAM). *Annals of Botany*, **93**, 629–52.
- Lüttge, U.** (2011) Photorespiration in Phase III of Crassulacean Acid Metabolism: evolutionary and ecophysiological implications. *Progress in Botany*, **72**, 371-384.

- Macduff, J. H. and Bakken, A. K.** (2003) Diurnal variation in uptake and xylem contents of inorganic and assimilated N under continuous and interrupted N supply to *Phleum pratense* and *Festuca pratensis*. *Experimental Biology*, **54**, 431–444.
- Maeshima, M.** (2000) Vacuolar H⁺-pyrophosphatase. *Biochimica et Biophysica Acta*, **1465**, 37–51.
- Mahadevan, R. and Schilling, C. H.** (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, **5**, 264–276.
- Malone, J. G., Mittova, V., Ratcliffe, R. G. and Kruger, N. J.** (2006) The response of carbohydrate metabolism in potato tubers to low temperature. *Plant and Cell Physiology*, **47**, 1309–1322.
- Masakapalli, S. K., Le Lay, P., Huddleston, J. E., Pollock, N. L., Kruger, N. J. and Ratcliffe, R. G.** (2010) Subcellular flux analysis of central metabolism in a heterotrophic Arabidopsis cell suspension using steady-state stable isotope labeling. *Plant Physiology*, **152**, 602–619.
- Masakapalli, S. K., Kruger, N. J. and Ratcliffe, R. G.** (2013) The metabolic flux phenotype of heterotrophic Arabidopsis cells reveals a complex response to changes in nitrogen supply. *The Plant Journal*, **74**, 569–82.
- Maurino, V. G. and Weber, A. P. M.** (2013) Engineering photosynthesis in plants and synthetic microorganisms. *Journal of Experimental Botany*, **64**, 743–51.
- McRae S. R., Christopher J. T., Holton J. A. M. and Smith J. A. C.** (2002) Sucrose transport across the tonoplast of *Ananas comosus* L. *Functional Plant Biology*, **29**, 717–724.
- Mintz-Oron, S., Aharoni, A., Ruppin, E. and Shlomi, T.** (2009) Network-based prediction of metabolic enzymes' subcellular localization. *Bioinformatics*, **25**, i247–52.
- Mintz-Oron, S., Meir, S., Malitsky, S., Ruppin, E., Aharoni, A. and Shlomi, T.** (2012) Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 339–44.
- Miyake, C.** (2010) Alternative electron flows (water-water cycle and cyclic electron flow around PSI) in photosynthesis: molecular mechanisms and physiological functions. *Plant & Cell Physiology*, **51**, 1951–63.

- Montagud, A., Navarro, E., Fernández de Córdoba, P., Urchueguía, J. F. and Patil, K. R.** (2010) Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. *BMC Systems Biology*, **4**, 156.
- Mueller, L. A., Zhang, P. and Rhee, S. Y.** (2003) AraCyc: a biochemical pathway database for Arabidopsis. *Plant Physiology*, **132**, 453.
- Müller, V. and Grüber, G.** (2003) ATP synthases: structure, function and evolution of unique energy converters. *Cellular and Molecular Life Sciences*, **60**, 474-494.
- Mullineaux, C. W. and Emlyn-Jones, D.** (2005) State transitions: an example of acclimation to low-light stress. *Journal of Experimental Botany*, **56**, 389-393.
- Munekage, Y. N., Genty, B. and Peltier, G.** (2008) Effect of PGR5 impairment on photosynthesis and growth in *Arabidopsis thaliana*. *Plant & Cell Physiology*, **49**, 1688–98.
- Murchie, E. H., Pinto, M. and Horton, P.** (2009) Agriculture and the new challenges for photosynthesis research. *The New Phytologist*, **181**, 532–52.
- Näsholm, T., Kielland, K. and Ganeteg, U.** (2009) Uptake of organic nitrogen by plants. *The New Phytologist*, **182**, 31–48.
- Neff, J., Chapin III, F. and Vitousek, P.** (2003) Breaks in the cycle: dissolved organic nitrogen in terrestrial ecosystems. *Frontiers in Ecology and the Environment*, **1**, 205–211.
- Neuhaus, H. E. and Schulte, N.** (1996) Starch degradation in chloroplasts isolated from C3 or CAM (crassulacean acid metabolism)-induced *Mesembryanthemum crystallinum* L. *The Biochemical Journal*, **318**, 945–53.
- Nguyen-Quoc, B., Krivitzky, M., Huber, S. C. and Lechamy, A.** (1990) Sucrose synthase in developing maize leaves: regulation of activity by protein level during the import to export transition. *Plant Physiology*, **94**, 516–23.
- Niewiadomski, P., Knappe, S. and Geimer, S.** (2005) The Arabidopsis plastidic glucose 6-phosphate/phosphate translocator GPT1 is essential for pollen maturation and embryo sac development. *The Plant Cell*, **17**, 760–775.
- Niklas, J., Schneider, K. and Heinzle, E.** (2010) Metabolic flux analysis in eukaryotes. *Current Opinion in Biotechnology*, **21**, 63–9.

- Noctor, G. and Foyer, C. H.** (1998) A re-evaluation of the ATP: NADPH budget during C₃ photosynthesis: a contribution from nitrate assimilation and its associated respiratory activity? *Journal of Experimental Botany*, **49**, 1895-1908.
- Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O. and Thiele, I.** (2012) Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 2678–83.
- Oberhardt, M. A., Palsson, B. Ø. and Papin, J. A.** (2009) Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology*, **5**, 320.
- Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M. and Palsson, B. Ø.** (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular Systems Biology*, **7**, 535.
- Osmond, C. B.** (1978) Crassulacean acid metabolism: a curiosity in context. *Annual Review of Plant Physiology*, **29**, 379–414.
- Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O. and Vonstein, V.** (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, **33**, 5691–702.
- Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., Wright, S. Y., Hinchliffe, E., Adams, J. L., Silverstone, A. L. and Drake, R.** (2005) Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nature Biotechnology*, **23**, 482–7.
- Papoutsakis, E. T.** (1984) Equations and calculations for fermentations of butyric acid bacteria. *Biotechnology and Bioengineering*, **26**, 174–187.
- Pedersen, A., Karlsson, G. B. and Rydström, J.** (2008) Proton-translocating transhydrogenase: an update of unsolved and controversial issues. *Journal of Bioenergetics and Biomembranes*, **40**, 463-473.
- Penning de Vries, F. W. T.** (1975) The cost of maintenance processes in plant cells. *Annals of Botany*, **39**, 77-92.

- Piques, M., Schulze, W. X., Höhne, M., Usadel, B., Gibon, Y., Rohwer, J. and Stitt, M.** (2009) Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in *Arabidopsis*. *Molecular Systems Biology*, **5**, 314.
- Poolman, M. G., Fell, D. A. and Raines, C. A.** (2003) Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. *European Journal of Biochemistry*, **270**, 430–439.
- Poolman, M.** (2006) ScrumPy: metabolic modelling with Python. *IEE Proceedings-Systems Biology*, **153**, 375–378.
- Poolman, M., Bonde, B., Gevorgyan, A., Patel, H. and Fell, D.** (2006) Challenges to be faced in the reconstruction of metabolic networks from public databases. *IEE Proceedings-Systems Biology*, **153**, 379–384.
- Poolman, M. G., Miquet, L., Sweetlove, L. J. and Fell, D. A.** (2009) A genome-scale metabolic model of *Arabidopsis thaliana* and some of its properties. *Plant Physiology*, **151**, 1570–1581.
- Poolman, M. G., Kundu, S., Shaw, R. and Fell, D.** (2013) Responses to light intensity in a genome-scale model of rice metabolism. *Plant Physiology*, **162**, 1060–1072.
- Price, N. D., Schellenberger, J. and Palsson, B. O.** (2004) Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. *Biophysical Journal*, **87**, 2172–86.
- Quek, L.-E., Wittmann, C., Nielsen, L. K. and Krömer, J. O.** (2009) OpenFLUX: efficient modelling software for ¹³C-based metabolic flux analysis. *Microbial Cell Factories*, **8**, 25.
- 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 Systems Biology*, **4**, 114.
- Raghavendra, A. S. and Padmasree, K.** (2003) Beneficial interactions of mitochondrial metabolism with photosynthetic carbon assimilation. *Trends in Plant Science*, **8**, 546–553.
- Raines, C. A.** (2011) Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: current and future strategies. *Plant Physiology*, **155**, 36–42.
- Ratcliffe, R.G. and Shachar-Hill, Y.** (2006) Measuring multiple fluxes through plant metabolic networks. *The Plant Journal*, **45**, 490–511.

- Riens, B. and Heldt, H. W.** (1992) Decrease of nitrate reductase activity in spinach leaves during a light-dark transition. *Plant Physiology*, **98**, 573–7.
- Rios-Esteva, R. and Lange, B. M.** (2007) Experimental and mathematical approaches to modeling plant metabolic networks. *Phytochemistry*, **68**, 2351–74.
- Rohwer, J. M.** (2012) Kinetic modelling of plant metabolic pathways. *Journal of Experimental Botany*, **63**, 2275–92.
- Rontein, D., Dieuaide-Noubhani, M., Dufourc, E. J., Raymond, P. and Rolin D.** (2002) The metabolic architecture of plant cells. Stability of central metabolism and flexibility of anabolic pathways during the growth cycle of tomato cells. *The Journal of Biological Chemistry*, **277**, 43948–43960.
- Roscher, A., Kruger, N. J. and Ratcliffe, R. G.** (2000) Strategies for metabolic flux analysis in plants using isotope labelling. *Journal of Biotechnology*, **77**, 81–102.
- Ruter, J. M.** (1993) High-temperature-induced electrolyte leakage from excised leaves and roots of 3 hollies. *Hortscience*, **28**, 927–928.
- 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**, e21784.
- Saha, R., Versepunt, A. T., Berla, B. M., Mueller, T. J., Pakrasi, H. B., Maranas, C. D.** (2012) Reconstruction and comparison of the metabolic potential of cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803. *PLoS One*, **7**, e48285.
- Schallau, K. and Junker, B. H.** (2010) Simulating plant metabolic pathways with enzyme-kinetic models. *Plant Physiology*, **152**, 1763–71.
- Scheibe, R.** (2004) Malate valves to balance cellular energy supply. *Physiologia Plantarum*, **120**, 21–26.
- Scheible, W.-R., Krapp, A. and Stitt, M.** (2000) Reciprocal diurnal changes of phosphoenolpyruvate carboxylase expression and cytosolic pyruvate kinase, citrate synthase and NADP-isocitrate dehydrogenase expression regulate organic acid metabolism during nitrate assimilation in tobacco leaves. *Plant, Cell and Environment*, **23**, 1155–1167.
- 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. Ø.** (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, **6**, 1290–1307.

- Scheurwater, I., Dunnebacke, M., Eising, R. and Lambers H.** (2000) Respiratory costs and rate of protein turnover in the roots of a fast-growing (*Dactylis glomerata* L.) and a slow-growing (*Festuca ovina* L.) grass species. *Journal of Experimental Botany*, **51**, 1089-1097.
- Schuetz, R., Kuepfer, L. and Sauer, U.** (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, **3**, 119.
- Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M. and Sauer, U.** (2012) Multidimensional optimality of microbial metabolism. *Science*, **336**, 601–604
- Schuster, S., Fell, D. A. and Dandekar, T.** (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, **18**, 326–32.
- Schuster, S., Fell, D. A. and Pfeiffer, T.** (2008) Is maximization of molar yield in metabolic networks favoured by evolution? *Journal of Theoretical Biology*, **252**, 497–504.
- Schwarzländer, M., Logan, D. C., Johnston, I. G., Jones, N. S., Meyer, A. J., Fricker, M. D. and Sweetlove, L. J.** (2012) Pulsing of membrane potential in individual mitochondria: a stress-induced mechanism to regulate respiratory bioenergetics in Arabidopsis. *Plant Cell*, **24**, 1188-1201.
- Schwender, J., Ohlrogge, J. B. and Shachar-Hill, Y.** (2003) A flux model of glycolysis and the oxidative pentosephosphate pathway in developing *Brassica napus* embryos. *The Journal of Biological Chemistry*, **278**, 29442-29453.
- 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**, 779–82.
- Schwender, J.** (2008) Metabolic flux analysis as a tool in metabolic engineering of plants. *Current Opinion in Biotechnology*, **19**, 131–7.
- Schwender, J. and Hay, J. O.** (2012) Predictive modeling of biomass component tradeoffs in *Brassica napus* developing oilseeds based on *in silico* manipulation of storage metabolism. *Plant Physiology*, **160**, 1218–36.
- Seaver, S. M., Henry, C. S. and Hanson, A. D.** (2012) Frontiers in metabolic reconstruction and modeling of plant genomes. *Journal of Experimental Botany*, **63**, 2247-2258.

- Sheikh, K., Förster, J. and Nielsen, L. K.** (2005) Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of *Mus musculus*. *Biotechnology Progress*, **21**, 112–21.
- Siebrecht, S., Herdel, K., Schurr, U. and Tischner, R.** (2003) Nutrient translocation in the xylem of poplar--diurnal variations and spatial distribution along the shoot axis. *Planta*, **217**, 783–93.
- Silvera, K., Neubig, K. M., Whitten, W. M., Williams, N. H., Winter, K. and Cushman, J. C.** (2010) Evolution along the crassulacean acid metabolism continuum. *Functional Plant Biology*, **37**, 995-1010.
- Smallbone, K. and Simeonidis, E.** (2009) Flux balance analysis: a geometric perspective. *Journal of Theoretical Biology*, **258**, 311–5.
- Smith, J.A.C. and Bryce, J.H.** (1992) Metabolite compartmentation and transport in CAM plants. In Tobin, A.K. (Eds.), *Plant Organelles* (pp. 141-167). Cambridge: Cambridge University Press.
- Smith, N. G. and Dukes, J. S.** (2013) Plant respiration and photosynthesis in global-scale models: incorporating acclimation to temperature and CO₂. *Global Change Biology*, **19**, 45–63.
- Stolyar, S., Van Dien, S., Hillesland, K. L., Pinel, N., Lie, T. J., Leigh, J. A., Stahl, D. A.** (2007) Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology*, **3**, 92.
- Suzuki, A. and Knaff, D. B.** (2005) Glutamate synthase: structural, mechanistic and regulatory properties, and role in the amino acid metabolism. *Photosynthesis Research*, **83**, 191–217.
- Sweetlove, L. J., Lytovchenko, A., Morgan, M., Nunes-Nesi, A., Taylor, N. L., Baxter, C. J., Eickmeier, I. and Fernie, A. R.** (2006) Mitochondrial uncoupling protein is required for efficient photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 19587–92.
- Sweetlove, L. J. and Ratcliffe, R. G.** (2011) Flux-balance modeling of plant metabolism. *Frontiers in Plant Science*, **2**, 38.
- Sweetlove, L. J. and Fernie, A. R.** (2013) The spatial organization of metabolism within the plant cell. *Annual Review of Plant Biology*, **64**, 723–46.
- Sweetlove, L. J., Williams, T. C. R., Cheung, C. Y. M. and Ratcliffe, R. G.** (2013) Modelling metabolic CO₂ evolution – a fresh perspective on respiration. *Plant Cell and Environment*, **36**, 1631-1640.

Szecowka, M., Heise, R., Tohge, T., Nunes-Nesi, A., Vosloh, D., Huege, J., Feil, R., Lunn, J., Nikoloski, Z., Stitt, M., Fernie, A. R. and Arrivault, S. (2013) Metabolic fluxes in an illuminated Arabidopsis rosette. *The Plant Cell*, **25**, 694–714.

Taira, M., Valtersson, U., Burkhardt, B. and Ludwig, R. A. (2004) *Arabidopsis thaliana* GLN2-encoded glutamine synthetase is dual targeted to leaf mitochondria and chloroplasts. *The Plant Cell*, **16**, 2048–58.

Talbott, L. D. and Zeiger, E. (1998) The role of sucrose in guard cell osmoregulation. *Journal of Experimental Botany*, **49**, 329–337.

Tanaka, R. and Tanaka, A. (2011) Chlorophyll cycle regulates the construction and destruction of the light-harvesting complexes. *Biochimica et Biophysica Acta*, **1807**, 968–76.

Tanz, S. K., Castleden, I., Hooper, C. M., Vacher, M., Small, I. and Millar, H. A. (2013) SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in Arabidopsis. *Nucleic Acids Research*, **41**, D1185–91.

Tcherkez, G., Mahé, A., Gauthier, P., Mauve, C., Gout, E. and Bligny, R. (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 Physiology*, **151**, 620–30.

Theodoulou, F. L. and Eastmond, P. J. (2012) Seed storage oil catabolism: a story of give and take. *Current Opinion in Plant Biology*, **15**, 322–8.

Thiele, I. and Palsson, B. O. (2010) A protocol for generating a high-quality genome-scale reconstruction. *Nature Protocols*, **5**, 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. *The Plant Cell*, **25**, 1197–1211

Tovar-Mendez, A., Miernyk, J. A. and Randall, D. D. (2003) Regulation of pyruvate dehydrogenase complex activity in plant cells. *European Journal of Biochemistry*, **270**, 1043–1049.

Varma, A. and Palsson, B. O. (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology*, **60**, 3724–3731.

- Vo, T. D., Greenberg, H. J. and Palsson, B. O.** (2004) Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *The Journal of Biological Chemistry*, **279**, 39532-39540.
- von Ballmoos, C., Wiedenmann, A. and Dimroth, P.** (2009) Essentials for ATP synthesis by F1F0 ATP synthases. *Annual Review of Biochemistry*, **78**, 649-672.
- von Caemmerer, S., Quick, W. P. and Furbank, R. T.** (2012) The development of C4 rice: current progress and future challenges. *Science*, **336**, 1671-2.
- Wang, C., Guo, L., Li, Y. and Wang, Z.** (2012) Systematic comparison of C3 and C4 plants based on metabolic network analysis. *BMC Systems Biology*, **6 Suppl 2**, S9.
- Wang, Y., Eddy, J. A and Price, N. D.** (2012) Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Systems Biology* **6**, 153.
- Weise, S. E., van Wijk, K. J. and Sharkey, T. D.** (2011) The role of transitory starch in C3, CAM, and C4 metabolism and opportunities for engineering leaf starch accumulation. *Journal of Experimental Botany*, **62**, 3109-3118.
- Weitzel, M., Nöh, K., Dalman, T., Niedenführ, S., Stute, B. and Wiechert, W.** (2013) 13CFLUX2--high-performance software suite for (13)C-metabolic flux analysis. *Bioinformatics*, **29**, 143-5.
- Wells, G. N. and Hageman, R. H.** (1974) Specificity for nicotinamide adenine dinucleotide by nitrate reductase from leaves. *Plant Physiology*, **54**, 136-41.
- Wiechert, W., Möllney, M., Petersen, S. and de Graaf, A. A.** (2001) A universal framework for 13C metabolic flux analysis. *Metabolic engineering*, **3**, 265-83.
- Wilkinson, T. L. and Douglas, A. E.** (2003) Phloem amino acids and the host plant range of the polyphagous aphid, *Aphis fabae*. *Entomologia Experimentalis et Applicata*, **106**, 103-113.
- Williams, T. C. R., Miguët, L., Masakapalli, S. K., Kruger, N. J., Sweetlove, L. J. and Ratcliffe, R. G.** (2008) Metabolic network fluxes in heterotrophic Arabidopsis cells: stability of the flux distribution under different oxygenation conditions. *Plant Physiology*, **148**, 704-18.

Williams, T. C. R., Poolman, M. G., Howden, A. J. M., Schwarzlander, M., Fell, D. A., Ratcliffe, R. G. and Sweetlove, L. J. (2010) A genome-scale metabolic model accurately predicts fluxes in central carbon metabolism under stress conditions. *Plant Physiology*, **154**, 311–323.

Winter, K. and Smith, J. A. C. (1996) Crassulacean acid metabolism: current status and perspectives. In Winter, K. and Smith J. A. C. (Eds), *Crassulacean acid metabolism. Biochemistry, ecophysiology and evolution* (pp. 389-426). Berlin: Springer-Verlag.

Wu, S. and Chappell, J. (2008) Metabolic engineering of natural products in plants; tools of the trade and challenges for the future. *Current Opinion in Biotechnology*, **19**, 145–52.

Yamamoto, H., Peng, L., Fukao, Y. and Shikanai, T. (2011) An Src homology 3 domain-like fold protein forms a ferredoxin binding site for the chloroplast NADH dehydrogenase-like complex in Arabidopsis. *The Plant Cell*, **23**, 1480–93.

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P. and Potrykus, I. (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science*, **287**, 303–5.

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**, i255–60.

Zamboni, N., Fischer, E. and Sauer, U. (2005) FiatFlux--a software for metabolic flux analysis from ¹³C-glucose experiments. *BMC Bioinformatics*, **6**, 209.

Zhao, Z., Kuijvenhoven, K., Ras, C., van Gulik, W.M., Heijnen, J. J., Verheijen, P. J. T. and van Winden, W. A. (2008) Isotopic non-stationary ¹³C gluconate tracer method for accurate determination of the pentose phosphate pathway split-ratio in *Penicillium chrysogenum*. *Metabolic Engineering*, **10**, 178-186.

Zhu, X.-G., Shan, L., Wang, Y. and Quick, W. P. (2010) C4 rice - an ideal arena for systems biology research. *Journal of Integrative Plant Biology*, **52**, 762–70.

Zhuang, K., Izallalen, M., Mouser, P., Richter, H., Risso, C., Mahadevan, R. and Lovley, D. R. (2011) Genome-scale dynamic modeling of the competition between *Rhodospirillum rubrum* and *Geobacter* in anoxic subsurface environments. *The ISME Journal*, **5**, 305–16.

9 Appendix

List of appendices: (Appendices S1-S12 are available on DVD only)

Appendix 1	Description of the ScrumPy commands for setting model constraints and objective functions specific to this study
Appendix 2	Reaction mapping between MFA model and genome-scale metabolic model
Appendix 3	Summary of energy-use stoichiometries for transporters and protein translocation
Appendix S1	ScrumPy codes for model construction and analysis
Appendix S2	Arabidopsis genome-scale metabolic model in ScrumPy and SBML formats
Appendix S3	Diel model constructed from the Arabidopsis genome-scale metabolic model in ScrumPy and SBML formats
Appendix S4	MFA flux maps of Arabidopsis heterotrophic cell culture under control and stress conditions
Appendix S5	Flux solutions of mature C ₃ leaf under different environmental or physiological conditions calculated from flux minimisation with random weighting factors
Appendix S6	Details on energy-use stoichiometries for transporters and protein translocation
Appendix S7	Flux solutions of Arabidopsis heterotrophic cell culture from 25 combinations of objective function and model constraints
Appendix S8	Calculation on maintenance costs in of Arabidopsis heterotrophic cell culture under control and stress conditions from MFA flux maps
Appendix S9	List of reactions with non-overlapping flux ranges between the diel model and the single steady-state model and their similarity measures
Appendix S10	Flux ranges of solutions of mature C ₃ and generic CAM leaves
Appendix S11	Flux solutions of a mature C ₃ leaf generated using flux minimisation as the objective function with 1000 different sets of random weighting factors
Appendix S12	Pearson correlation coefficients between reactions calculated from flux solutions of a mature C ₃ leaf

Appendix 1

Description of the ScrumPy commands for setting model constraints and objective functions specific to this study

Modelling codes written for setting up the specific model constraints and objective functions used in this study are contained mainly in “LPARA.py” (Appendix S1) unless state otherwise. A prerequisite is that a linear programming object, “lp”, was created using “lp=LPARA.lp(m,‘culture’)” for modelling heterotrophic cell culture and “lp=LPARA.lp(m,‘lightdark’)” for modelling diel leaf metabolism. The biomass constraints for heterotrophic cell culture (chapter 4) was imported into ScrumPy by “fd=lp.FluxDics()” where “fd” is a dictionary with the experimental condition labels, “U” – control, “H” – heat and “M” – hyperosmotic, as keys and flux dictionaries, “fluxdic”, as values. A flux dictionary can then be used to set a fixed biomass constraint using “lp.SetFixedFlux(fluxdic)”. This requires the information on biomass constraint to be in a file with path “./analysis/biomasstom.txt” in a tabular format as in “biomasstom.txt” in Appendix S1. NADPH maintenance was assumed to be distributed equally in the cytosol, the plastid and the mitochondrion throughout this study. This constraint was applied by “nadphreac=lp.SetNADPHoxReac()”, which uses the function “lp.SetReacsFixedRatio()” described in section 2.5, where “nadphreac” is the reaction name of the overall generic NADPH maintenance reaction. The proportion of inorganic nitrogen input as nitrate, versus ammonia, was set by the function “lp.SetNRatio()”. When modelling heterotrophic cell culture, the transport costs can be remove by the command “lp.RemoveTransportCosts()” which can be reversed by “lp.IncludeTransportCosts()”.

For modelling leaf metabolism (chapters 5 and 6), the C₃- and CAM-specific constraints were applied using “lp.SetC3()” and “lp.SetCAM()” respectively. The proportion of Rubisco flux through the oxygenase reaction was set using the function “lp.SetPRRatio()”. In the diel model, the proportion of nitrate uptake during the day, versus at night, was set by “lp.SetNO3in Ratio()”. “lp.SetLightDarkExport()” and “lp.SetLightDarkMaintenance()” set the export of sucrose and amino-acids and the ATP and NADPH maintenance costs in

the proportions stated in chapter 5. For modelling leaf growth, a reaction for the export of sucrose and amino-acids into the phloem in the proportion described in chapter 6 was created using the command `"exportreac=lp.SetLightDarkExportReac()"` where `"exportreac"` is the reaction name of the export reaction. Similarly, a reaction for biomass synthesis was created by `"biomassreac=lp.SetLightDarkBiomassReac()"` where `"biomassreac"` is the reaction name of the biomass synthesis reaction. This requires the information on the biomass composition to be in a file with path `"/analysis/biomassaragem.txt"` (Appendix S1).

The five objective functions explored in chapter 4 were set using the following functions: `"lp.SetMinFluxObj()"` for minimising total flux, `"lp.SetMinGLCObj()"` for minimising glucose consumption, `"lp.SetMaxBiomassObj()"` for maximising biomass synthesis, `"lp.SetMaxATPaseObj()"` for maximising ATP production and `"lp.SetMaxNADPHoxObj()"` for maximising NADPH production. Note that maximising biomass requires the definition of a single biomass reaction. This was done using the code in `"SMPlus.py"` (Appendix S1). Instead of loading the model using `"m=ScrumPy.Model()"` as described in section 2.2.5, the model was loaded using the command `"m=SMPlus.StoModel()"` after importing `"SMPlus.py"`. The biomass reaction, with reaction name `"Biomass_tx"` by default, was then added to the model by `"m.AddBiomassReaction(fluxdic)"` where `"fluxdic"` is a flux dictionary containing information on the biomass composition which can be generated from `"lp.FluxDics()"`. A linear programming objective was then created by `"lp.LPAra.lp(m)"` as described in section 2.5.

Appendix 2

Reaction mapping between MFA model and genome-scale metabolic model

¹³ C-MFA reaction identifier	Genome-scale model reaction identifier
cpex	- G6P_Pi_pc
chex2	$[(6\text{PFRUCTPHOS-RXN_c} + 2.7.1.90\text{-RXN_c} - \text{F16BDEPHOS-RXN_c}) + \text{F16ALDOLASE-RXN_c}] / 2$
phex1	$[(6\text{PFRUCTPHOS-RXN_p} - \text{F16BDEPHOS-RXN_p}) + \text{F16ALDOLASE-RXN_p}] / 2$
cphex	$[(\text{GAPOXNPHOSPHN-RXN_c} + \text{GAPOXNPHOSPHN-RXN_p} - 1.2.1.13\text{-RXN_p}) + (-\text{PHOSGLYPHOS-RXN_c} - \text{PHOSGLYPHOS-RXN_p})] / 2 + 1.2.1.9\text{-RXN_c}) + (3\text{PGAREARR-RXN_c} + 3\text{PGAREARR-RXN_p}) + (2\text{PGADEHYDRAT-RXN_c} + 2\text{PGADEHYDRAT-RXN_p})] / 3$
chex3	$(\text{PEPDEPHOS-RXN_c} - \text{PYRUVATEORTHOPHOSPHATE-DIKINASE-RXN_c})$
phex2	$(\text{PEPDEPHOS-RXN_p} - \text{PYRUVATEORTHOPHOSPHATE-DIKINASE-RXN_p})$
cmex = tca1 = tca2	$[- \text{PYRUVATE_mc} + \text{PYRUVDEH-RXN_m} + \text{CITSYN-RXN_m}] / 3$
tca3	$[(\text{ACONITATEDEHYDR-RXN_c} + \text{ACONITATEDEHYDR-RXN_m}) + (\text{ACONITATEHYDR-RXN_c} + \text{ACONITATEHYDR-RXN_m}) + (\text{ISOCITRATE-DEHYDROGENASE-NAD+ -RXN_m} + \text{ISOCITDEH-RXN_c} + \text{ISOCITDEH-RXN_m} + \text{ISOCITDEH-RXN_x})] / 3$
tca4a+b	$[(2\text{OXOGLUTARATEDEH-RXN_m}) + (\text{SUCCCOASYN-RXN_m} + \text{SUCCINATE—COA-LIGASE-GDP-FORMING-RXN_m})] / 2$
tca5a+b	$(\text{FUMHYDR-RXN_c} + \text{FUMHYDR-RXN_m})$
ppp1	$[(\text{GLU6PDEHYDROG-RXN_c} + \text{GLU6PDEHYDROG-RXN_p}) + (6\text{PGLUCONOLACT-RXN_c} + 6\text{PGLUCONOLACT-RXN_p}) + (6\text{PGLUCONDEHYDROG-RXN_c} + 6\text{PGLUCONDEHYDROG-RXN_p})] / 3$
ppp3 = (ppp2 - ppp4)	$- 1\text{TRANSKETO-RXN_p}$
ppp4 = (ppp2 - ppp3)	2TRANSKETO-RXN_p
ana1	$(\text{PEPCARBOX-RXN_c} - \text{PEPCARBOXYKIN-RXN_c})$
ana3	$(\text{MALIC-NADP-RXN_c} + \text{MALIC-NADP-RXN_p} + \text{MALIC-NAD-RXN_m})$

Reactions by common names

Enzyme Name	Genome-scale model reaction identifier	¹³ C-MFA reaction identifier
Glycolysis		
glucose 6-phosphate transporter (plastidial)	G6P_Pi_pc	cpex
phosphofructokinase (ATP) (cytosol)	6PFRUCTPHOS-RXN_c	chex2
phosphofructokinase (PPi) (cytosol)	2.7.1.90-RXN_c	chex2
phosphofructokinase (ATP) (plastid)	6PFRUCTPHOS-RXN_p	phex1
aldolase (cytosol)	F16ALDOLASE-RXN_c	chex2
aldolase (plastid)	F16ALDOLASE-RXN_p	phex1
GADPH (cytosol + plastid) & phosphoglycerate kinase (cytosol + plastid) and GADPH (non-phosphorylating) (cytosol)	(GAPOXNPHOSPHN-RXN_c + GAPOXNPHOSPHN-RXN_p + PHOSGLYPHOS-RXN_c + PHOSGLYPHOS-RXN_p)/2 + 1.2.1.9-RXN_c	cphex
phosphoglycerate mutase (cytosol + plastid)	3PGAREARR-RXN_c + 3PGAREARR-RXN_p	cphex
enolase (cytosol + plastid)	2PGADEHYDRAT-RXN_c + 2PGADEHYDRAT-RXN_p	cphex
pyruvate kinase (cytosol)	PEPDEPHOS-RXN_c	chex3
pyruvate kinase (plastid)	PEPDEPHOS-RXN_p	phex2
TCA cycle		
pyruvate transporter (mitochondrial)	PYRUVATE_mc	cmex
pyruvate dehydrogenase (mitochondrion)	PYRUVDEH-RXN_m	tca1
citrate synthase (mitochondrion)	CITSYN-RXN_m	tca2
aconitate hydratase (cytosol + mitochondrion)	(ACONITATEDEHYDR-RXN_c + ACONITATEDEHYDR-RXN_m) or (ACONITATEHYDR-RXN_c + ACONITATEHYDR-RXN_m)	tca3

Enzyme Name	Genome-scale model reaction identifier	¹³ C-MFA reaction identifier
isocitrate dehydrogenase (cytosol + mitochondrion + peroxisome)	ISOCITRATE-DEHYDROGENASE-NAD+-RXN_m + ISOCITDEH-RXN_c + ISOCITDEH-RXN_m + ISOCITDEH-RXN_x	tca3
α-ketoglutarate dehydrogenase (mitochondrion)	2OXOGLUTARATEDEH-RXN_m	tca4a+b
succinate succinate-CoA ligase (mitochondrion)	SUCCCOASYN-RXN_m + SUCCINATE—COA-LIGASE-GDP-FORMING-RXN_m	tca4a+b
fumarase (cytosol + mitochondrion)	FUMHYDR-RXN_c + FUMHYDR-RXN_m	tca5a+b
Pentose Phosphate Pathway		
glucose 6-phosphate dehydrogenase (cytosol + plastid)	GLU6PDEHYDROG-RXN_c + GLU6PDEHYDROG-RXN_p	ppp1
6-phosphogluconolactonase (cytosol + plastid)	6PGLUCONOLACT-RXN_c + 6PGLUCONOLACT-RXN_p	ppp1
phosphogluconate dehydrogen (cytosol + plastid)	6PGLUCONDEHYDROG-RXN_c + 6PGLUCONDEHYDROG-RXN_p	ppp1
transketolase 1 (plastid)	1TRANSKETO-RXN_p	ppp3
transketolase 2 (plastid)	2TRANSKETO-RXN_p	ppp4
transaldolase (plastid)	TRANSALDOL-RXN_p	ppp5/ppp6
Others		
PEP carboxylase (cytosol) and PEP carboxykinase	PEPCARBOX-RXN_c – PEPCARBOXYKIN-RXN_c	ana1
malic enzyme (cytosol + mitochondrion + plastid + peroxisome)	MALIC-NADP-RXN_c + MALIC-NADP-RXN_p + MALIC-NAD-RXN_c + MALIC-NAD-RXN_m + MALIC-NAD-RXN_x	ana3

Appendix 3

Summary of energy-use stoichiometries for transporters and protein translocation

Membrane energisation

Plasma membrane	1	H ⁺ per ATP
Mitochondrial membrane	3.33	H ⁺ per ATP
Tonoplast	2	H ⁺ per ATP

Metabolite transport

Transport metabolite		Stoichiometry	Membrane
Sucrose	1	H ⁺ per molecule	Plasma membrane
Glucose	1	H ⁺ per molecule	Plasma membrane
Nitrate	2	H ⁺ per molecule	Plasma membrane
Amino-acid	1	H ⁺ per molecule	Plasma membrane
Ammonium	0	H ⁺ per molecule	Plasma membrane
Sulphate	3	H ⁺ per molecule	Plasma membrane
Pi	3	H ⁺ per molecule	Plasma membrane
K	1	H ⁺ per molecule	Plasma membrane
Ca	1	H ⁺ per molecule	Plasma membrane
Mg	1	H ⁺ per molecule	Plasma membrane
Sucrose	0	H ⁺ per molecule	Tonoplast
Glucose	1	H ⁺ per molecule	Tonoplast
Nitrate	0.5	H ⁺ per molecule	Tonoplast
Ammonia	0	H ⁺ per molecule	Tonoplast
Organic acids	0	H ⁺ per molecule	Tonoplast
K	1	H ⁺ per molecule	Tonoplast
Ca	3	H ⁺ per molecule	Tonoplast
Mg	2	H ⁺ per molecule	Tonoplast
Pi	1	H ⁺ per molecule	Mitochondrial membrane
Pyruvate	1	H ⁺ per molecule	Mitochondrial membrane
Fatty acid	2	ATP per molecule	Peroxisomal membrane

Protein translocation

ATP per amino-acid (in protein) translocated	2.89017341
Proportion of protein translocated	50%