Modelling Metabolic Interactions in the Legume-Rhizobia Symbiosis

Thomas Pfau

Diplom Bioinformatik, Eberhard-Karls-Universität Tübingen

A thesis presented for the degree of Doctor of Philosophy at the University of Aberdeen

2013

Summary

With the emergence of "omics" techniques, it has become essential to develop tools to utilise the vast amount of data produced by these methods. Genome-scale metabolic models represent the mathematical essence of metabolism and can easily be linked to the data from omics sources. Such models can be used for various analyses, including the investigation of metabolic responses to changing environmental conditions. Legumes are known for their ability to form a nitrogen-fixing symbiosis with rhizobia, a vital process that provides the biosphere with the majority of its nitrogen content. In the present thesis, a genome-scale metabolic model for the legume *Medicago truncatula* was reconstructed, based on the annotated genome sequence and the MedicCyc database. A novel approach was employed to define the compartmentalisation of the plant's metabolism. The model was used to calculate the biosynthetic costs of biomass precursors (e.g. amino acids, sugars, fatty acids, nucleotides), and its capability to produce biomass in experimentally observed ratios was demonstrated using flux balance analysis. Further investigation was carried out into how the biosynthesis fluxes and costs change with respect to different nitrogen sources. The precise charge balancing of all reactions in the model allowed the investigation of the effects of charge transport over the cellular membrane. The simulations showed a good agreement with experimental

data in using different sources of nitrogen (ammonium and nitrate) to minimise the charge transport of the membrane. To allow the investigation of the symbiotic relationship, two rhizobial models were used. The first model, for *Sinorhizobium meliloti*, was reconstructed from the MetaCyc database (MC-model); the second model was a recently published model for *S. meliloti* specialised for symbiotic nitrogen fixation (SNF-model). Combined models were created for both rhizobial networks using a specialised nodule submodel of the plant model. Potential interactions were extracted from the literature and investigated, with the analysis suggesting that oxygen availability is the main limitation factor in symbiotic nitrogen fixation. Within the analysis the SNF-model appeared to be too restricted and lacking the potential for sufficient nitrogen fixation; therefore, further analysis was carried out using the MC-model, upon which it was observed that the availability of oxygen can also influence how nitrogen is supplied to the plant. At high oxygen concentrations ammonia is the primary form of nitrogen supplied by the rhizobium. However, the simulations, in accordance with experimental data, show that at lower concentrations of oxygen, alanine takes precedence. The findings also support the concept of amino acid cycling as a potential way to improve nitrogen fixation. The more flexible MetaCyc based model has allowed other potential genetic engineering approaches for higher nitrogen fixation yields to be proposed.

Declaration

I hereby declare that this thesis has been composed by the named author. It has not been accepted in any previous application for a degree. The work described herein has been done by the author and any contributions from other persons have been acknowledged in the text. Quotations have been distinguished by quotation marks and all sources of information have been specifically acknowledged.

Acknowledgements

I would like to thank Oliver for offering me the possibility to do this work and for supporting me during my time in Aberdeen. He provided the possibilities to attend conferences which have inspired parts of this work and without him I would have had a much harder time. I want to express my gratitude for the help by Mark in starting this project and thank all the people in his group, in particular Hassan and Maurice, for the fruitful discussions. I would like to thank all the PhD students and PostDocs at the ICSMB for the seminars and in particular Jens and Luca for organising them. A special thank you to Beth for proofreading and providing valuable comments. I am grateful for the advice and support that Mamen gave me. I would like to acknowledge the help I got from the people in the lab of Lee Sweetlove, while doing my biomass experiments. In particular I would like to thank Shyam for providing so much help and being so patient with me in the lab. I am grateful to all the people in the lab of the late Gail Ferguson, who helped me up here in Aberdeen and were supportive even when I was just showing up without prior notice. A special thank goes to Andreas, who went to some length in organising the seeds for my plants, which in the end he got from Australia.

And finally and above all, I am eternally grateful for the support of my wife and all the help she provided me with during my PhD. Without her I doubt that this thesis would have become the piece it now is and I would likely still struggle with the very first sentences.

Contents

List of Figures

List of Tables

Nomenclature

Acronyms

2KG 2-ketoglutarate

6PG 6-phosphogluconate

ADP Adenosine diphosphate

ATP Adenosine triphosphate

CoA Coenzyme A

DCT Dicarboxylic acid transporter

DHAP Dihydroxyacetone phosphate

DPG 1,3-bisphosphoglycerate

DW Dry weight

E4P Erythrose 4-phsophate

ED End of day

EFM Elementary flux mode

EN End of night

EP Extreme pathway

F16BP Fructose 1,6-bisphophate

F6P Fructose 6-phosphate

FBA Flux balance analysis

FW Fresh weight

G3P 3-phosphoglycerate

G6P Glucose 6-phosphate

GAP D-glyceraldehyde 3-phosphate

GOGAT Glutamate synthase

GS Glutamine synthase

MCS Minimal cut set

MILP Mixed integer linear program

NAD Nicotinamide adenine dinucleotide

NADH Reduced nicotinamide adenine dinucleotide

NADP Nicotinamide adenine dinucleotide phosphate

NADPH Reduced nicotinamide adenine dinucleotide phosphate

PEP Phosphoenolpyruvate

PFK Phosphofructokinase

PPI Pyrophosphate

PPP Pentose phosphate pathway

PYC Pyruvate Carboxylase

Ri5P Ribose 5-phosphate

Ru5P Ribulose 5-phosphate

RuBisCO Ribulose bisphosphate carboxylase oxygenase

S7P Sedoheptulose 7-phosphate

TCA Tricarboxylic acid cycle

X5P Xylose 5-phosphate

Chapter 1

Introduction

I found it difficult to devise a way to introduce the project of this thesis. The main reason is the high specificity of the organisms investigated along with the complexity of general principles governing this field of research. When asked for a one sentence description of my work, I would describe it as:

> "An investigation of the metabolism of *Medicago truncatula* and its metabolic interactions with its symbiont *Sinorhizobium meliloti*."

However, there are several key fields in this sentence which need explanation to allow the reader to place it into context. The first of those fields is metabolism, and many textbooks cover this topic in extensive chapters [1, 2]. I will only give an overview of the parts of metabolism most relevant for this work (Section 1.1.2) and will discuss some general principles to help with understanding of the functions of metabolism in general (see Section 1.1.1). I will mainly focus on the descriptions of the energy and nitrogen metabolism of the two investigated organisms, as these are central to my

work. Those organisms are *Medicago truncatula*, a clover, and *Sinorhizobium meliloti*, a soil bacterium, and I will provide an overview of their properties and particularities in sections 1.2.1 and 1.2.2. The symbiosis between this organisms is a model for the nitrogen fixing symbiosis of legumes and rhizobia, with the importance of this symbiosis stemming from the importance of nitrogen within biological systems. Nitrogen atoms are present in a large fraction of important biomolecules (i.e. molecules synthesized in a living organism). They are part of DNA; the genome or the blueprint of most known organisms (with the exception of some viruses). It is also essential for the amino acids that build the proteins forming the working entities within cells. However, while nitrogen makes up 78% of the earth's atmosphere [3], its availability in nutrients is rather restricted due to the inherent inertia of nitrogen. Lavoisier called it an azotic gas, from the negated Greek word for life (i.e. lifeless) [4]. There are a limited number of ways that incorporate nitrogen into the biosphere with the most important being:

- lightning ($\sim$10%)
- artificial fertilizers ($\sim$25%)
- biological nitrogen fixation ($\sim$60%) [5]

The most dominant biological process is the legume-rhizobial symbiotic nitrogen fixation [6]. Agricultural nitrogen fixation by legume crops was about 40 Tg/year in 1994 [7]. Supplying this nitrogen through artificial fertilizers would require an equivalent of 5 million tons of fossil fuel, consumed during the production of the fertilizer when using methods like the Haber-Bosch Process [8]. Therefore, a better understanding of this symbiosis can lead to:

- Higher agricultural yields by better fertilization;

- A reduced amount of fossil fuels needed to produce artificial fertilizer; and
- A natural way to cultivate barren zones.

The potential impact of these outcomes is a convincing reason to pursue the investigation of these organisms and their symbiosis. The symbiotic properties will be presented in detail in Section 1.2.4, where I will provide an overview of the symbiotic process between legumes and rhizobia.

The final part of this introduction will concern the field I use to investigate these organisms. I will introduce the concepts used and the assumptions made in stoichiometric metabolic modelling and will give an overview of the possibilities and restrictions of this approach.

1.1 Metabolism

"**Metabolism** is the overall process through which living systems acquire and utilize the free energy they need to carry out their various functions."

Voet and Voet [1, p. 557].

"[...]**metabolism** , the unceasing, complex network of thousands of chemical reactions by which cells grow and reproduce, take up foods and excrete wastes, move, and communicate with each other"

Metzler [2, p. 1]

These two statements focus their view of metabolism on the important processes that keep cells running and thriving: energy generation and growth. The main problem with investigating metabolism can be seen in the second quote. With thousands of highly interconnected reactions a metabolic network (see Figure 1.1) is hard to conceptualise. Even in the complex structure of the metabolic network of a eukaryotic cell there is a directly observable order; each cluster formed by a layout algorithm represents the part of metabolism in one of the compartments of a cell. Still, even this reaction network is only a fraction of the processes that control and regulate metabolism. In a living cell, most of these processes have to be performed by enzymes, which in turn are coded for in the cells' genome. Enzymes themselves are a product of metabolism since they are produced by combining amino acids at the expense of energy. For most modern work, it is essential to know which capabilities a given organism has. To define these capabilities the "Central Dogma of Molecular Biology" [9, 10] has proven to be of great importance. It states that one gene codes for one RNA and one RNA for one protein. This dogma holds true in prokaryotes but is somewhat invalidated by

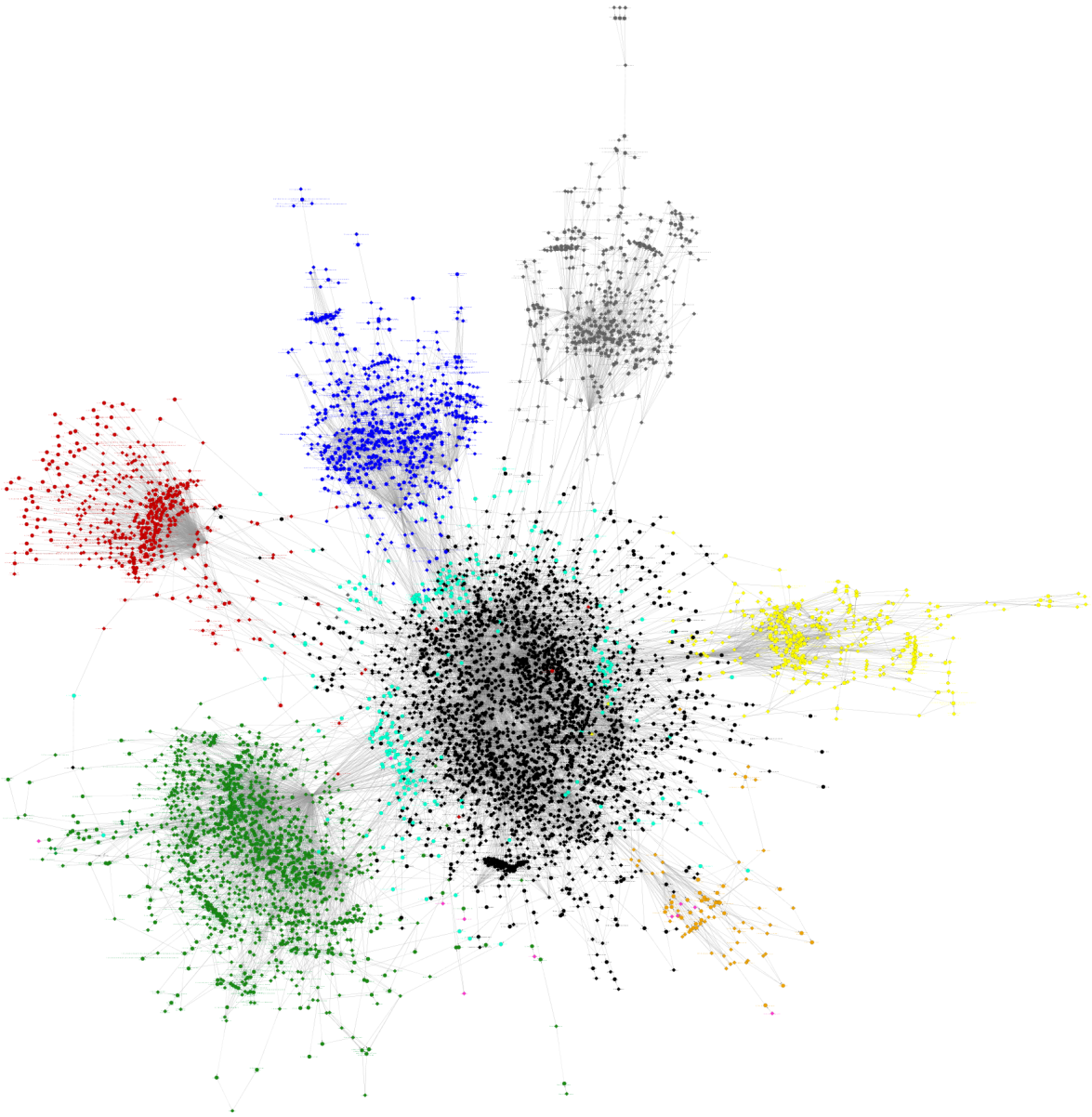


Figure 1.1: A simple representation of the genome scale metabolic network of *M. truncatula* laid out with the organic layout from yTools in Cytoscape. The compartments of the network nicely cluster and are depicted in different colours (black - cytosol; green - plastid; blue - mitochondrion; yellow - peroxisome; grey - vacuole; red - endoplasmic reticulum; orange - Golgi apparatus; cyan - transporters; magenta - external metabolites).

alternative splicing in many eukaryotes [11]. Since enzymes are the active forms of proteins, the dogma is of high importance for metabolic reconstruction. However, not all enzymes catalyse one highly specific reaction. Instead, they can often utilize similar substrates and perform the same general type of reaction on all of these substrates. However, even with this diversity, important information can be retrieved from the genome of an organism. By single protein analysis, it is possible to determine what kind of reaction an enzyme catalyses. Due to the vast amount of available data, this process can be automated. There exist well-established methods to infer probable enzymatic activity by homology analysis (e.g. using BLAST [12, 13]) from either the protein sequence or the genomic sequence. Although this provides a huge repertoire of information on single enzymes, the information on all enzymes needs to be combined to investigate the complete metabolism of an organism. A single cell contains thousands of enzymes, and with each catalysing one or more reactions, the metabolism becomes a complex network. It is therefore necessary to bring structure to this network in order to communicate findings more efficiently. To do so, the concept of pathways has proven useful, even though it is likely to create the impression that there is more order in metabolism than actually present. This concept and some examples will be explained in Section 1.1.2. To understand the metabolic processes contained in these pathways, it is important to know the driving forces of metabolism, to which Voet and Voet give a good indication in their description of metabolism above. As they say, "Living organisms are not at equilibrium" [1, p. 559] and therefore require energy to maintain their internal processes.

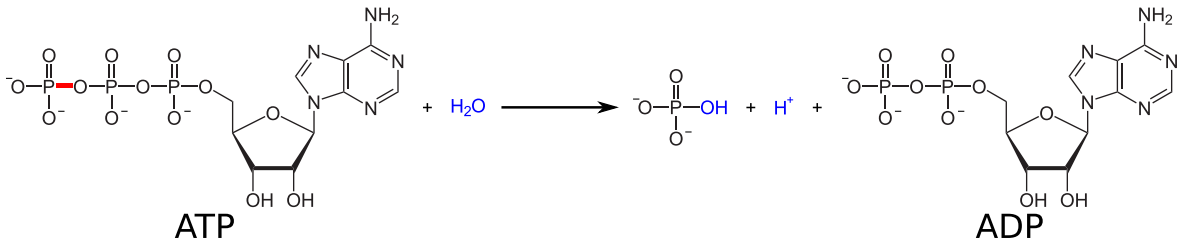


Figure 1.2: Hydrolysis of adenosine triphosphate (ATP). The red bond is broken and releases 20.5 kJ/mol. This reaction is part of many metabolic processes, and provides energy for endergonic reactions. Water is not necessarily the only substrate for the transfer of the phosphate group and other substrates can retain the energy or only release smaller amounts of energy. For clarification, the atoms originating from water in the substrate and product are marked in blue. Figure created with Inkscape.

1.1.1 The forces behind metabolism

The purpose of metabolism is to provide energy to the organism. To understand the processes involved, it is essential to realise how an organism handles energy. To sustain its present state or to grow, a cell needs to expend energy in order to counter the universal drive to entropy. While energy is available in many forms, a common energy currency is necessary to facilitate energy transfer between different enzymes within a cell. Without such a currency it would be infeasible to transfer energy between different enzymes. While energy can be stored in large molecules on a long term basis, on a shorter time scale it is necessary to have a fast "battery" which most enzymes can readily use. This battery is present in the form of adenosine triphosphate (ATP, Figure 1.2). Hydrolysing (i.e. splitting the bond with water) the phosphate bond of ATP yields about 20.5 kJ/mol [14]. Per mole of ATP this is approximately double the amount of energy stored in one AA battery. However, there is only a very small amount of ATP present in an organism and it has to be constantly regenerated by metabolic processes. ATP functions as an energy source for many endergonic reactions and is of-

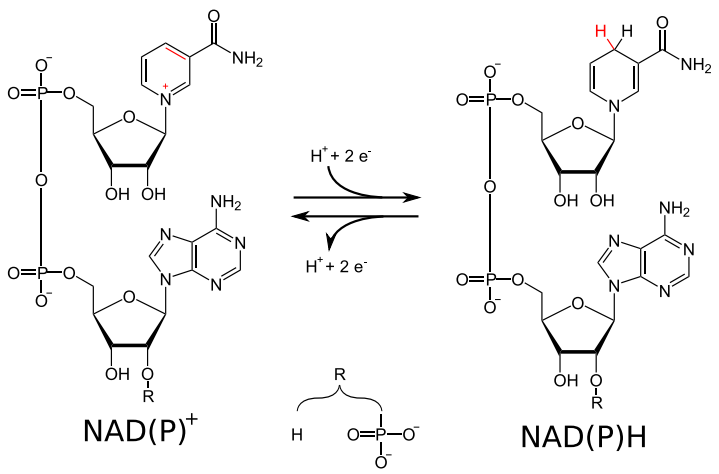


Figure 1.3: Reduction of nicotineamide adenine dinucleotide (phosphate) $NAD(P)^+$ to NAD(P)H. Along with two electrons, one proton is transferred from a donor to NAD(P). Some enzymes can use both substrates while others are specific to one of the molecules. Figure created with Inkscape.

ten described as the "energy currency" of a cell. The second important currency within a cell are electron carriers. These, in contrast to ATP, come in many different varieties. They allow an organism to transfer electrons from one molecule to another, which is essential for most biosynthetic processes. Even more importantly, some of them are used to drive the electron transfer chain which regenerates ATP from adenosine diphosphate (ADP, Figure 1.2) in photosynthesis or oxidative phosphorylation (see sections 1.1.2.4 and 1.1.2.3). The most important electron carriers nicotineamide adenine dinucleotide (NAD^+) and its phosphate ($NADP^+$) can be seen in Figure 1.3. However, NADH and NADPH do not only carry electrons; in contrast to several other carriers the electrons are accompanied by protons. Reduced electron carriers (like NADPH) are often called reductants, as their purpose in most reactions is to reduce the oxidisation state of another molecule. Reductants and energy (in the form of ATP) are not exclusively used for metabolic purposes - energy, for example, is needed for transportation of small

molecules and macromolecules. While the transport of small molecules is often only indirectly linked to ATP hydrolysis [15], macromolecules require large amounts of direct ATP hydrolysis for transport over membranes [16, 17]. Furthermore, energy is necessary for motility [18] and many other mechanic processes. This means that there is a permanent drain on the pool of available ATP from outside the metabolism and a permanent influx of ADP and phosphate (Pi) into the metabolism. Reducing power is also in high demand, as it is, for example, required to repair damage to the cell caused by external oxidants. This means that much of the potentially available energy from nutrition is diverted towards maintenance and repair processes. In fact, in plants, between 40 and 70% of the potential energy (and reductant) is not used for metabolic processes but diverted to satisfy other demands [19, 20]. This turnover demand has to be met, and the following sections will provide examples of energy regenerating pathways, after an introduction of the pathway concept.

1.1.2 Metabolic pathways

A single enzyme catalyses one (or one kind of) reaction which may be denoted as:

$$A + B \leftrightharpoons C + D \tag{1.1}$$

If several reactions are necessary to convert a compound A into a compound D, the concept of pathways is useful to illustrate this process and to allow simultaneous discussions about larger parts of metabolism. Conceptually, metabolism can be divided into reactions which are essential for the survival of the species (primary metabolism), and those which are likely to be beneficial but which are not essential (secondary metabolism). There are several metabolic pathways belonging to basic metabolism

which are common to many organisms. A selection of these pathways is detailed here, as they have shown to be relevant to many processes investigated in this thesis.

1.1.2.1 Glycolysis

Perhaps the most important and best studied pathway is the glycolytic pathway [1, p. 593ff]. Sugars ($C_nH_{2n}O_n$) are the main source of energy and carbon for many organisms. In plants, glucose ($C_6H_{12}O_6$) is stored as starch via the formation of ADP-glucose and transfer of the glucose moiety to the starch granule during the day. This starch granule is then degraded and the glucose is recovered during the night [21, p. 182ff]. The initial process employed to extract the energy stored in the glucose molecule is called glycolysis. It converts one molecule of glucose into two molecules of pyruvate (see Figure 1.4) while regenerating ATP and reductants. The pathway can be split into two phases, a preparation phase, from glucose to two glyceraldehyde 3-phosphate (GAP), and a pay-off phase from GAP to pyruvate. In the preparation phase ATP is needed to phosphorylate glucose to glucose 6-phosphate and fructose 6-phosphate to fructose 1,6-bisphosphate (F16BP).

The latter reaction is commonly catalysed by phosphofructokinase (PFK). In plants, an alternative enzyme is present that also catalyses this reaction. Instead of ATP, the alternate plant enzyme can use pyrophosphate as phosphate donor [23]. Following these two phosphorylation steps, F16BP is split into two triose-phosphates, dihydroxyacetone phosphate (DHAP) and GAP. The final step in the preparation phase is the conversion of DHAP into a second GAP. In total, two phosphate bonds have to be broken and transferred to glucose for this preparation, in energy consuming steps. In the pay-off phase, this energy is not only recovered but a surplus is achieved, as all of the following reactions are carried out twice per glucose molecule. In the initial reaction of the pay-off

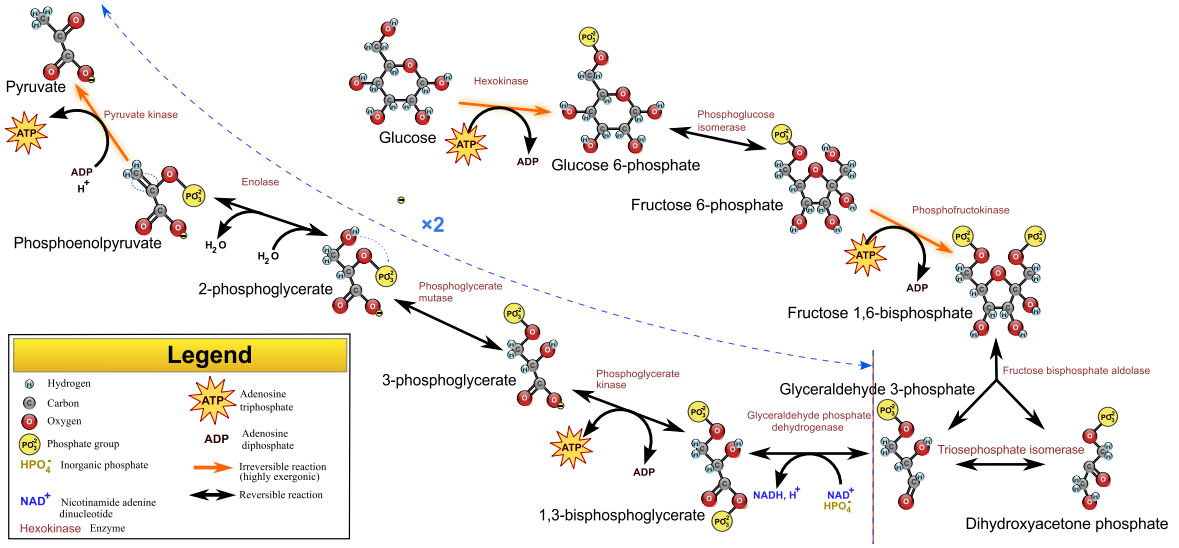


Figure 1.4: Glycolysis from glucose to pyruvate. In the initial steps two ATP are consumed while in the later steps a total of four ATP and two NADH are regenerated per glucose. Image adapted from Yassine Mrabet under the Creative Commons License 3.0 BY-SA [22].

phase, GAP is oxidized and at the same time phosphorylated with inorganic phosphate by the GAP dehydrogenase to yield 1,3-bisphosphoglycerate (DPG). This step yields an NADH which can later be converted into additional energy (see Section 1.1.2.3). In the last four steps, from DPG to pyruvate, two ATP are regenerated using the energy and phosphate from the phosphate bonds of DPG and phosphoenolpyruvate. In total, the pay-off phase yields four ATP and two NADH per glucose. The overall glycolytic balance is therefore

$$Glucose + 2\,ADP + 2\,Pi + 2\,NAD^{+} \implies 2\,Pyruvate + 2\,NADH + 2H^{+} + 2\,ATP \tag{1.2}$$

Along with pyruvate being an important precursor for many biomolecules (e.g. in amino acid biosynthesis [24]) it is also the source of further energy generation in the tricarboxylic acid cycle [21, p. 262ff] (see Section 1.1.2.2). Glycolysis therefore provides

the organism both with an important biomolecule precursor, as well as energy and reductants for maintenance and biosynthetic processes.

1.1.2.2 Tricarboxylic acid cycle

The tricarboxylic acid cycle (TCA) is the main source of carbohydrate oxidation, and therefore reductant regeneration in most cells, and provides precursors for a multitude of biomolecules [1, p. 789ff]. Figure 1.5 gives an overview of the TCA and its intermediates. The cycle is driven by pyruvate obtained from glycolysis which is decarboxylated and attached to coenzyme A (CoA) yielding acetyl-CoA and regenerating one NADH.

Acetyl-CoA is then combined with oxaloacetic acid leading to the formation of citrate. By water addition and elimination citrate is converted into iso-citrate. At this point the energy extraction starts with dehydrogenating and decarboxylating reactions (from iso-citrate to α-ketoglutarate and from the latter to succinyl-CoA). The latter of the two reactions, α-ketoglutarate dehydrogenase, also ligates CoA to the resulting C4-carboxylic acid. The energy from this bond is used to catalyse the formation of ATP from ADP in plants or from guanosine diphosphate to guanosine triphosphate in animals. In the final steps the C4-dicarboxylic acids are dehydrogenated twice, yielding an additional NADH and producing coenzyme QH_2 which can be used in the electron transfer chain (see Section 1.1.2.3).

1.1.2.3 The electron transfer chain and oxidative phosphorylation

The electron transfer chain is the last step before the regeneration of ATP by oxidative phosphorylation and both processes have been described extensively [1, 2, 21, 26]. The mechanism is common to both prokaryotic and eukaryotic organisms, with current thought suggesting that eukaryotic cells have at some point taken up prokaryotes

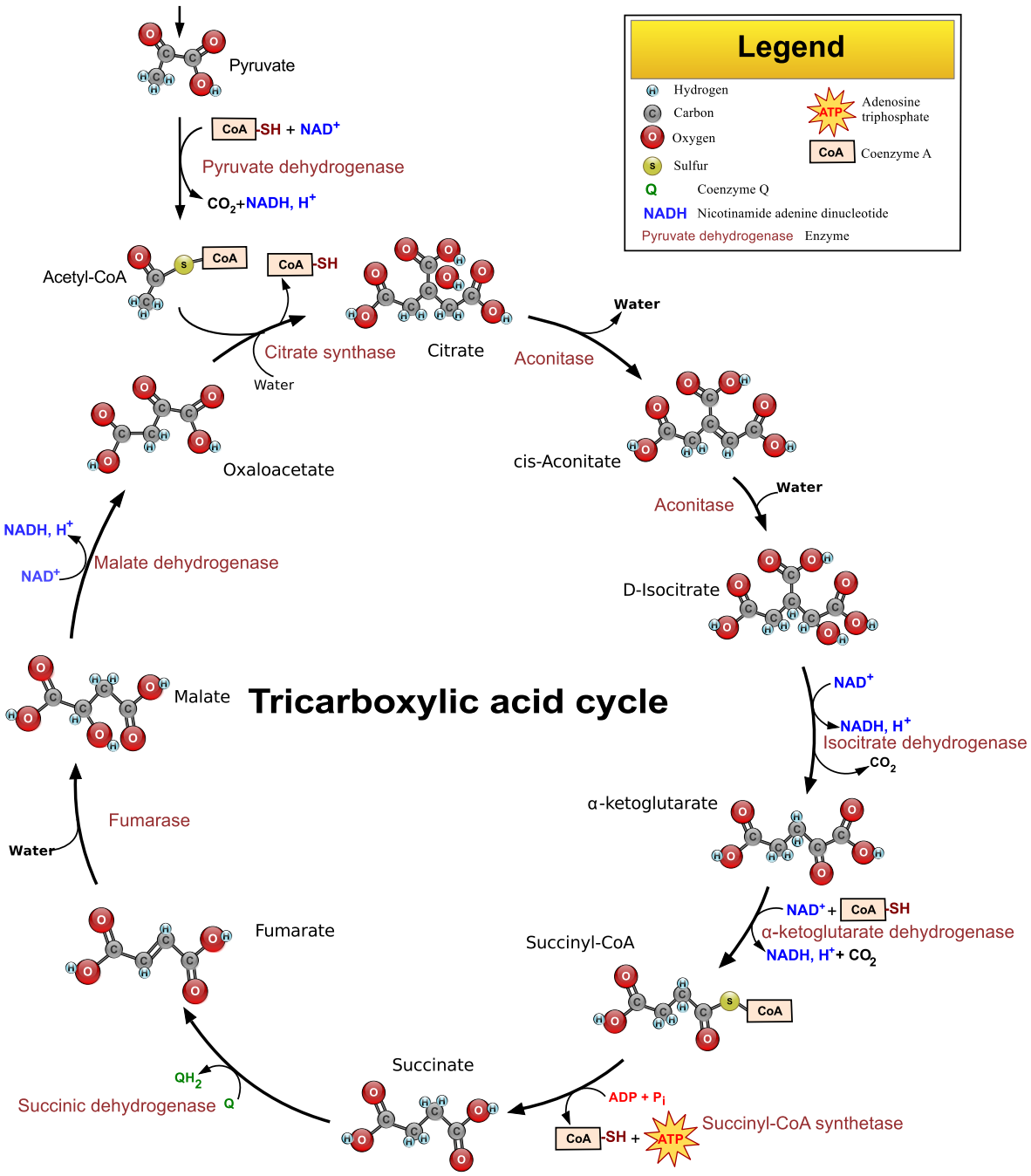


Figure 1.5: The Krebs or tricarboxylic acid cycle (TCA). An acetyl group is attached to coenzyme A (CoA) by decarboxylation of pyruvate, while simultaneously NADH is regenerated. Acetyl-CoA and oxalacetic acid are combined to form citrate, which is converted to iso-citrate and dehydrated to form the next NADH, the energy being obtained from splitting of another CO_2. Two additional NADH are regenerated by ketoglutarate dehydrogenase, and malate dehydrogenase and one GTP is obtained by splitting of Co-enzyme A from succinate. The image is a modified version of work by Narayanese, WikiUserPedia, YassineMrabet, and TotoBaggins licensed under Creative Commons 3.0 BY-SA [25].

to perform this task for them [27]. Over time these prokaryotes have evolved into the mitochondria of eukaryotic cells, which are now known as the power plants of eukaryotic cells. An overview of the processes involved in the electron transfer chain can be seen in Figure 1.6. The electron transfer chain is essential in that it creates a chemical gradient over a membrane thus allowing the ATP synthase to regenerate ATP. The membrane involved is either the mitochondrial inner envelope membrane or the cellular membrane in eukaryotes and prokaryotes, respectively. The energy for the gradient generation is obtained by gradual transfer of electrons to lower energy levels. Four enzyme complexes (I-IV) are involved in the chain: NADH-coenzyme Q oxidoreductase (complex I), succinate dehydrogenase (complex II), coenzyme Q-cytochrome *c* oxidoreductase (complex III) and cytochrome *c* oxidase (complex IV). There are two main points of entry into the chain, at complex I or at complex II by electron transfer from NADH and succinate respectively. The most important enzymes are complexes I, III and IV, as they are each transporting four protons over the membrane in the process of electron transfer. The generated proton gradient is then exploited by the ATP synthase (or complex V) by which 3-4 protons (depending on the species) are transported back over the membrane, and the energy from this equilibrating step is used to convert Pi and ADP into ATP [29, 30].

1.1.2.4 Photosynthesis

Photosynthesis is the process by which plants and some bacteria assimilate carbon and generate energy [21, 31]. There are two major parts of the photosynthetic process, which commonly occurs in specialised compartments, the chloroplasts:

1. Energy generation by the light reactions; and

2. Carbon fixation by the Calvin cycle.

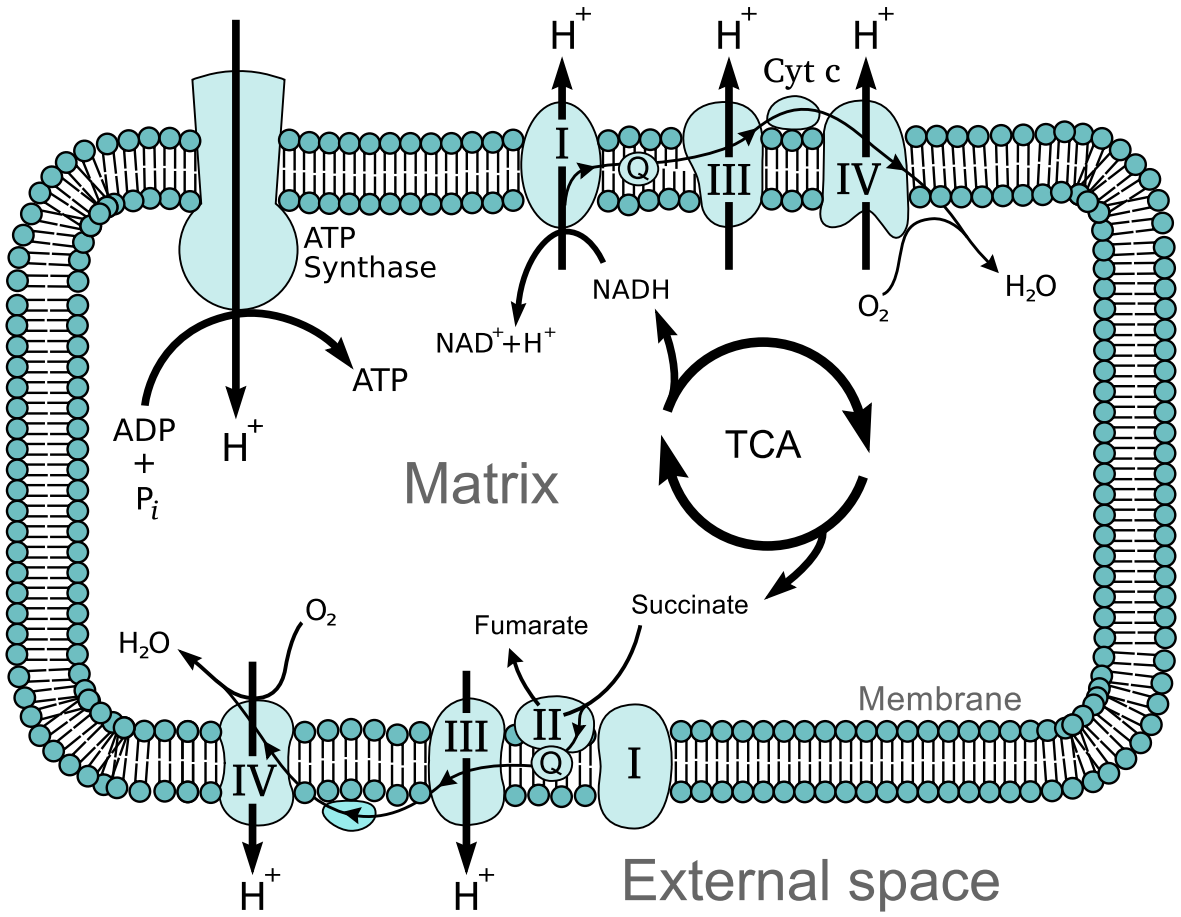


Figure 1.6: Electron transfer chain and ATP synthesis. There are two entry points into the electron transfer chain. Complex I uses NADH to reduce coenzyme Q and at the same time transports four protons over the membrane. The other entry point is complex II (succinate dehydrogenase) where the electrons are obtained from the reduction of succinate to fumarate. Complex III transfers the electrons further to cytochrome *c*, translocating an additional four protons from the matrix to the external space. The final transfer from cytochrome *c* to oxygen is performed by complex IV, again translocating four protons. The translocated protons can be employed by the ATP synthase to generate ATP. In prokaryotes, the matrix is the cytosol of the organism, while in eukaryotes it is the internal matrix of the mitochondrion. The external space is the inter-membrane space between mitochondrion and cytosol in eukaryotes and the periplasm space in prokaryotes, respectively. Image modified from public domain work by Fvasconcellos [28]

This section will cover the specifics of plant photosynthesis and illustrate the importance of this process in plant metabolism.

Light reactions Similar to the electron transfer in the mitochondrial electron transfer chain, the light reactions of photosynthesis generate a proton gradient by using the energy gradients between different types of electron carriers [32]. The membrane used for this electron transfer is the membrane of the thylakoid inside the chloroplast. It follows a four step process with electron transfer from water to plastoquinone, plastoquinol (the reduced form of plastoquinone) to plastocyanine, plastocyanine to ferredoxin and ferredoxin to NADPH (see Figure 1.7b). The first protein complex involved is photosystem II, which oxidises water to oxygen and uses light energy to increase the potential of the extracted electron (see Figure 1.7a). At the same time, it collects protons from the stroma (the inside of the chloroplast) and transfers them, along with the obtained electrons, to plastoquinone. Plastoquinol is then transferred to the cytochrome b_6f complex. This enzyme uses a two step process to transfer the electrons from plastoquinol to plastocyanine and translocate additional protons from the stroma to the lumen (the inside of the thylakoids). In the first step, two electrons from one plastoquinol are taken up and the remaining protons are released into the lumen ($PQH_2 \rightarrow PQ + 2\ H^+_{Lumen} + 2\ e^-$). One of these electrons is directly transferred to plastocyanine while the other is transferred to a plastoquinone ($PQ + e^- \rightarrow SPQ^-$), leading to the formation of a semiquinone. In the second step, another plastoquinol is oxidized, transferring two further protons into the lumen. The electrons are split again between plastocyanine and the earlier created semiquinone. In addition, two protons from the stroma are collected and transferred to the plastosemiquinone leading to the formation of plastoquinol, which can be used again in the first step ($PQH_2 + SPQ^- + 2\ H^+_{Stroma} \rightarrow 2\ H^+_{Stroma} + e^- + PQ + PQH_2$). Overall, four protons are released

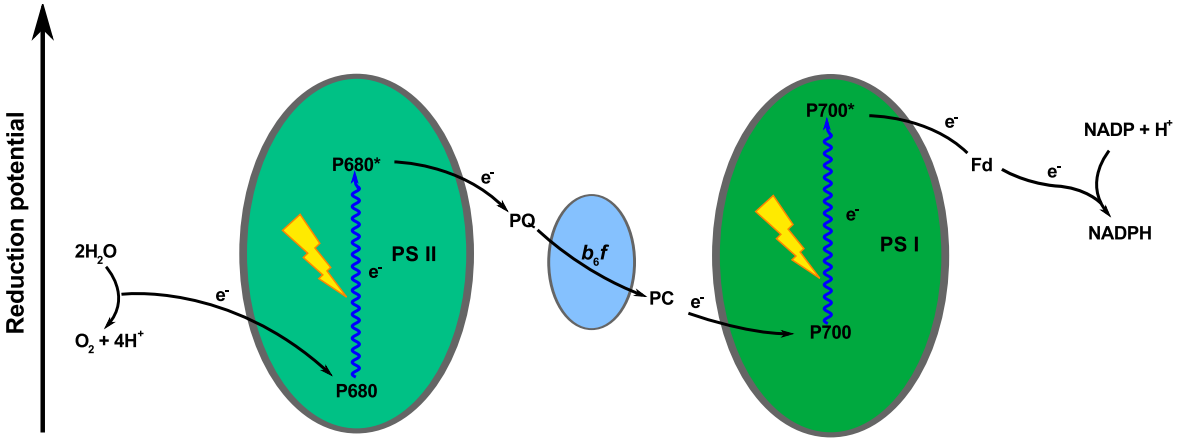
into the lumen and two are taken up from the stroma for each electron pair transferred. The third step is catalysed by photosystem I where light is used to increase the electrons' potential and allow the transfer of these electrons to a reductant with a higher reduction potential, from plastocyanine to ferredoxin (see Figure 1.7a). Finally, the electrons are transferred to NADPH by the ferredoxin-NADP reductase. Similarly to the mitochondrial electron transfer chain, the generated proton gradient is used to drive the chloroplast ATP synthase. The proton per ATP ratio for this ATP synthase is commonly assumed to be dictated by the protein structure as 14:3 [21, p. 151]. Instead of transfer to ferredoxin, electrons can also be transferred back from P700 in photosystem I to the cytochrome b_6f complex using plastoquinol as carrier. By this mechanism, additional protons are transferred over the membrane [21, p.147]. This leads to the possibility of regenerating ATP without any NADPH regeneration.

The Calvin cycle Photosynthesis is not only a ready source for energy, but also encompasses fixation of carbon dioxide. The most important enzyme in this process is the ribulosebisphosphate carboxylase oxygenase enzyme (RuBisCO), the most abundant protein in leaves with an abundance of up to 65% of the total soluble protein fraction [33]. RuBisCO catalyses the carbon-dioxide fixation step in the Calvin cycle by the following reaction:

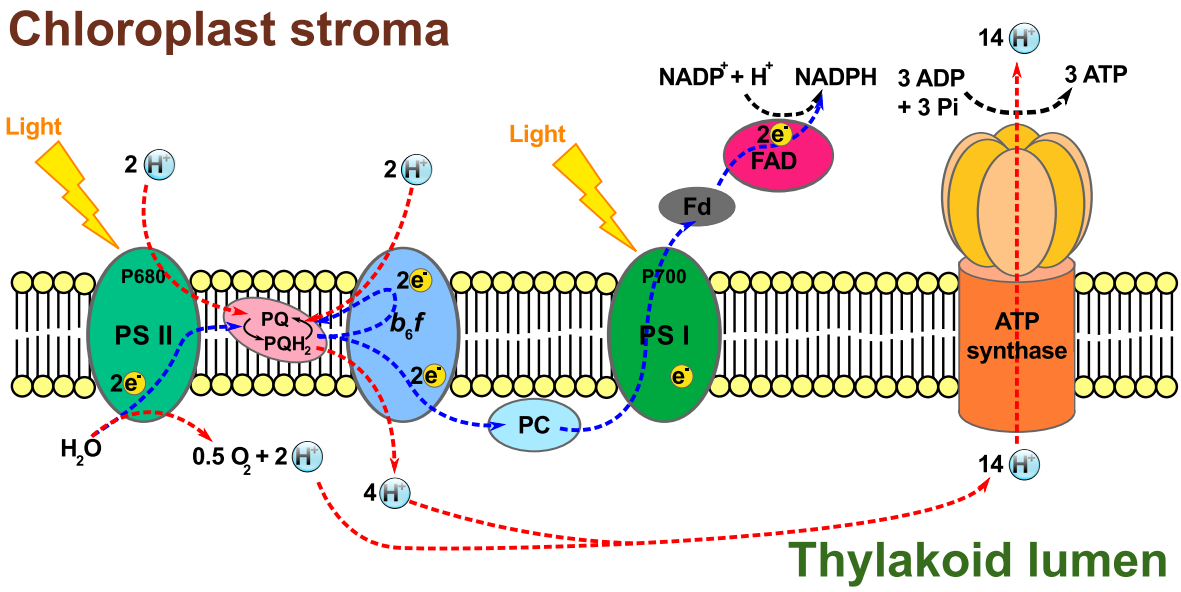
$$CO_2 + D\text{–}ribulose\ 1.5\ bisphosphate + H_2O \longrightarrow 2\ 3\text{–}phosphoglycerate \tag{1.3}$$

The Calvin cycle uses the product of RuBisCO (G3P) and converts it into its substrate ribulose 1.5-bisphosphate (R15BP). The pathway is displayed in Figure 1.8.

The initial reactions from 3-phosphoglycerate (G3P) are shared with the glycolytic pathway. In the Calvin cycle, however, these reactions are acting in the reverse di-



(a) Z-scheme



(b) Light reactions

Figure 1.7: Enzymes involved in photosynthesis and their action in the electron transfer chain. (a) Electron potentials are repeatedly increased by light to allow the transfer to more potent reductants. (b) Simultaneously, protons are transferred into the lumen and a proton gradient is built. The gradient is used by the ATP synthase to regenerate ATP. The final products of the light reactions are reductants in the form of NADPH and energy in the form of ATP . PS II: Photosystem II; PS I: Photosystem I; b_6f: Cytochrome b_6f; PQ(H_2): Plastoquinone/ol ; PC : Plastocyanine; Fd : Ferredoxin; FAD : Ferredoxin NADPH reductase.

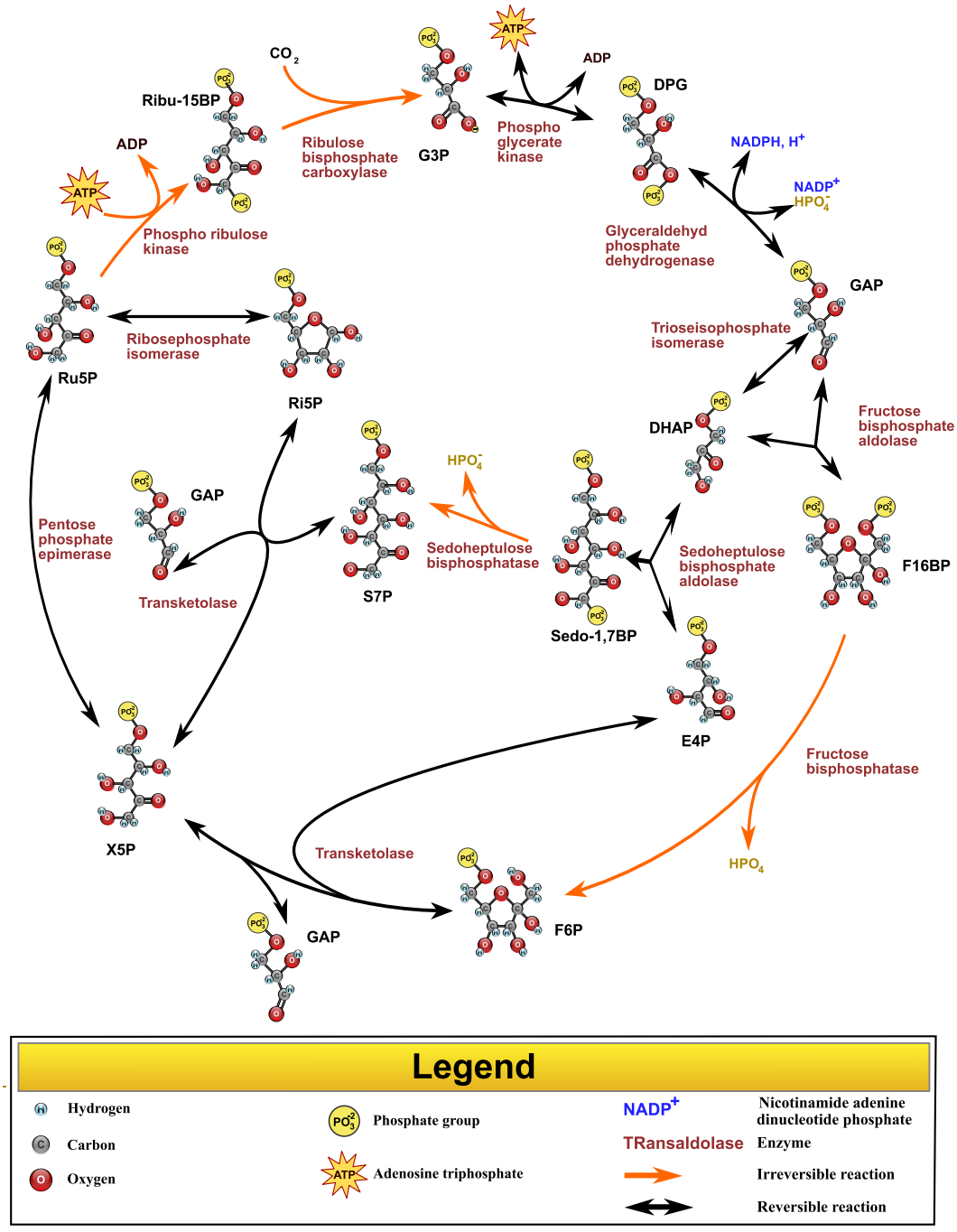


Figure 1.8: The Calvin cycle pathway assimilating carbon dioxide. The pathway consumes three ATP and two NADH per fixed carbon dioxide.
Legend and structures based on work from Y. Mrabet published under the Creative Commons License 3.0 BY-SA [22]

rection from G3P to GAP. At the expense of ATP, G3P is converted into DPG and reduced to obtain GAP, which forms a hub of the Calvin cycle. A total of five initial GAP molecules are necessary to drive the fixation of three carbon dioxide into one new G3P. Two GAP are converted into fructose 1,6-bisphophate (F16BP) by the reverse action of triosephosphate isomerase and fructose bisphosphate aldolase. F16BP is then dephosphorylated to provide fructose 6-phosphate (F6P). By the action of transketolase, F6P and an additional GAP are converted to xylose 5-phosphate (X5P) and erythrose 4-phosphate (E4P). The fourth GAP combines with E4P to form sedoheptulose 1,7-bisphosphate, which is subsequently dephosphorylated to sedoheptulose 7-phosphate (S7P). By a second reaction, catalysed by transketolase, S7P and the final GAP are converted into another molecule of X5P and one molecule of ribose 5-phosphate (Ri5P). A total of two X5P and one Ri5P are converted into ribulose 5-phosphate, which is phosphorylated to provide the original three R15BP which are needed to generate six G3P (and thereby fixing three carbon dioxide). The process costs a total of nine ATP and six NADH. This ratio of consumption is similar to the ratio of production provided by the light reactions. The overall reaction catalysed by the Calvin cycle is:

$$3\,CO_2 + 9\,ATP + 6\,NADPH + 5\,H_2O \longrightarrow GAP + 9\,ADP + 8\,Pi + 6\,NADP + 3\,H^+ \tag{1.4}$$

1.1.2.5 Pentose phosphate pathway

The final pathway presented here is the pentose phosphate pathway (PPP), displayed in Figure 1.9. It is one of the main sources for reductant in metabolism [34]. The substrate for the PPP is glucose 6-phosphate (G6P) or fructose 6-phosphate (F6P) from

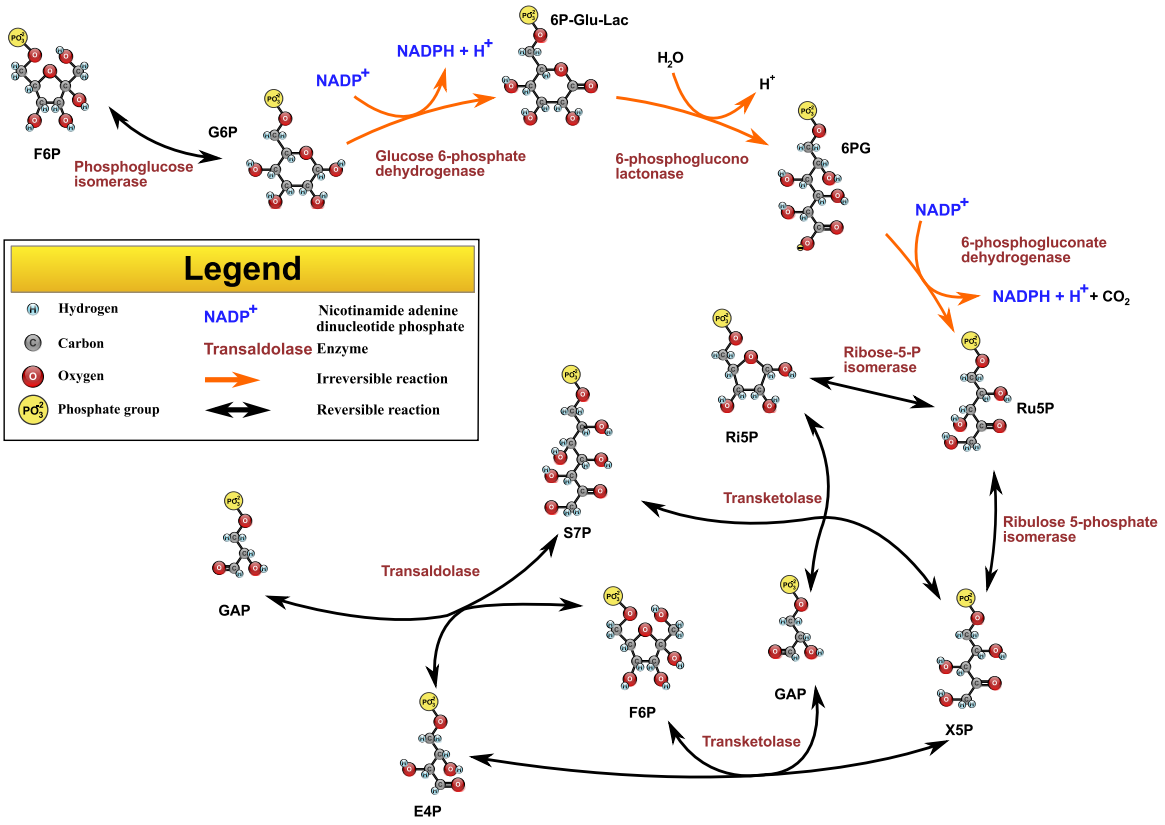


Figure 1.9: The pentose phosphate pathway is fuelled by fructose phosphate or glucose phosphate. It converts one fructose into three carbon dioxide plus one GAP while regenerating six NADPH. Legend and structures based on work from Y. Mrabet published under the Creative Commons License 3.0 BY-SA [22].

the initial glycolytic reactions. G6P is reduced and the resulting product hydrated to yield 6-phosphogluconate (6PG). 6PG in turn is dehydrogenated and decarboxylated to ribulose 5-phosphate (Ru5P), the first five-carbon sugar in the pathway. This is followed by two isomerisation reactions converting two molecules of Ru5P into one molecule of xylulose 5-phosphate and one molecule of ribose 5-phosphate. By the actions of transketolases and transaldolases (both transferring carbon fragments between the sugars) these 5 carbon sugars are converted into a 2:1 ratio of F6P and GAP. The

overall reaction of the pentose phosphate pathway is:

$$F6P \ + \ 6 \ NADP^{+} \ + \ 3 \ H_2O \ \longrightarrow \ GAP \ + \ 6 \ NADPH \ + \ 6 \ H^{+} \ + \ 3 \ CO_2 \quad (1.5)$$

Thus the pentose phosphate pathway is a fast route to obtain a maximum output of reductant.

1.2 *Medicago truncatula* and *Sinorhizobium meliloti*

For agriculture the legume-rhizobia symbiosis is likely to be the best way for a sustainable fertilisation of soils with nitrogen. With ever-increasing energy costs and the depletion of fossil fuel reservoirs, artificial nitrogen fertilizer production is likely to become infeasible in the future. Different legume species are already used for both forage crop production (alfalfa - *Medicago sativa*) or as a source of oilseed production, with over 50% of the world production coming from soy bean (*Glycine max*) [35]. Other species of the legume family include clover (*Trifolium* spp) and groundnut (*Arachis hypogaea*), which have a long history in crop rotation schemes, or vegetables like beans (*Phaeseolus vulgaris*) or peas (*Pisum sativum*). Models for the symbiotic interactions are therefore essential to gain a better understanding of the processes involved and allow a better focus of research in this field. Similar to *Arabidopsis thaliana* as model for plant growth and genetics [36, 37], *M. truncatula* has emerged as model for the legume rhizobial interaction [35]. The next sections will give an overview of this plant and its most important symbiont *S. meliloti*, and will explain the processes involved in forming the symbiosis. The basics of the metabolic interactions between both systems will be discussed and finally the properties of the most important enzyme of this symbiosis, the nitrogenase enzyme, will be explained.

Figure 1.10: *M. truncatula* shoot (left) and roots (right). The plants were grown in hydroponic solution on styrofoam rafts.

1.2.1 The plant host: *M. truncatula*

M. truncatula or barrel clover (see Figure 1.10), is native to the Mediterranean but is now used as a crop legume in many parts of the world, especially in Australia [38]. It is a cool season legume from the galegoid clade of the papilinoids, which also include other important legumes like alfalfa (*Medicago sativa*), clover (*Trifolium spp.*), or peas (*Pisum sativum*) [39]. *M. truncatula* was chosen as a model organism for the symbiosis due to its small diploid genome and its rapid generation time [35, 40, 41]. Even though the whole genome sequence of *M. truncatula* was only assembled recently [42], there is plenty of information available for the organism. It has been extensively studied in recent years, with detailed proteomics analysis (reviewed in Colditz and Braun [43]), extensive gene expression studies (reviewed in Young and Bharti [39]) and multiple metabolomic studies [44, 45, 46]. The Mt3.5v5 genome annotation lists

a total of 64123 genes with 66% of these genes being assigned to one of the eight chromosomes [47]. There are databases for gene expression profiles of different plant parts [48, 49], a database containing metabolic reactions [50], and many established experimental protocols for studies and growth [51]. With a large scale proteomic analysis, Watson et al. investigated the proteomic differences between several tissues of the plant [52]. They showed that the different tissues share 61% of all proteins, while only 39% of proteins are tissue-specific. However, this implies that a substantial amount of metabolic functions are similar over a vast fraction of the tissues, with differences most likely only in branches of secondary metabolism.

Beside its function as a model for the legume-rhizobial symbiosis, *M. truncatula* has also been used in studies on several other interactions. It is used as a plant partner for studies in the plant-mycorrhiza symbiosis [53, 54] and was suggested as a model for parasite-root interactions [51]. With this multitude of available information, it becomes ever more important to supplement this data with computational models of the underlying networks and analyse these computational models.

1.2.2 The rhizobial partner: *S. meliloti*

Rhizobia are gram negative soil bacteria which have the very rare ability to fix nitrogen [55]. However, they can only perform this task efficiently in symbiosis with plants from the legume family, due to the irreversible deactivation of the nitrogenase enzyme by oxygen [56]. While most rhizobia cannot create a low oxygen environment on their own, many plants from the Leguminosae family can provide this environment. Legumes, when interacting with rhizobia, form specialised organs called nodules. In these nodules the concentrations of free oxygen are extremely low [57]. More detailed descriptions of the plant nodule and the symbiotic process and interactions are provided in Section 1.2.4. *S. meliloti*, formerly known as *Rhizobium meliloti*, is a symbiont for *M. truncatula* and other legumes of the *Medicago, Melilotus* and *Trigonella* genera [58]. Its genome was sequenced in 2001 [59] and multiple studies on gene expression [60, 61, 62, 63] have been performed since, investigating its symbiotic and non symbiotic properties. Along with the genetic information available, proteomic studies [64, 65] and metabolic profiling [66, 67] provide a strong basis for an investigation of the metabolic interactions between this organism and its host. Indeed, a metabolic model for *S. meliloti* was recently published by Zhao et al. [68], and BioCyc contains databases for different biovariants of *S. meliloti* [69].

1.2.3 The nitrogenase enzyme

Nitrogenase is at the very heart of the symbiosis between legumes and rhizobia. While there are some rhizobia which can fix nitrogen in free living cultures, they are unable to grow on N_2 as the sole nitrogen source. In fact, the nitrogen fixation in free living cells occurs in specialised non-growing cells [70]. This is mainly due to the low-oxygen environment needed by nitrogenase, which is inconsistent with the necessary energy requirement for growth (requiring oxygen for the oxidative phosphorylation). However, the environment provided by legume nodules allows for growth and nitrogen fixation by nitrogenase. In rhizobia, nitrogenase is encoded by the genes *nifH*, *nifD* and *nifK*, which together form the subunits of the enzyme [71]. Other genes from the *nif* family are necessary for the biosynthesis of important co-factors for the protein [72]. The expression of these genes is under tight regulatory control, tied to the oxygen concentration. They are expressed when the oxygen concentration is low and repressed when too much oxygen is present [72]. This tight control makes it even more important for the plant to provide an environment that allows the expression of the *nif* genes, as otherwise no nitrogen fixation would occur.

1.2.4 The legume-rhizobia symbiosis

The symbiotic process between rhizobia and legumes is probably one of the best investigated symbioses, with publications dating back to 1888 [73, 74, 75]. It is the focus of extensive research due to its importance in biological nitrogen supply, and its long history of application in agriculture. This section will explain the processes involved in the formation of the legume-rhizobia symbiosis and detail available information about the metabolic processes and the exchange of metabolites during the symbiosis.

1.2.4.1 Formation of the symbiosis

The legume-rhizobia symbiosis is formed by a mutual effort of plant and bacterium, shown in Figure 1.11. During the process a specialised plant organ, the nodule, is formed. There are two different, though similar, types of nodules: determinate and indeterminate. The type of nodule formed depends on the involved species; *M. truncatula* and *S. meliloti* form indeterminate nodules [41] and therefore this section will concentrate on these. The reader is referred to Ferguson et al. [76] for more information on the differences between indeterminate and determinate nodules and the specifics of determinate nodules.

The symbiosis starts with a complex exchange of signalling molecules (step (1); Figure 1.11). The signals exchanged are specific to the involved plant and rhizobium species, and this is the reason for the specificity of rhizobial species for certain plant species [77]. The initial signal is the release of flavonoids by the plant, which act on rhizobia in multiple ways. They attract the rhizobia, induce the production of nodulation genes (nod) and enhance their growth [78]. It has been shown that the attractive influence of specific flavonoids can be lost when nod genes are mutated [79]. Finally, they induce the production of nodulation factors (NF) by the rhizobium. The

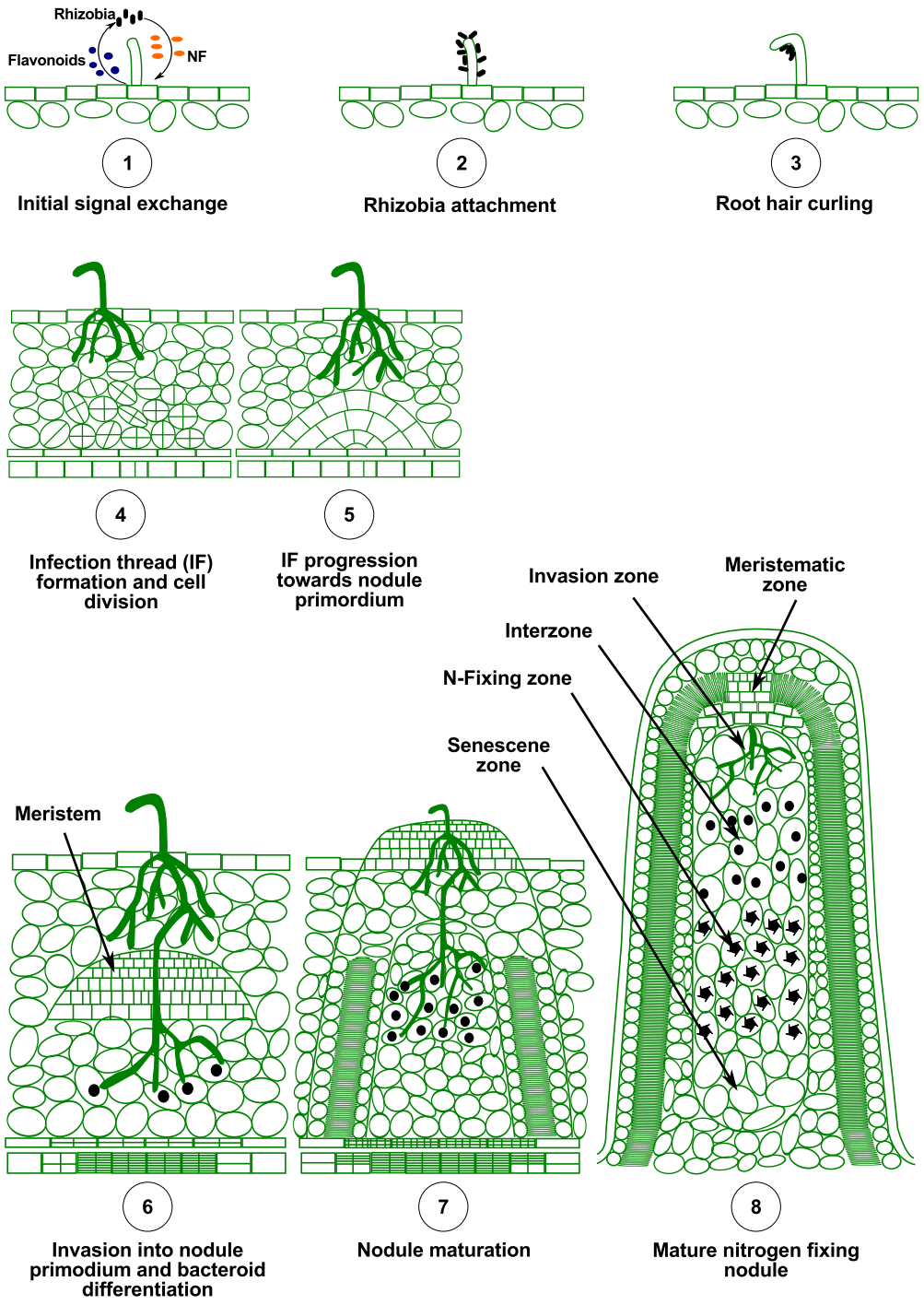


Figure 1.11: The process of plant invasion and nodule formation in indeterminate nodules. Flavonoids from the plant attract rhizobia, while nodulation factors from the rhizobia induce root hair curling. An infection thread is formed and the plant root is invaded. Simultaneously, the plant starts growing at the nodulation site creating the nodule organ. The nodules keep growing while nitrogen is fixed in the nitrogen-fixing zone. Figure based on Ferguson et al. [76] with permission from John Wiley and Sons license number (3199421180965)

NF are lipo-chito-oligosaccharides [80] and are specific for the host plant [81]. Even though rhizobia can enter the plant root through cracks in the epidermal tissue, the most common route of infection is through root hairs [76]. In reaction to the attractive flavonoid influence, rhizobia attach to root hairs (step 2; Figure 1.11). The influence of NFs leads to root hair curling (step 3; Figure 1.11) which encloses bacteria [82]. NFs also induce preparatory cell divisions in the inner cortical cells which finally lead to the formation of the nodule primordium [83, 84]. Whether by crack invasion or by trapping in root hairs, once the rhizobia are surrounded by plant cell walls, they are transported and replicate in an inward growing infection thread (IT) (reviewed in Brewin [85] and shown in step 4; Figure 1.11). It is likely that the forming primordium is stimulated by the increasing levels of NFs produced by the invading rhizobia and finally forms a meristem growing towards the IF [76]. The growth of the IF is driven by the replication of rhizobia [86]. At this point the plant seems to take control, as rhizobia are imported into cells by an endocytosis-like mechanism (step 6 and 7 in Figure 1.11) [87].

In the investigated partners *S. meliloti* and *M. truncatula*, the rhizobia in these vesicles (or symbiososmes) terminally differentiate [88]. This seems to be predominant in the galegoid glade but is not observed in e.g. the *Rhizobium leguminosarum-Phaseolus vulgaris* or *Mesorhizobium loti-Lotus japonicus* symbioses [88]. The nodule itself keeps growing and can be divided into five zones (step 8; Figure 1.11):

1. The meristematic zone
2. The infection zone
3. The interzone
4. The nitrogen fixing zones
5. The senescence zone

The meristematic zone is the zone of growth in nodules. Plant cells constantly divide providing new cells for invasion. The infection zone is the place where rhizobia are taken up by plant cells. As the nodule keeps growing the infection thread network has to follow the new plant cells for infection [89]. In the interzone the rhizobia can replicate a few times and mature into bacteroids (4-7 times larger than normal bacteria [90]). There are different types of bacteroids depending on the host-rhizobium combination and a detailed analysis can be found in Haag et al. [91]. Bacteroids are surrounded by a plant membrane, the peribacteroid membrane, which regulates the exchange of metabolites and thereby plays an essential role in the symbiosis [92, 93]. The nodules provide the conditions essential for nitrogen fixation:

- A low oxygen concentration
- Enough oxygen for energy production by oxidative phosphorylation

These are contradictory but necessary conditions, as the nitrogenase enzyme is highly susceptible to oxygen [91]. At the same time the nitrogen fixing process is very energy demanding [2, p. 1360ff]. As seen in Section 1.1.2.3, the main source of energy comes from the electron transfer chain, which uses oxygen as a final electron acceptor. Thus, while oxygen is damaging the nitrogenase, it is also needed to allow a sufficient energy supply to drive its function. In nodules, this contradiction is solved by using leghemoglobin [94]. Like hemoglobin, these proteins can bind oxygen and are used as means to minimise free oxygen. Simultaneously, they allow its delivery to the relevant enzymes of the electron transfer chain and have been shown to be crucial for symbiotic nitrogen fixation [95]

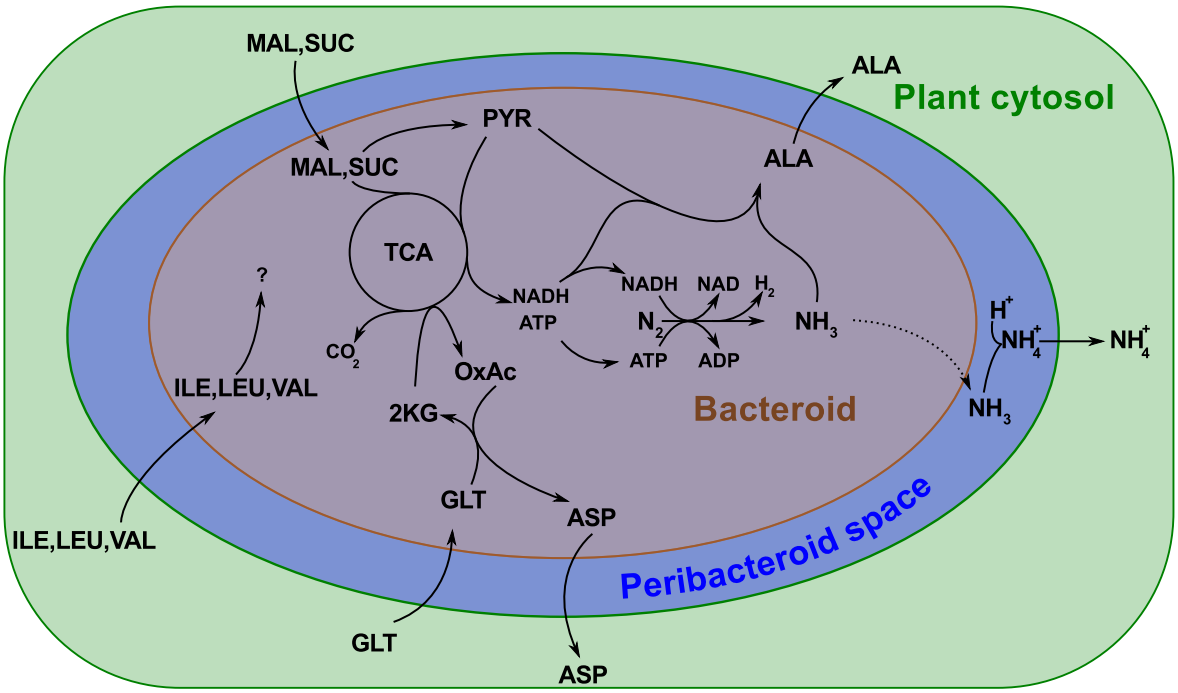


Figure 1.12: Metabolic exchange between bacteroid and plant. Dicarboxylic acids are supplied as main carbon source to drive nitrogen fixation. Ammonia or amino acids are the main sources of fixed nitrogen supplied to the plant by the bacteroid. Amino acid cycling (e.g. supply of glutamate by the plant and export of aspartate from the bacteroid) can provide necessary precursor metabolites and carbon to the rhizobium.

1.2.4.2 Metabolic exchange in the legume-rhizobia symbiosis

Multiple models for metabolite exchange between rhizobia and legumes during symbiosis have been proposed and Figure 1.12 displays the most important suggested exchanges. These metabolic exchanges were extensively reviewed in Udvardi and Day [92], Prell and Poole [96] and Udvardi and Poole [93].

The supply of dicarboxylic acids is commonly accepted and its essentiality was proven in *S. meliloti* [97] and *R. leguminosarum* [98]. As the corresponding transporter (the DCT - dicarboxylic acid transporter) has high affinities to malate, succinate and fumarate [92] it is not clear whether all of them or only a single dicarboxylic acid

is transported. The dicarboxylic acids commonly assumed to be most important are malate [93, 99, 100] or succinate [101, 102].

The form in which nitrogen is supplied to the plant is still disputed. The most common assumption is that ammonia diffuses through the bacteroid membrane into the peribacteroid space, where it is protonated to ammonium and then transported into the plant cytosol [93, 103]. This is mostly due to the evidence that ammonia assimilation by the glutamine synthase (GS) / glutamate synthase (GOGAT) system is inhibited by high ammonia concentrations [92, 104]. It has been shown in free living *Rhizobium etli* that removing the GS/GOGAT activity leads to increased export of ammonia from the cells [105], which was suggested as a possible approach of improving symbiotic efficiency. There is evidence that bacteroids still assimilate nitrogen into different amino acids [99], which contradicts the paradigm that bacteroids only reduce dinitrogen to ammonia and export it without assimilation [106]. It was even suggested that alanine, not ammonia, is the sole nitrogen compound exported from bacteroids [107], which is also disputed [108, 109]. However, alanine export is observed even though the aldehyde dehydrogenase for *de novo* alanine biosynthesis is not essential [109].

Along with alanine export, amino acid cycling between plant and rhizobium was suggested to be essential for nitrogen fixation [100]. It was supposed that glutamate is supplied as both nitrogen donor and carbon source, in addition to malate, and in return some of the glutamate nitrogen is exported as aspartate [100]. Following these initial amino acid transport studies, Prell et al. [110] showed that in some rhizobia the transport of branched chain amino acids (like leucin, valine or isoleucine) from plant to symbiont is essential for nitrogen fixation. They also showed that this is not evident in a symbiosis between *S. meliloti* and alfalfa [111]. However, there is evidence that leucine can be provided by the plant in *S. meliloti*, while valine and isoleucine are not

provided and have to be synthesised by the rhizobium [112].

1.3 Computational background

The growing amount of experimental biological data leads to a necessity for computational methods to analyse and evaluate this information [113, 114]. While for a long time the focus of metabolic models has been in understanding small systems, pathways, or even single enzymes, the emergence of multiple "omics" techniques has shifted this to a more comprehensive view of biological systems [115]. Multiple genome-scale metabolic models have been published over the last few decades, including model organisms such as *Escherichia coli* [116, 117, 118, 119] and *Arabidopsis thaliana* [120, 121, 122], or even the highly complex network of *Homo sapiens* [123, 124]. This section details the process of metabolic reconstruction and highlights available methods aiding in this process. It deals with the analysis of the reconstructed models and gives an overview of the methods employed beginning with the fundamental concepts of metabolic modelling and stoichiometric analysis methods in Section 1.3.2. Section 1.3.3 will then detail the applications of flux balance analysis (FBA) and provide an overview of the capabilities of these kinds of methods.

1.3.1 Metabolic network reconstruction

The quality of any result of a metabolic analysis is restricted by the quality of the underlying network. Network reconstruction is therefore a crucial step and for a new reconstruction it might take years to assemble and curate all available data. The increasing availability and quality of genome sequences and annotations for multiple organisms allows the reconstruction of genome scale metabolic models. Thiele and Palsson [125] published a protocol for network reconstruction in which they estimate the total time necessary for the reconstruction of a network to be between one and 18

month.

Metabolic models for flux balance analysis are generally based on finding appropriate reaction stoichiometries for annotated enzymatic activities of the target organism. This can be achieved manually by looking up reaction stoichiometries for specific enzymatic activities in databases like KEGG [126, 127], MetaCyc [128, 69] or BRENDA [129, 130]. Another approach is an initial automated reconstruction step using tools like PathoLogic from the PathwayTools program suite [131] or metaSHARK [132]. In both instances the initial reconstruction is a list of reactions which often represent disconnected sub-networks (e.g. due to missing annotations or spontaneous reactions). Multiple issues arise at this point:

- It has to be ascertained that all of the included reactions really belong to the metabolic capabilities of the organism;
- Gaps in the network have to be filled;
- Potential metabolic capabilities have to be considered; and
- The thermodynamic consistency of the model needs to be verified.

Automatic reconstruction often leads to the inclusion of reactions which are unlikely to be present in the investigated organism. This stems from the issue that many enzymes are only annotated with their general function (mostly due to lack of information on their specific substrates). Unfortunately, databases often contain enzyme links which point to such general annotations (like dehydrogenase or kinase). Automatic reconstruction therefore often leads to the inclusion of many reactions which have weak evidence and need to be manually removed prior to further curation. While manual

curation can avoid the inclusion of these false positives, it is easy to miss important reactions when selecting reactions using the available annotation.

To fill the gaps left after this step, multiple automated methods have been proposed. Automated reconstructions (like PathoLogic) often use pathway information to infer additional reactions (e.g. if a majority of a pathway is present, the remaining reactions from that pathway are included). Kharchenko et al. [133] proposed a method using gene expression data to find candidate genes for metabolic functions, which were missing in the network. Other approaches use the available multi-organism databases like KEGG and MetaCyc to infer possible missing reactions on a network basis [134, 135]. However the way this information is retrieved differs significantly. Kumar et al. [134] first try to relax the network by changing reversibility constraints to allow production of all metabolites. If this is insufficient, reactions from MetaCyc are added. This selection of reactions is achieved by solving a mixed integer linear program. The program aims at finding the smallest possible set of additional reactions which allows the production of all metabolites.

In contrast, Christian et al. [135] use a network-based approach called Network Extension. They define a set of products as target and a set of nutrients as seed. The initial network gets embedded in a reference network (e.g. all reactions in KEGG or MetaCyc) and the algorithm iteratively removes reactions from the reference network, ensuring that all compounds in the target are still producible. As the resulting network is not unique, the method allows for the incorporation of genomic data to determine the order of reaction removal (with more likely reactions being removed later on). The selection of metabolites for the target necessitates the incorporation of additional metabolomic data. A first step in obtaining this data is the determination of a biomass composition.

This determination often yields results including multiple macromolecular species, like cellulose, hemicellulose or protein. While determination of these macromolecular quantities is rather quick, they can not directly be incorporated into metabolic models. Instead, they have to be split into the contained monomers (sugars, amino acids etc.) as the internal variability would otherwise lead to an explosion in the number of metabolites. Proteins have to be split into their amino acid composition and the monomer composition of cell wall sugars needs to be assessed.

The formulation of a biomass composition often only reflects major metabolites and some important cofactors, while other compounds are neglected. However, if a compound is found in a metabolomic study, and is not a xenobiotic, the system needs to be able to biosynthesise this compound. This issue has been recognised and methods to account for this fact have been proposed [136, 137]. For curation this means that additional reactions need to be found in the literature to allow the production of these metabolites.

Another issue often arising in models is thermodynamic inconsistency. While in principle all reactions are reversible using the right concentrations of products and substrates (potentially with the exception of the light harvesting reactions in photosynthesis), many reactions need to be considered as irreversible due to thermodynamic constraints. The physiological reversibility however is not obvious and many databases have either no or contradicting information on the reversibility of a specific reaction.

When assuming reversibility as the default mode for any reaction, thermodynamically infeasible cycles can easily occur, which lead to the production of energy (see Figure 1.13). There are also other cycles which do not create energy but can carry flux without any net conversion, called futile cycles. Futile cycles are thermodynamically infeasible [138], as action is happening without energy being spent, and they should be

(a) (b)

Figure 1.13: The two reactions are catalysed by phosphofructokinase (F6P → F16BP) and fructose 1,6-bisphosphatase (F16BP → F6P). If no reversibility constraints are assumed these two reactions can regenerate ATP without any further input by this cycle as in (a). However, thermodynamics tells us, that both reactions are irreversible and therefore only allow the consumption of energy as in (b).

eliminated. There are however instances of futile cycles, especially in large networks, which can not be initially eliminated without restricting the overall function of the model. Those cycles therefore have to be considered in calculations.

Price et al. [139] proposed a method to restrict the maximal and minimal fluxes in reactions involved in futile cycles. They suggested restricting one of the reactions in the cycle to zero and determining the maximal and minimal fluxes the other reactions of the cycle can carry. These maxima and minima can then be used as constraints to the respective fluxes, allowing the minimisation of the internal activity of these cycles without inhibiting other activities of the network.

For prokaryotic systems a model that has reached this stage of curation can be used for an initial analysis. For eukaryotic models however, another issue needs to be addressed. Eukaryotic cells exhibit a highly organised structure with several compartments that split the metabolism into several distinct networks (see Figure 1.1). Information about sub-cellular localisation of specific enzymes has to be retrieved as compartmentalisation can have a high impact in a metabolic reconstruction. Multiple

methods have been proposed to address this challenge and to achieve the distribution of reactions. Protein localisation information can often be found in databases like SUBA [140] for *A. thaliana* or LOCATE [141] for mammalian organisms. If a database for the reconstructed organism exists it can directly be used to assign localisation information. Otherwise, localisation of homologous enzymes from other organisms in the databases can be used for inference. This, along with organism-specific proteomic studies, gives rise to an initial compartmentalisation of the reactions. However, either the localisation of many proteins is still unknown, or they are not expressed in high enough quantities to be detected in an untargeted study. Prediction algorithms like SherLoc [142] or PSort [143] allow the assignment of localisation for proteins for which only their sequence data is known. This process can also be automatically performed using tools like the Subliminal Toolbox [144], even though predictions can be misleading.

There are reactions which either have no enzyme association or for which no subcellular localisation can be obtained after application of predictive tools. One way to distribute those reactions was developed by Mintz-Oron et al. [145]. They proposed a network-based approach, which incorporates data from different sources and minimises the necessary amount of metabolite transport while activating all localised reactions. The approach was successfully used for *A. thaliana*, for which explicit localisation databases like SUBA exist, and which is undoubtly the presently best studied plant. While the extent of available data for *A. thaliana* allows the successful application of this approach, the amount of available data for most other organisms is comparatively restricted, therefore manual curation or other methods are necessary for compartmentalisation.

1.3.2 Network analysis

As detailed kinetic information for many enzymes is still scarce, algorithms to investigate the stoichiometric properties of metabolic networks are of high importance. The stoichiometric properties of any metabolic network can easily be captured by the $m \times n$ stoichiometric matrix S. The rows in S represent the m metabolites of the network while its columns represent the n reactions present in the network. Each entry $S_{i,j}$ represents the stoichiometric coefficient of metabolite i in reaction j. A simple network and its stoichiometric matrix can be seen in Figure 1.14.

Using this matrix the following equation completely defines the metabolism of the organism:

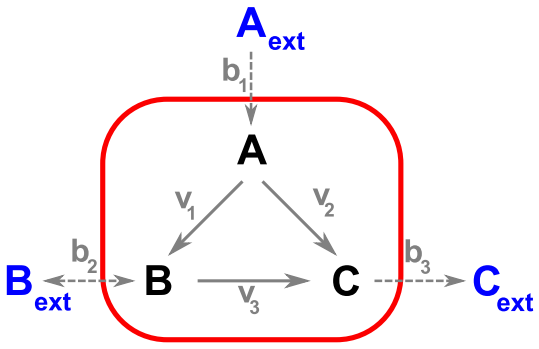
$$S \cdot \vec{v} = \frac{dC}{dt} \tag{1.6}$$

where $\vec{v}$ is the vector of fluxes through the reactions of the network and $\frac{dC}{dt}$ is the change in concentrations of each metabolite over time. However, the change of concentrations is highly dependent on the kinetics of the reactions, a piece of information missing for many enzymes. The "quasi-steady-state assumption" bridges this gap and allows the investigation of this system without information about enzyme kinetics [146]. The assumption is made that the internal system is in a steady state and that the concentrations of all internal metabolites (black in Figure 1.14) do not change. To reflect this steady-state assumption in our model the stoichiometric matrix has to be split into two parts:

- The $m_{int} \times n$ internal matrix N (the lower part of the stoichiometric matrix)
- The boundary $m_{ext} \times n$ matrix B (the upper part of the stoichiometric matrix)

	b_1	b_2	b_3	v_1	v_2	v_3
A_{ext}	-1	0	0	0	0	0
B_{ext}	0	-1	0	0	0	0
C_{ext}	0	0	-1	0	0	0
A	1	0	0	-1	-1	0
B	0	1	0	1	0	-1
C	0	0	1	0	1	1

(a) Stoichiometric matrix



(b) Network

Figure 1.14: A simple stoichiometric network and its stoichiometric matrix. b_{1-3} are the boundary fluxes and v_{1-3} are the internal fluxes. The internal stoichiometric matrix N is the lower half of the matrix displayed in (a). Internal metabolites are displayed in black, while external metabolites are depicted in blue.

This allows the reduction of Equation 1.6 into a system of linear equations:

$$N \cdot \vec{v} = 0 \tag{1.7}$$

where all internal metabolite concentrations have to be constant. The internal stoichiometric matrix N can then be investigated to gain insight into internal processes.

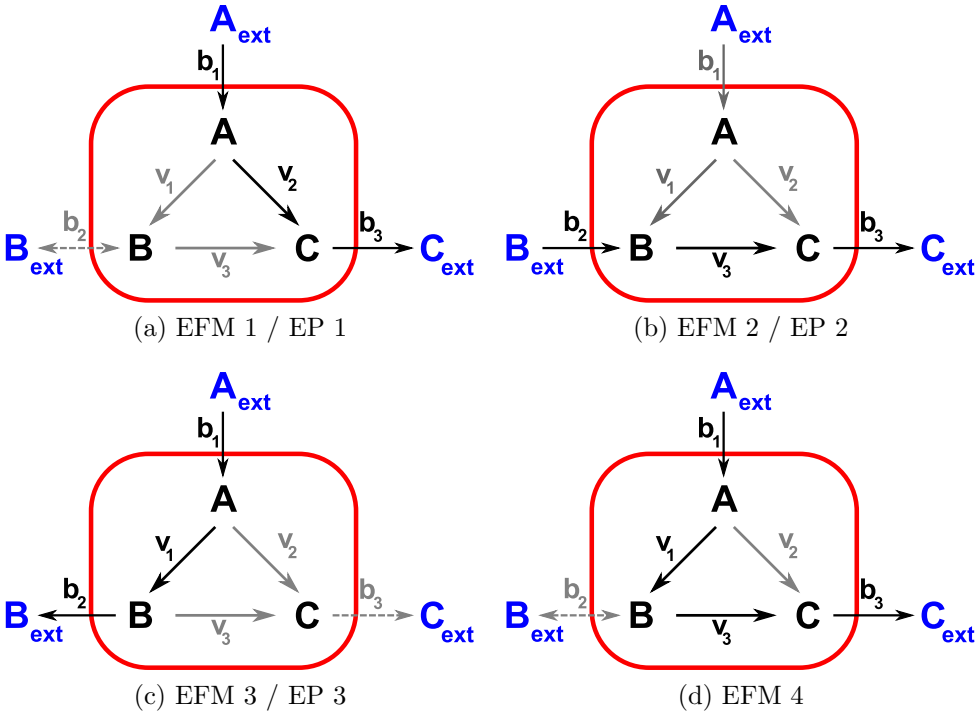
The solutions of a linear system as in Equation 1.7 can be directly calculated using Gaussian elimination [147]. A system of linear equations has either no solutions, exactly one solution or infinitely many solutions. If there are no solutions, there are equations which impose contradicting constraints to the system rendering it unsolvable. It is then called inconsistent. If $m = n$ and all row vectors in S are linearly independent (i.e. there is no row which is a linear combination of the other rows) then the system has exactly one solution [147]. If there are less linearly independent equations than variables, and the system is not inconsistent, the system has infinitely many solutions.

While solving a general linear equation system is straightforward, analysis of the

solution can be rather complicated. The most powerful method of direct analysis of the set of solutions is by investigating its null spaces [148]. The null spaces are the spaces that contain all solutions for $\vec{v}$ in Equation 1.7 (the right null space) and to the equation $\vec{w} \cdot S = 0$ (the left null space). The former contains all viable flux distributions that fulfil the steady-state assumption (also called the flux space), while the latter contains all pools of conserved metabolites (i.e. groups of metabolites for which the sum of concentrations is constant) [149, p. 154ff]. However, as a space can commonly be described in an infinite number of ways, just calculating one set of basis vectors for the null spaces is insufficient for the analysis of the generating network. Driven by the idea of metabolic pathways, the approaches to investigate these null spaces therefore focus on finding pathways or subnets within the null spaces.

Three alternative but equally useful methods were used for this purpose: elementary flux modes (EFM) [150, 151, 152, 153, 154, 155], extreme pathways (EP) [156, 157, 149] and metabolic cut sets (MCS) [158]. An EFM is a flux distribution which fulfils the steady-state condition (1.7) and is minimal in the sense that no reaction carrying a flux can be removed without violating this condition. Furthermore, the elementarity refers to the property that no EFM may contain another EFM [114]. EPs further restrict this requirement making the set of EPs a convex basis set. Thus, if there are reversible fluxes the set of EPs can be smaller than the set of EFMs [149]. An example of EPs and EFMs of the network in Figure 1.14b is provided in Figure 1.15.

A third way to characterise the null space is using minimal cut sets (MCS). They are minimal sets of reactions that need to be removed to force a given reaction to be unable to carry flux. They are a dual representation of EFMs [159], i.e. the set of MCS can be determined directly from the set of EFMs and vice versa. MCSs, EFMs, and EPs suffer from the problem that their number grows rapidly with the size of the



(a) EFM 1 / EP 1

(b) EFM 2 / EP 2

(c) EFM 3 / EP 3

(d) EFM 4

Figure 1.15: Extreme pathways and elementary flux modes of the network in Figure 1.14b. Active fluxes are marked in black. Inactive fluxes are marked in grey. The fourth EFM is minimal in the EFM sense but since it is a linear combination of EP 2 and EP 3 and therefore not linearly independent it is not an EP.

investigated network [160]. This makes it difficult to calculate all possible EFMs or EPs for large networks [161, 152, 162]. There were attempts by de Figueiredo et al. to overcome this restriction by only calculating a subset of "k-shortest" EFMs [163]. Their method uses a mixed integer linear program to compute solutions with a minimal number of reactions under the constraint that they do not contain any solution found in a previous iteration. By the definition of EFMs the solutions are again EFMs, as they are minimal (the objective is to minimise the amount of active fluxes) and do not contain other, smaller, EFMs (no reaction can be removed to obtain another EFM).

1.3.3 Linear programming and constraint-based modelling

Elementary modes and extreme pathways are a useful tool to investigate the structure of a network and to find important links between reactions and metabolites. Their potential to find internal cycles (i.e. modes which do not have any boundary reactions) is powerful for network reconstruction and curation. The limitation of their application is due to the size of the network, with large networks containing too many EFMs and EPs. With thousands of EPs and EFMs in a genome-scale metabolic network, their ability to provide structural information diminishes.

Flux balance analysis (FBA), also called constraint-based modelling (CBM), bridges this obstacle by using linear programming to gain further insights into the flux distribution $\vec{v}$ of the steady state solutions [164]. Linear programming is a technique that solves a system of linear equations while optimising a given objective function [165]. The steady-state equation (Equation 1.7) is such a system of linear equations.

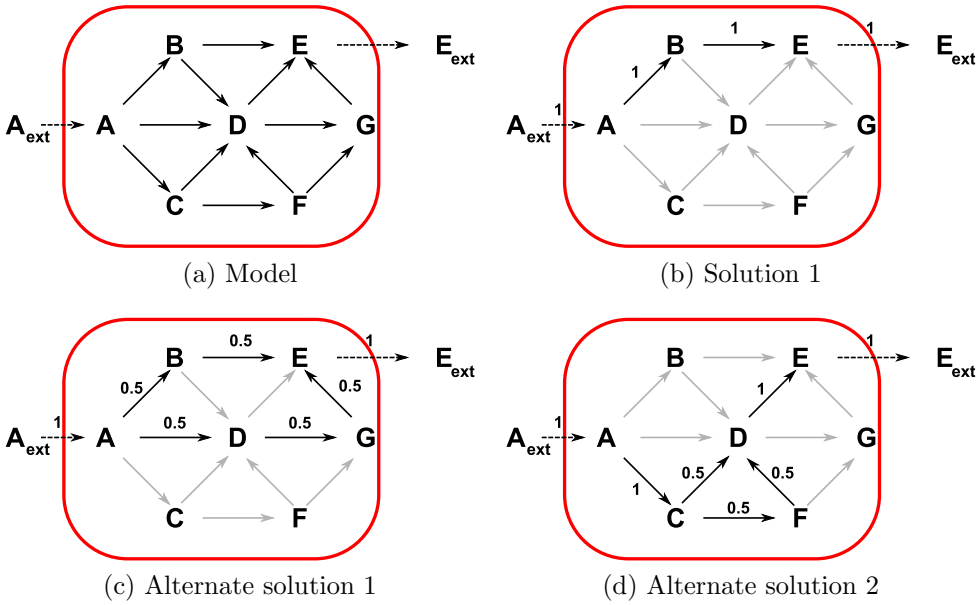
For a metabolic network, the number of solutions is commonly infinite, as there are usually more reactions (the variables in columns) than equations (the metabolites in rows). Linear programming allows the introduction of constraints on any flux in the network which offers the possibility to incorporate experimental data. Biomass composition and growth rates allow the restriction of the exporting boundary reaction fluxes (exporters) while measured uptake rates allow the limitation of importing boundary

fluxes (importers). With these constraints the problem can be written as:

$$\begin{aligned} optimise : & \sum_{i=0}^{n} c_i \cdot v_i \\ & s.t. \\ & N \cdot \vec{v} = 0 \\ & a_i \leq v_i \leq b_i \ ; \ \forall i \ \in \ \{1 \dots n\} \end{aligned} \tag{1.8}$$

where c_i is the objective coefficient for reaction flux v_i, while a_i and b_i are the lower and upper bounds for flux v_i, respectively. N is the internal stoichiometric matrix. Linear programming optimises the fluxes towards a given objective $\sum_{i=0}^{n} c_i \cdot v_i$, which can be either a single flux (only one $c_i \neq 0$) or a more complex combination of fluxes. The choice of this objective function is crucial for the results obtained by the optimisation. Numerous different objective functions have been used and a review has been published by Schuetz et al. [166]. The most common objective function is biomass optimisation (reviewed by Feist and Palsson [167]), which assumes that the metabolic objective of the investigated organism is growth, and has been used successfully for *E. coli* [116, 168] and other microorganisms [169, 170]. The limitation of this objective is the complexity of the investigated organism. While most prokaryotic and single-cell eukaryotic organisms tend to optimise for growth, higher organisms, like plants or mammals, have different cell types which fulfil very specialised functions. An example in plants is the difference between a young and an old leaf, the latter (not growing any more) assimilating much of the carbon the former needs for growth.

Flux minimisation is an objective that tries to address this issue on a system-wide level by imposing a fixed production constraint, based either on experimental growth



(a) Model (b) Solution 1

(c) Alternate solution 1 (d) Alternate solution 2

Figure 1.16: Alternative optima pose an issue in FBA. When optimising the network in (a) for production of metabolite E with a restriction on the influx of metabolite A of 1, a solver might give the solution displayed in (b). However, many other solutions are also feasible and optimal for production of E. Two exemplary alternate solutions are shown in (c) and (d). These alternative solutions have to be investigated and recognised.

rates or as an arbitrary biomass unit, while minimising the total amount of flux needed to achieve this objective. The rationale is that minimisation of total flux will allow the organism to save energy for protein biosynthesis and maintenance as less enzymes are needed to achieve the imposed constraints. This method was successfully applied to models of *A. thaliana* [120], *Oryza sativa* [171] and *Saccharomyces cerevisiae* [172].

One general problem when analysing FBA results is the variability of optimal solutions (see Figure 1.16). Even optimising for a given objective still allows the network to reach the optima by various alternative solutions [173, 174, 175]. Several methods have been proposed to assess these alternate optima [173, 175]. Flux variability anal-

ysis provides a method to assess the flexibility of the optimal solution space without assessing all possible solutions [174, 176]. This allows the modeller to find reactions which are inflexible and can potentially be bottlenecks, providing a target for genetic engineering.

Finding genetic engineering strategies has also prompted the development of other methods like OptKnock [177]. OptKnock tries to determine reactions that need to be deleted from the model (i.e. knocked out in the organism) to improve the production of a desired by-product. The approach uses a bi-level mixed integer linear program (MILP), optimising the production of a given chemical compound under the condition of optimal growth for each possible knock-out.

The assumption of biomass optimisation in a perturbed system has been challenged by Tepper and Shlomi who proposed a method to search for all knock-outs showing a production of the target molecule in all possible solutions [178]. While these methods work under the assumption that the organism will adapt quickly to the knock-outs, other methods, like minimisation of metabolic adjustment (MOMA), minimise the metabolic change necessary to adapt to a gene knock-out [179]. MOMA employs a quadratic programming approach which minimises the Euclidean distance between the unperturbed flux distribution, and the flux distribution with the knock-out in question. Regulatory ON/OFF minimisation (ROOM) simplifies the problem solved by MOMA to a MILP by only considering significantly changing fluxes [180]. The rationale is that while the system will try to adapt, it will do so by a few significant changes in comparison to the more distributed adaptation suggested by MOMA. These methods show the power of flux balance analysis in genetic engineering and provide tools for modellers when investigating metabolic models.

Besides metabolic engineering, another field of application for FBA is to assess

which compounds an organism is capable of producing. It is possible to do this by null-space analysis, e.g. by adding an exporter for each metabolite, and testing whether there are basis vectors containing this exporting flux. There can however be additional compounds in the biosynthesis route that are needed to produce the compound, and which are not in fact producible. An attempt to address this problem is the calculation of the scope of a network [181]. The scope of a network is the set of all metabolites, that the network can produce from a given set of nutrients. The calculation process starts by selection of a set of nutrients (the seed). In an iterative way, all products from reactions for which all substrates are part of the current seed are again added to the seed. This process is repeated until no further metabolites can be added. The obtained set of metabolites is the scope of the network. One problem using this approach, however, is that there are plenty of compounds which are needed for their own production, with the most prominent example being ATP. This issue is solved by adding "allowed" cofactor-pairs to the seed that allow the consumption of the respective compound if it is only as a cofactor, but do not allow *de novo* biosynthesis without a proper route being present.

Another attempt to address the issue of compound producibility was proposed by Benyamini et al. [137]. They proposed to constrain the FBA problem to account for metabolite dilution due to growth. The idea is to add exporters for all metabolites and force the export of small amounts of each metabolite which is present in a reaction in the solution. In both methods the problem of macromolecular carriers needs to be considered, as ferredoxin or flavodoxins, for example, cannot be accounted for explicitly, but only implicitly as part of the amino acid composition of the protein fraction of the biomass.

Chapter 2

Materials and methods

There are multiple ways how a model can be built and interrogated and an overview has been provided in the introduction. This chapter will detail the data sources used in the reconstruction of the metabolic model for *M. truncatula* and provide an overview of the tools used for network analysis. Finally it will detail the experimental procedures used to obtain the biomass composition of *M. truncatula*.

2.1 Model reconstruction

Model generation commonly follows several steps in which an increasing depth of detail is achieved. After generation of a draft network, the model is curated to eliminate and add reactions based on experimental evidence. Reactions need to be adjusted to correct inconsistencies and, depending on the model, need to be distributed between multiple compartments. A layout of the process used in this thesis is displayed in Figure 2.1 This section will illustrate the reconstruction steps and cover information on how data sources were included in the model and which methods were used to incorporate them.

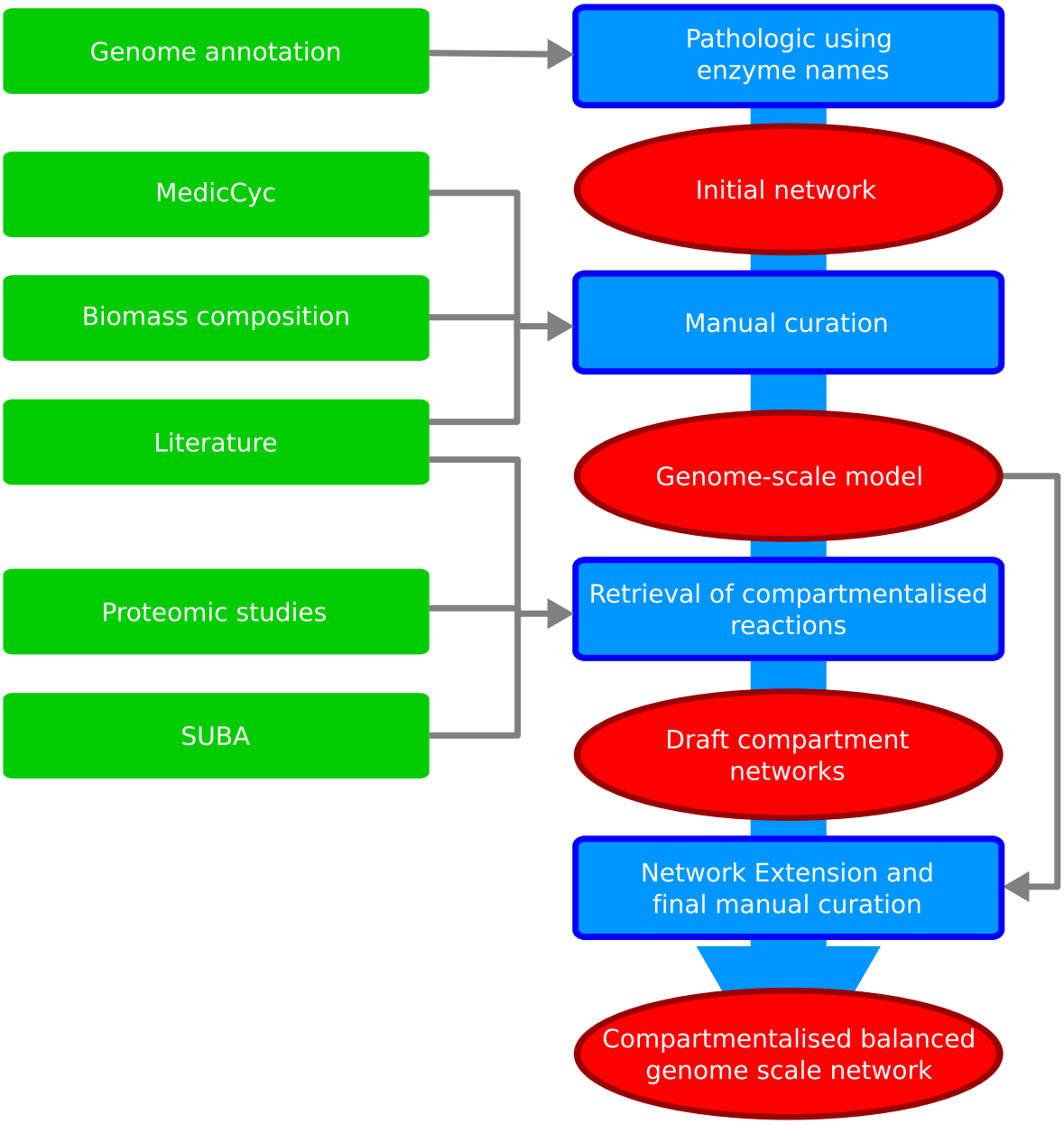


Figure 2.1: Illustration of the model reconstruction process. After an automated initial model reconstruction, literature and database information is included to obtain a charge- and atom-balanced genome-scale model. Proteomic studies and homology are used to create a draft compartmentalisation. The compartments are extended by the Network Extension algorithm and adjusted by manual curation using information from the literature. Most manual curation steps are semiautmatic. An computational method is used to detect the errors while they are corrected manually.

2.1.1 Initial model generation

The present model is based on version 3.5v5 of the Medicago genome annotation downloaded from the J. Craig Venter Institute [47]. This annotation was used to extract an initial network with the PathoLogic algorithm from PathwayTools 15.5 [131]. The PathoLogic reconstruction algorithm extracts reactions from MetaCyc based on matching enzyme names. The MetaCyc database used was version 15.5. PathoLogic creates a database in the BioCyc format [69] which was used to store the model information. The automated enzyme matching led to the inclusion of many reactions which were catalysed by unspecific enzymes (e.g. alcohol dehydrogenase). The obtained reactions in the database were therefore filtered and reactions matching all of the following conditions were removed.

- There are metabolites in the reaction which are not present in any other reaction;
- The reaction could not carry any flux under the condition that all nutrients can be imported and all metabolites can be exported; and
- There exists no evidence for the presence of the reaction beyond a match of unspecific reaction names.

The first condition identifies disconnected metabolites. This is either because these metabolites are the product of a pathway or because there is no information on their further metabolism. A product of such a pathway could be a secondary metabolite that is excreted by the organism or a compound which should be part of the biomass. To avoid the removal of these valid kinds of products, the second condition restricts the removed reactions to those unable to carry any flux. This assures that the removed reactions are disconnected from the remaining network. Thus a reaction fulfilling the

two initial conditions is unlikely to be present in a functional biosynthetic pathway. The third condition is necessary to avoid eliminating reactions which have specific evidence. These reactions are usually present in degradation pathways for xenobiotics, or catalyse reactions between highly specific compounds for which biosynthesis routes are unknown.

2.1.2 Manual curation

The database was matched against the MedicCyc database (version 1.0) [50] and information not present in the current database was included. The inclusion was restricted to reactions for which the respective gene annotation had not changed between the annotation used for MedicCyc 1.0 and the current gene annotation. The original MedicCyc database (v1.0) was constructed using annotated transcripts from an Affymetrix GeneChip. The Affymetrix gene identifiers were used as gene identifiers in MedicCyc. To incorporate the MedicCyc information, homology analysis (using BLAST) between the GeneChip sequences and the current genome annotation was used.

The combination of this data with the initial BioCyc database, created by PathoLogic, resulted in a list of enzyme-catalysed reactions converting mainly small molecules. However, the database also includes a number of reactions which contain macromolecules. Since metabolic networks are commonly limited to small molecules, the inclusion of macromolecular species posed a problem when trying to establish the atomic consistency of the reconstructed model. The reason for the exclusion of many macromolecules from metabolic models is due to their highly diverse structure. Instead of incorporating them explicitly, they are implicitly considered as part of the biomass or as conserved quantities in carrier molecules.

Modification reactions for macromolecules, which put a requirement for energy or

reductant on the metabolism, can therefore not be modelled explicitly for each single macromolecule but have to be considered as a general requirement in implicit reductant or energy consuming reactions. Two such reactions were included in the model:

- An ATP consuming reaction (ATPase):

$$ATP \; + \; H_2O \; \rightleftarrows \; ADP \; + \; Pi \; + \; H^+ \tag{2.1}$$

- A reductant consuming reaction (DEHOG) :

$$0.5\; O_2 \; + \; NADH \; + \; H^+ \; \rightleftarrows \; NAD^+ \; + \; H_2O \tag{2.2}$$

For analysis described in this work the DEHOG reaction was not allowed to carry any flux, unless explicitly mentioned. Similarly the ATPase was only allowed to carry flux in the forward direction (thus consuming energy). Even when turning of the ATPase reaction, the system would have plenty possibilities to consume ATP (e.g. the ATP driven cycle in Figure 1.13). Apart from macromolecule modifications, there are several reactions encompassing macromolecules which require further consideration. Those include reactions processing important storage molecules like starch as well as proteins, cell wall poly-sugars, DNA or RNA. All of these macromolecules are part of the biomass and therefore need a biosynthetic pathway.

This is achieved by treating macromolecules as their respective precursor molecules. However, when handling these molecules it has to considered that polymerisation commonly is achieved by some form of ligation reaction, usually accompanied by dehydration. This requires that the stored monomer is not of the same atomic composition as the monomeric compound but lacks one water molecule. In the present model this

is considered, e.g. assuming a monomer composition of glucose in starch as $C_6H_{10}O_5$. Many reactions in MetaCyc include these compounds as a polymer of a non-specific degree, starch(n), on the left hand side and an altered degree of size, starch(n+1), on the right hand side. In these reactions the unspecific polymers were replaced by the monomers, whose composition lacks one water molecule, such that they are restricted to only appear on one side of the reaction equation.

An additional group of macromolecules that cannot be excluded are carrier molecules like the acetyl-carrier proteins in fatty acid biosynthesis or ferredoxin electron carriers. As essential cofactors or carriers, these molecules have to be explicitly included in the reaction stoichiometries, but their synthesis pathways are outside the scope of a metabolic network model. These carriers, however, form conserved quantities that allow the introduction of pseudo-atoms to account for their core. This provides a way to ascertain atomic balance. The atomic balance was tested by using the method of Gevorgyan et al. [182]. The only non-conserved metabolite detected by this algorithm is light. This is to be expected, as light is considered solely as a source of energy driving photosynthesis and not as a particle, thus being consumed without any atomic production.

2.1.2.1 Charge balancing

The present model was intended to be able to simulate both photosynthetic and heterotrophic growth. As proton gradients are essential in both regimes, charge balance in all reactions was essential. Using MetaCyc charge information all reactions were checked for their charge balance. Apparently several reactions in MetaCyc (especially those in the phospholipid desaturation pathway) were curated to achieve mass balance by adding protons, a scheme commonly seen in many models. This is also observable

for many other reactions which are catalysed by dehydrogenase, or mixed function oxidase activities. Where possible, these reactions were adjusted by assuming common dehydrogenase or mixed function oxidase activities. In general, where protons had been added for mass balance reasons they were replaced by NADPH + H^+/$NADP^+$ pairs to achieve charge balance. This replacement was done by applying the general rules which are laid out in Table 2.1.

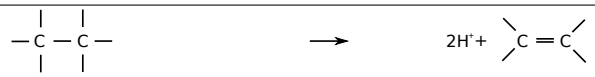
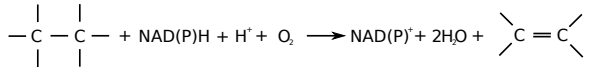
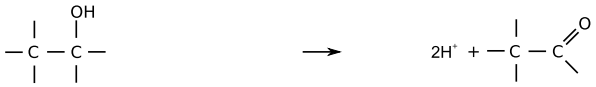
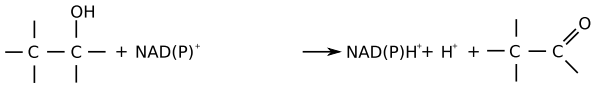
	Replacement example	Assumed enzymatic activity
Original reaction		desaturase
Corrected reaction		
Original reaction		alcohol
Corrected reaction		dehydrogenase

Table 2.1: Examples for the replacement of charge unbalanced reactions by conceptual enzymatic activities.

2.1.2.2 Gap filling

To address the gaps present in the network, data from the literature was collected. Metabolomic studies for *M. truncatula* were searched to obtain information on produced metabolites which might be missing from the reconstruction [44, 183, 45, 184, 185, 186]. For each of the detected metabolites a linear program was solved to establish whether a flux distribution exists, allowing the production of this specific metabolite from the provided substrates. When this failed, or the metabolite was not yet present in the network, MetaCyc was searched for plausible reactions which allowed a produc-

tion of the metabolite in question. Where present, gene information associated with the identified reactions was extracted and blasted against the *M. truncatula* genome. If multiple potential ways of biosynthesis were found, the most likely (according to gene similarities) was selected and manually added to the model.

Photosynthesic reactions Photosynthesis was included into the model using the following stoichiometries:

$$\begin{aligned}
PhotosystemII : 2\ H_2O_{stroma} \ +\ 4\ Light \ +\ 2\ PQ_{lumen} \ +\ 4\ H^+_{stroma} \\
\longrightarrow\ O_{2\ stroma} \ +\ 2\ PQH_{2,lumen} \ +\ 4\ H^+_{lumen} \\
Cytochrome\ b_6f : PQH_{2,lumen} \ +\ 2\ PC_{ox,lumen} \ +\ 2\ H^+_{stroma} \\
\longrightarrow\ PQ_{lumen} \ +\ 2\ PC_{red,lumen} \ +\ 4\ H^+_{lumen} \\
PhotosystemI : Ferredoxin_{ox,lumen} \ +\ 2\ PC_{red,lumen} \ +\ Light \\
\longrightarrow\ PC_{ox,lumen} \ +\ 2\ Ferredoxin_{red,lumen} \\
ATPase : ADP_{stroma} \ +\ Pi_{stroma} \ +\ 14/3\ H^+_{lumen} \\
\longrightarrow\ ATP_{stroma} \ +\ H_2O_{stroma} \ +\ 11/3\ H^+_{stroma}
\end{aligned} \tag{2.3}$$

Since the cyclic electron transfer is directly coupled from P700 to cytochrome b_6f a irreversible reaction providing ferredoxin to the stroma was included. This ensures that reductants cannot simply be transferred into the photosystem but allows the use of ferredoxins for the cyclic electron transfer flow. Cyclic electron flow is then incorporated as follows:

$$\begin{aligned}
CyclicElectronFlow : PQ_{lumen} \ +\ 2\ Ferredoxin_{red,lumen} \ +\ 2\ H^+_{lumen} \\
\longrightarrow\ PQH_{2,lumen} \ +\ 2\ Ferredoxin_{ox,lumen}
\end{aligned} \tag{2.4}$$

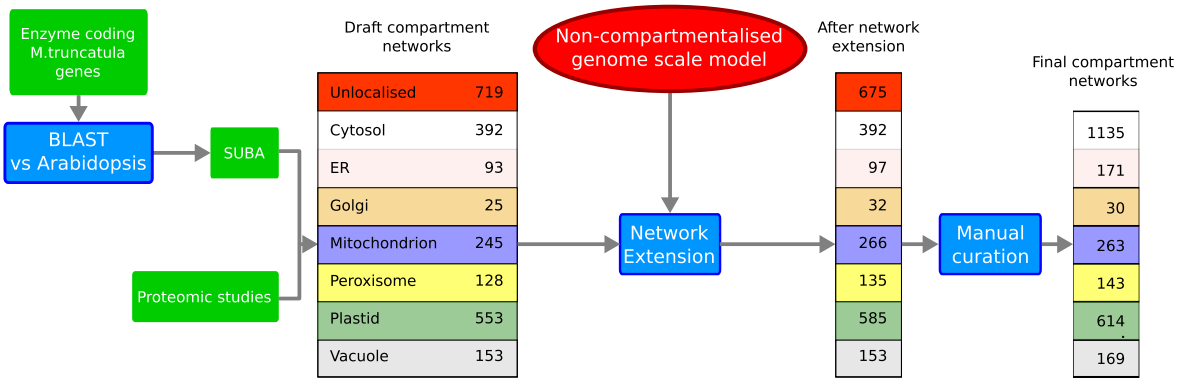


Figure 2.2: Process of compartment localisation. After a BLAST search for *A. thaliana* homologues and obtaining proteomic data [187, 188], Network Extension is used to fill in missing reactions. Manual curation ensured producibility of all compounds that were producible in the uncompartmentalised network but which were not part of the seed and target sets for the Network Extension algorithm. The genome-scale model is the uncompartmentalised model initially generated and used as reference network for Network Extension.

Thus $Ferredoxin_{lumen}$ in the model can be assumed to represent the electrons in P700.

2.1.3 Compartmentalisation

In order to realistically describe the metabolism of a plant cell, it is important to develop a fully compartmentalised model. Each compartment fulfils specialised tasks and key processes, such as ATP generation, are only possible due to gradients established over intercompartmental membranes. Key challenges when building a compartmentalised model are a) to realistically assign compartments to each reaction, b) to define transport reactions over intracellular membranes and c) to ensure that the resulting model is not able to generate and maintain gradients which are thermodynamically infeasible. The steps followed to address these challenges are detailed in Figure 2.2 and explained in the following paragraphs.

Assigning compartments to reactions To overcome the first challenge, all available protein sequences for annotated genes were scanned against the *A. thaliana* genome

on TAIR [189] and information about localisation of the homologues was extracted from the SUBA database [140]. Only information based on experimental evidence, such as GFP localisation, MS identification, TAIR data, and swissprot data, was extracted to localise reactions. Since computational predictions are known to result in a large number of false positives, the predicted localisations in SUBA were not included during this step.

In addition, two recent proteomic studies for root plastids [187] and mitochondria [188] in *M. truncatula* were used to assign compartments to reactions. The proteomic studies used identifiers from different versions of the *Medicago truncatula* Gene Index (MtGI) [190]. These identifiers were matched to identifiers of version 3.5v5 of the *M. truncatula* genome sequence using BLAST to obtain the corresponding genes, which were then matched to the corresponding reactions. The sources were combined such that a reaction was assigned to a compartment if there was at least one source of evidence assigning a corresponding gene to that compartment. The process resulted in a set of draft compartment networks with a large number of reactions without localisation remaining.

In the next step, an established network modelling approach [135] was used to ensure that each compartment contains a self-consistent metabolic network: the Network Extension method developed by Christian et al. identifies a minimal set of reactions from a reference network which need to be added to a draft network in order to ensure that a set of 'seed' metabolites (commonly the nutrients) can be metabolised into a set of 'targets' (e.g. molecules found in metabolomic studies). While for a whole organism, the seed can directly be inferred from defined growth media, and the targets can be obtained by metabolomic data, their definition is not straightforward for single compartments. Seed and target sets for each compartment were obtained by extensive

literature research for metabolites which are known to be transported into or produced by the compartment.

This step was performed for all compartments except the cytosol and the vacuole. The procedure is difficult for the cytosol, because it is in direct contact with all other compartments. This renders the definition of a seed and target set infeasible as too many targets would also be present in the seed and thereby automatically be produced. The vacuole, in turn, is a main location for degradation and storage, and thereby suffers from the same problem. Stored compounds are later exported again putting them into both seeds and targets, while degradation often leads to the very same compounds which are normally stored. Therefore, the initial reactions from gene localisation were used to search for degradation pathways including these reactions. Additional reactions from these pathways were then assigned to the vacuole if a majority of reactions of the respective pathway was already present.

In the Network Extension method for whole organism networks, a network containing all reactions stored in a biochemical database is usually used as a reference. The procedure for compartment prediction uses the genome scale uncompartmentalised model as reference instead. This is done to avoid the inclusion of new reactions into the model. The lists of seed and target metabolites used for this process and the resulting reactions added to each compartment can be found in Appendix B. All reactions which remained without assigned compartment after applying Network Extension to all compartments were subsequently assigned to the cytosol. An exception is the sterol metabolism, which is performed by a " channeling of substrates/products through microdomains of the endoplasmic reticulum" [191] and commonly assumed to happen in this compartment [192].

Transporters Describing transport processes of metabolites over membranes is a challenge in all compartmentalised metabolic networks. The main difficulty is that exact information about transported metabolites is still missing for many compartments. Due to localisation of biosynthetic enzymes, the existence of several transporters has to be assumed without experimental evidence, as otherwise these enzymes would lack their substrates. Recently, Linka and Weber reviewed many characterised transport systems for plants [193]. However, even this comprehensive review must necessarily miss transporters for which no molecular characterisation exists yet, but which are essential for production of different metabolites.

In addition to the transports listed in Linka and Weber [193] further transporters were added to allow the production of each metabolite that was originally producible in the uncompartmentalised network. To establish which metabolites were the most likely to be transported, the seed information from the network extension step along with several other sources [194, 195, 196] was used, and otherwise the biologically most plausible transporters were included. All transporters, with the exception of the reactions involved in the electron transfer chain, were assumed to be carrying uncharged metabolites. This ensured that no charge could be transported over membranes which would allow the regeneration of ATP (e.g. by the mitochondrial ATPase) without consuming some form of nutrients.

Maintaining thermodynamic integrity As touched upon in the paragraph above, protonation states and charge balancing are of great importance to achieve a thermodynamically consistent model. This is particularly true for the electron transport chains in which a charge imbalance could lead to the free regeneration of ATP. As the model requires a net charge uptake and export of zero, a proton exchanger and a bicarbonate exchanger were introduced to balance the charged uptake of inorganic nutrients.

Another issue caused by thermodynamic inconsistency is the occurrence of energy producing cycles within the model. These are likely to arise from reactions in the database which do not have an assigned directionality. Commonly these reactions are included as reversible in a model, as most metabolic reactions are reversible when the substrate and product concentrations are in specific ranges. Energy-producing internal cycles however are violating the first law of thermodynamics. As there is no consumption of substrates they would be able to act as an infinite energy source. To address these cycles the model was forced to produce energy without any import. This was achieved by setting the ATPase reaction to a fixed value greater than zero, while constraining all importers to zero. If a solution was found, the reactions involved in the solution were investigated for their reversibility. The reactions which were most unlikely to be functioning in the direction indicated by the solution (e.g. by comparing their respective ΔG using the eQuilibrator Tool provided by the Milo lab [197]) were selected and directionality and reversibility were adjusted to prohibit the obtained solution without restricting the other biosynthetic capabilities of the network.

2.2 Model analysis

Flux balance analysis (FBA) is a powerful tool to investigate the metabolic capabilities of an organism, for which a good reconstruction is available. However, single solutions of the formulated linear problem often provide only a glimpse of the metabolic capabilities. This section gives an overview of the kinds of algorithmic approaches used to investigate the adaptation of the system to changing environments and summarises the employed FBA methods.

General linear programming details The common linear problem solved in FBA as detailed in Equation 1.8 uses maximisation or minimisation of a linear combination

of fluxes as objective. When trying to minimise the absolute flux however, which is commonly referred to as flux minimisation, this method cannot be used as the absolute function is not linear. A linear function has to satisfy the additivity condition:

$$f(x\ +\ y) = f(x)\ +\ f(y) \tag{2.5}$$

The absolute function clearly violates this condition (with a simple example being $abs(-2\ +\ 3)\ = 1 \neq 5\ =\ abs(-2) + abs(3)$). The absolute function itself can therefore not be applied in linear programming and other approaches have to be taken. A simple addition of all fluxes however is insufficient, as reversible reactions can carry negative flux. As the aim of flux minimisation is to minimise the total amount of enzymes needed be minimising the total absolute flux, negative fluxes would contradict these aim. To overcome this issue all reversible reactions have to be split into a forward and backward reaction. This leads to all flux, regardless of the reaction directionality, being positive and allows the minimisation of the absolute flux by summing all fluxes. To reflect this, the stoichiometric matrix N is extended by one column per reversible reaction. For each reversible reaction column $\vec{r_i}$ this additional column $\vec{r_j}$ is set to $-\vec{r_i}$. Let k be the number of reversible reactions in the model and let l be the number of irreversible reactions. The resulting stoichiometric matrix N_{split} then contains $2 \cdot k + l$ columns which can be ordered as:

- forward reversible reactions
- backward reversible reactions
- irreversible reactions

with $r_{i+k} = -r_i \; \forall \; i \; \in \; 1 \ldots k$. In addition the stoichiometric constraints have to be adjusted such that no reaction is allowed to carry negative flux, as negative fluxes would contradict the objective of absolute flux minimisation. All negative flux constraints have to be converted into positive flux constraints for the backward reaction of the reversible reaction. Thus a constraint $a \leq p \leq b$ would be converted to two constraints $max(0, a) \leq p_{forward} \leq max(0, b)$ and $max(0, -b) \leq p_{backward} \leq max(0, -a)$. This translates the FBA formulation from Equation 1.8 to the following linear problem:

$$
\begin{aligned}
minimise : & \sum_{i=0}^{2k+l} c_i \cdot v_{split,i} \\
& s.t. \\
& N_{split} \cdot \vec{v_{split}} = 0 \\
& max(0, a_i) \leq v_i \leq max(0, b_i) \; ; \; \forall i \; \in \; \{1..k\} \\
& max(0, -b_i) \leq v_{2i} \leq max(0, -a_i) \; ; \; \forall i \; \in \; \{1..k\} \\
& a_j \leq v_j \leq b_j \; ; \; \forall j \; \in \; \{2k..2k+l\}
\end{aligned}
\tag{2.6}
$$

Using this split formulation for flux minimisation has the benefit that it is not necessary to restrict the two directions of each reversible reaction to only carry flux if the opposing direction does not, as this is automatically achieved by flux minimisation. If both forward and backward reaction would carry a flux A, there would be a flux distribution with a total minimal flux that is at least $2 \cdot A$ smaller than the current solution leaving the current solution non-optimal. This formulation is used throughout the analysis and has to be kept in mind when considering maximisation attempts on reversible reactions. While minimisation benefits from this approach, it is necessary to turn off the reverse flux when maximising a reaction flux. Furthermore, when the maximisation leads to

an infeasible solution, a minimisation of the negative flux is required. This will provide the maximal flux this reaction can carry even though the maximal flux is negative.

2.2.1 Energy demands

To establish the energy demands for different biomolecules the ATPase and DEHOG reactions (see equations 2.1 and 2.2) were allowed to carry flux in the reverse direction. The target metabolite exporter was set to one, all carbon imports beside the starch importer were set to zero and the system objective was set to minimise the uptake of starch. This ensures that only the required carbon is taken up and all other energy and reductant demand is matched by the ATPase and DEHOG reaction respectively. In the next step, the starch uptake was set to the determined minimum and the system was minimised for DEHOG activity. If the minimal DEHOG activity was zero, DEHOG was set to only carry positive flux and the system was maximised for DEHOG activity. The restriction to positive flux is necessary due to the split nature of the underlying linear problem, and this maximisation provides the surplus of reductant available. In the final step the DEHOG activity was fixed to the determined amount and the minimal ATPase activity was calculated in the same way as the DEHOG activity. It is crucial to first determine the DEHOG activity as reductant can easily be converted into ATP while this is not possible in the opposite direction.

2.2.2 FBA scanning

While single FBA solutions can provide an interesting impression and allow assessment of maxima and minima for given fluxes, scans provide a much better overview of the response of the network to a change in demand or environment. A scan in this context is a set of solutions of the linear program with the values of a specific constraint or

variable of the problem being varied from a minimum to a maximum. In this thesis two types of scans are used:

- flux scans
- weight scans

To perform a flux scan an objective function is chosen and constraints are applied to the linear program. An additional constraint γ for the target flux v_i is introduced and the target flux is set to γ. The problem is then solved multiple times while γ gets assigned values between a chosen minimum and maximum. The solution for each value of γ is stored and analysed for changing fluxes in response to the alteration of γ. This kind of scan can be applied to investigate effects of increasing energy demand (by using the ATPase reaction as v_i).

Weight scans do not pose a fixed constraint to the problem but rather alter the objective. For a given flux v_j the corresponding objective coefficient c_j is assigned a variable ϕ. As before, this variable is altered from a minimum to a maximum value. This kind of scan can illustrate the effects of an increasing cost of a certain enzymatic activity or compound (if the chosen reaction is an importer).

2.2.3 Programs used

The PathwayTools interface was used for maintaining the *M. truncatula* BioCyc database. The ScrumPy modelling framework was used for model reconstruction and analysis. A CPLEX interface was implemented based on the glpk interface in ScrumPy and CPLEX was used for FBA analysis. Matlab was used for visualisation and plotting and the implementation of the algorithm by Shlomi et al. [198] in the COBRA toolbox was used for submodel generation. Inkscape was used for the generation of figures of

metabolic pathways, and Cytoscape for network visualisation.

2.3 Experimental methods

This section details the methods used in the experiments performed to obtain the biomass composition of *M. truncatula*.

Hydroponic culture of *M. truncatula* Seeds were scarified for 8 minutes using 97% H_2SO_4. After washing they were incubated in a hydrochloric solution (Millipore tablets 2WCL01F50) for 30 minutes and finally washed with a large surplus of sterilized dH_2O. The seedlings were transferred to agar plates (0.8% (w/v)) and left in the dark at 25 °C for 48 hours. Plants were grown hydroponically for eight weeks at 21 °C in a 16/8h light/dark cycle. The nutrient solution was prepared according to the *Medicago truncatula* Handbook [51] with a final ammonium nitrate concentration of 8 mmol/L and no potassium nitrate. The high nitrogen concentration used should ensure that any symbiotic activity is inhibited. The tubs were aerated with air stones. Plants were harvested in two intervals; one harvest at the end of the light period, the other at the end of the dark period. After harvest, plants were submerged in liquid nitrogen and freeze dried till further use. Plant material was pooled and divided into root, stem and shoot material.

Extraction of soluble metabolites Soluble metabolites were extracted from lyophilized tissue using methanol/chlorofom according to the protocol of Lisec et al. [199]. Tissue samples (leaves, root, stem) were transferred to a mortar and ground to a fine powder under liquid nitrogen. A known mass (25 mg) was weighed into a pre-cooled 15 mL Falcon tube and 7 mL of 100% methanol (pre cooled at −20 °C) was added and vortexed for 10 s. 300 µL of ribitol (0.2 mg/mL) was added as internal stan-

dard at this point. Tubes were shaken for 10 min at 70 °C. 730 µL was transferred to a 2 mL Eppendorf tube and centrifuged at 11 000 g for 10 min to pellet insoluble material. 375 µL chloroform and 750 µL water (precooled to −20 °C and 4 °C respectively) were added. The tubes were briefly vortexed and centrifuged for 15 min at 2200 g to allow the non-polar and polar phases to separate. For GC-MS analysis aliquots of the polar phase were taken and dried down. Lipids were extracted using hexane/isopropanol extraction according to Hara and Radin [200] and quantified by weight.

Protein extraction, precipitation and hydrolysis Protein extraction was performed as described in Masakapalli [201]. 6 mL of 6 M urea/2 M thiourea solution was added to ground, lyophilized, samples (100 mg). The samples were carefully homogenized by a polytron and the extract divided into two 2 mL eppendorfs (1.5 mL of extract each) and centrifuged for 30 min at 20 000 g. The supernatants were collected and the remaining pellets were resuspended in 500 µL urea/thiourea. The supernatants from the second centrifugation and the first centrifugation were pooled and centrifuged again to remove all debris. The protein present in the supernatant was precipitated by adding 100% (w/v) trichloroacetic acid so that its final concentration was 15% (v/v). Samples were vortexed and stood on ice overnight in a cold room. The precipitated protein was pelleted by centrifugation of the samples for 40 min at 4 °C and 3100 rpm. The supernatant was discarded. The pellets were washed twice with 4 mL of ice-cold acetone and again centrifuged as above to remove acetone. The pellets were allowed to stand in a flow hood to evaporate any acetone left. Protein weights were taken after drying up overnight. The protein pellets were resuspended in 2 mL of 6 M HCl and incubated in a heat block at 100 °C for 24 hours for hydrolysis. Samples were cooled to room temperature and an aliquot (50 µL) of the hydrolysates (containing amino acids) was transferred to Eppendorf tubes and dried in a vacuum dryer overnight to remove

HCl. The dried samples were subjected to GC-MS analysis.

Nucleic acids RNA and DNA were extracted using TRIzol/DNAse and Chloroform/isoamylalcohol extraction methods respectively (adapted from [202]). Quantification was performed photospectrometrically using a NanoDrop (ND-1000 Spectrophotometer; NanoDrop Technologies Inc.; Wilmington, DE, USA).

Starch Starch was extracted according to an established protocol [20, 201] using the insoluble pellet obtained after the methanol chloroform extraction as sample. According to the protocol, "each sample was gelatinized by autoclaving the residue in 3 mL of 25 mM sodium acetate for 3 h at 121 °C, 1.04 bars pressure followed by enzymatic digestion with 20U α-amylase (Sigma-Aldrich) and 5U amyloglucosidase (Sigma-Aldrich) for 16 h at 37 °C. The samples were centrifuged and the supernatant containing the glucose monomers was collected, freeze dried and stored at −80 °C until further analysis"[201]. Starch was determined with an enzymatic assay.

Gas chromatography-mass spectrometry The GC-MS measurements were performed according to protocols in Masakapalli et al. [203] and Williams et al. [20], "on an Agilent 79890 GC coupled to an Agilent 5975 quadrupole MS detector, electron impact ionisation (70 eV) equipped with an Agilent HP 5-ms column (30 mm, 0.25 mm inner diameter) at the facility in the Department of Plant Sciences, University of Oxford, UK. Samples (5 µL to 50 µL) to be derivatised were dried in 1.5 mL Eppendorf tubes using a speed-vac system to ensure complete removal of water. Authentic commercial standards were analysed concurrently by GC-MS, to obtain the elution profiles and to identify mass fragments of all metabolites."[201]

The protein hydrolysate extracts were derivatised by TBDMS [204] whereas the soluble extracts were derivatised by MeOX TMS [199]. "To obtain the TBDMS deriva-

tives, the dried samples were first dissolved in 25 µL of pyridine and incubated at 37 °C, shaking at 900 rpm for 30 min. Then 35 µL of MtBSTFA + 1% t-BDMCS (N-methyl-N-(t-butyldimethylsilyl) trifluoroacetamide + 1% t-butyl-dimethylchlorosilane, Regis Technologies Inc) was added and the mixture was incubated at 60 °C, shaking at 900 rpm for 30 min."[201] To obtain the MeOX TMS derivatives, 40 µL of 20 mg/mL methoxyamine hydrochloride (Sigma) in pyridine was added to the sample and incubated at 37 °C, shaking at 900 rpm for 2 h. Then 70 µL MSTFA (N-methyl-N-(trimethylsilyl)trifluroacetamide (HiChrom)) was added and the mixture was incubated at 37 °C for 30 min at 900 rpm. The derivatised samples were transferred to 8 mm glass vials (Chromacol) and sealed with a septum cap. The samples were run through GC-MS. The mass spectra of all the samples were acquired for m/z 146-600 by scanning (at 4.38 or 3.19 scans s^{-1}). A solvent delay of 10 min was set so that the mass spectrometer was turned on after elution of the bulk of the solvent from the column.

Chapter 3

A genome-scale model of *M. truncatula*

Definitions Some definitions are necessary to understand the following description. The term conversion will refer to any reaction which converts one or more substrates into one or more distinct products. Care has to be taken that this is not confused with the term reaction which will refer to any metabolic process (transport, import/export or conversion). Metabolites are considered to be all compounds (micro or macromolecular) that are involved in reactions. To simplify the description, this includes macromolecular structures which function only as carrier molecules.

3.1 Model description

The information collected on *M. truncatula* is available in two distinct formats; as a PathwayTools database and as a metabolic reconstruction model in ScrumPy and SBML file formats. The PathwayTools database was created at the beginning of reconstruction (as detailed in Section 2.1.1) and alterations to the model were also incor-

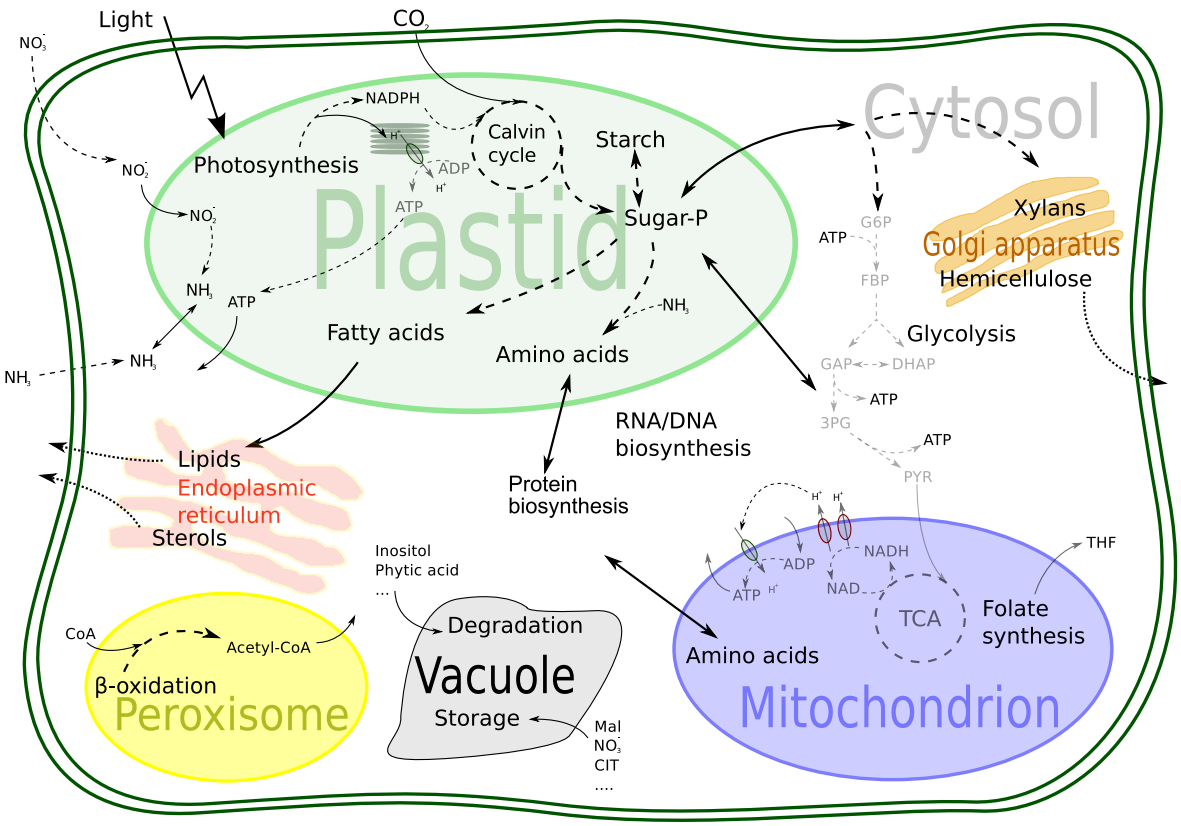


Figure 3.1: Model layout of the constructed *M. truncatula* metabolism. General localisation of important metabolic function is presented. Solid arrows indicate transport over compartment boundaries while dashed arrows represent intra compartment activity. The model comprises seven compartments plus a thykaloid subcompartment for the photosynthetic electron transfer chain.

porated into the database using the PathwayTools program suite [131]. The database contains additional reactions which were removed from the model. Among these are reactions for protein or DNA modification removed during the generation procedure (see Section 2.1.2). The enzymatic coverage (i.e. reactions with an associated gene-protein-reaction relationship) within this database is about 70% of all reactions. However, only about 65% of the conversions in the model have genetic evidence. When comparing the database to the original MedicCyc database (version 1.0) [50] this is an increase

of 15% in genetic evidence with an increase in size from 1509 to 2244 reactions, and an increase in the number of metabolites from 1168 to 1442. The high amount of genetic evidence allows the usage of tools described in Section 1.3.3, as many tools are dependent on including proteomic or transcriptomic data. This data can only be included if the corresponding reaction can be determined, which makes these gene-protein-reaction relationships so essential. Furthermore, the current database contains at least 190 pathways not present in the MedicCyc (1.0) database. In addition to these pathways, there were multiple changes to the information in MedicCyc which are detailed in Appendix A.

The main changes concern updates in MetaCyc identifiers and changed genome annotation information. In addition, several pathways with little or no evidence were removed or updated to more current versions. The larger fraction of reactions with evidence in the database compared to the model is due to protein modification reactions. In contrast to other reactions, which are likely to be inferred as missing steps in a pathway, protein modification reactions are not automatically inferred but only included in the database if they have direct genetic evidence. However, those protein modification reactions are commonly not part of a metabolic reconstruction and were therefore removed from the model. In the final model for *M. truncatula* there are a total of 1640 unique conversions, after removal of unspecific and protein modification reactions (see Section 2.1.1). The model conversions are distributed over seven compartments, with the cytosol being the largest of those compartments, containing 1135 conversions, or roughly 39% of all reactions (45% of all conversions). The general distribution can be seen in Figure 3.2.

As the MetaCyc categorization of the 1640 conversions shows, a majority of all conversions (1117) are involved in biosynthetic processes and 294 are involved in degrada-

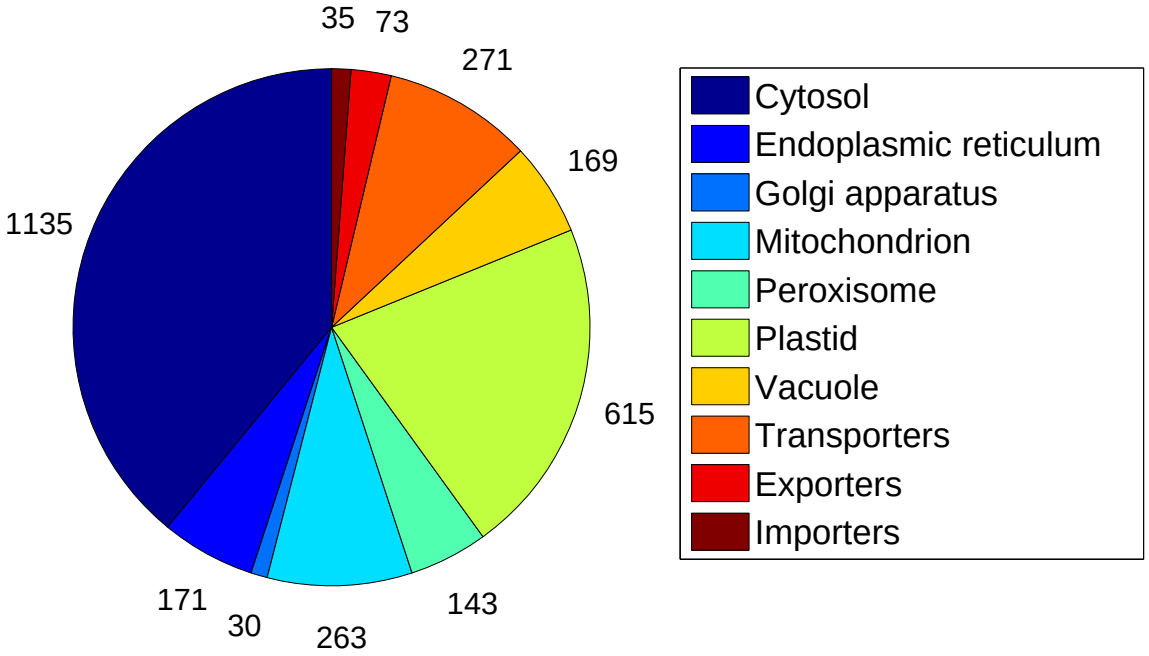


Figure 3.2: Distribution of the reactions over different compartments. Transporters include both single molecule transporters as well as symporters and antiporters. The largest compartment is the cytosol with about 39% of all reactions. The allocation of all unassigned reactions to this compartments is one of the major reasons for its size.

tion processes. Figure 3.3 displays the distribution of reactions over different metabolic processes. With over 400 conversions dedicated to secondary metabolism (∼150 occur in flavonoid biosynthesis), the model covers an extensive amount of the metabolic regime which is essential for forming the symbiotic relationship.

3.1.1 Reaction distribution

The model reactions are distributed over seven compartments (cytosol, plastid, mitochondrion, vacuole, peroxisome, endoplasmic reticulum, and Golgi apparatus). These compartments fulfil many different functions (see Figure 3.1). While there are a total of 572 conversions with duplicates in different compartments (e.g. initial parts of the TCA cycle in cytosol, plastid, and mitochondria; glycolysis in both cytosol and

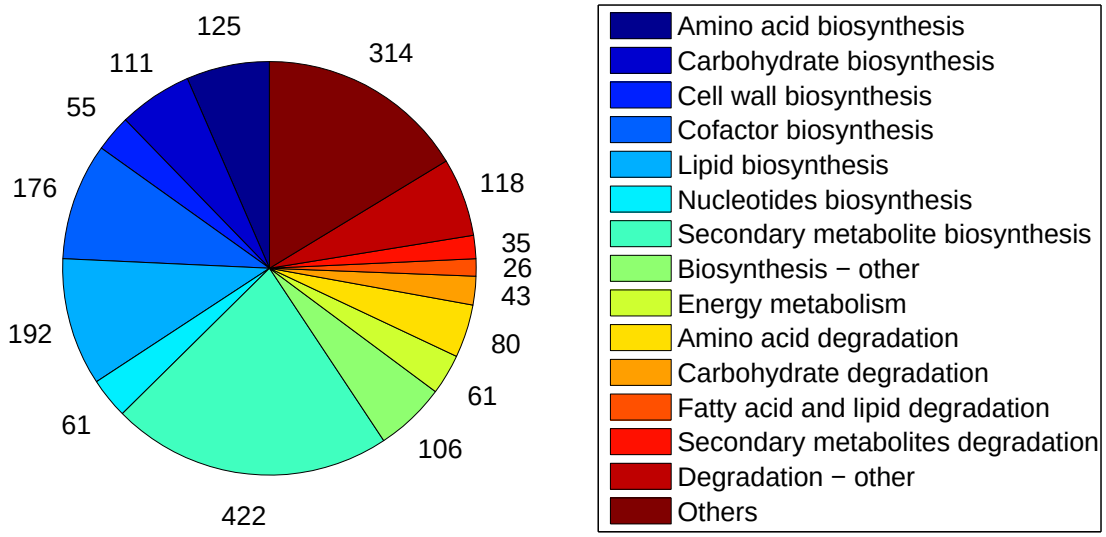


Figure 3.3: Detailed distribution of reactions to different metabolic processes. While there are a total of 1117 biosynthetic reactions, several reactions are present in multiple biosynthetic processes. The same applies for degradation where there are a total of 293 reactions.

plastid), some functions are solely located in specific compartments. The most important of these functions are photosynthesis in plastids, and the electron transfer chain in mitochondria, the two main energy sources for the organism. This is reflected by the comparatively high amounts of reactions involved in energy metabolism in these two compartments (see Figure 3.4). The plastid is also the major biosynthetic site beside the cytosol. While there are fewer biosynthetic conversions in the plastid than in the cytosol (417 to 762 respectively), the ratios of active reactions in a biomass producing solution are almost reversed. Of a total of 430 flux carrying reactions in a flux minimal solution for growth on starch, 217 conversions are present in the plastid, while only 115 conversions in the cytosol are employed. The remaining reactions are mainly transports between these two compartments and a few mitochondrial reactions from the TCA along with the electron transfer chain. All major biomass compounds

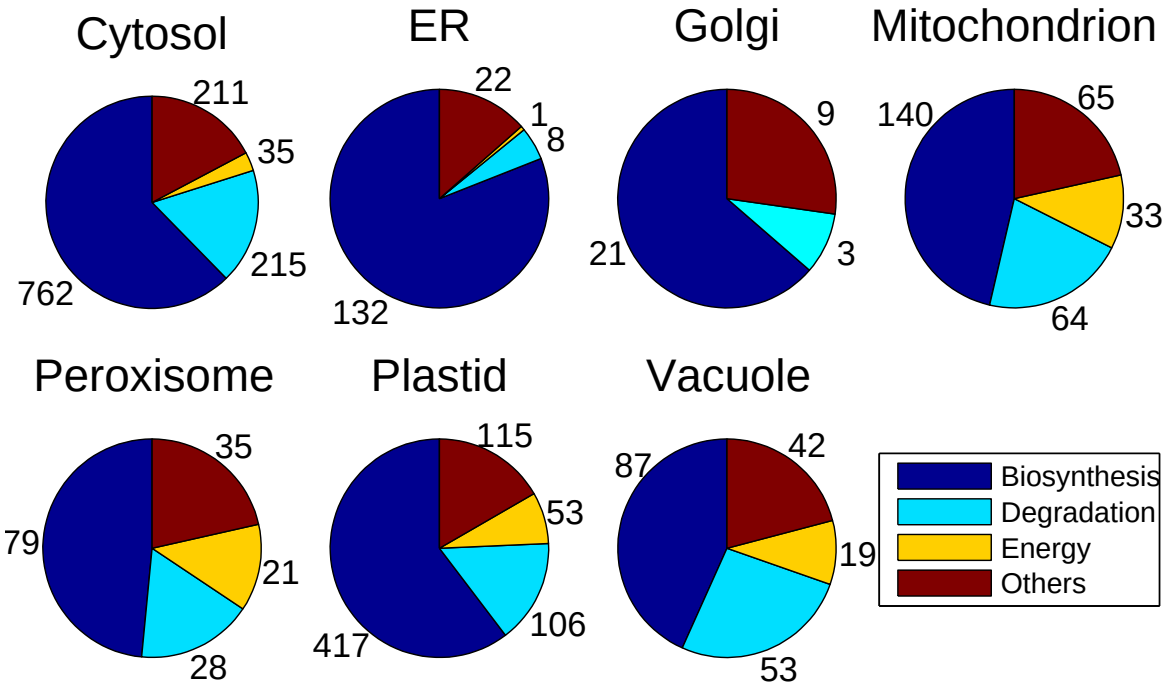


Figure 3.4: Distributions of the types of conversions in the different compartments. The assignments are based on MetaCyc pathway categories, with others being either uncategorized reactions or reactions which do not fall into the three explicit categories. Conversions can be present in multiple groups (leading to a sum of reactions for each compartment larger than the number of reactions in that compartment.

can be produced using only these three compartments. The purpose of the vacuole as storage and degradation compartment is reflected by the largest fraction (31% of all conversions) of degradation reactions of all compartments.

3.1.2 Metabolite distribution

The model contains a total of 1370 distinct metabolites over half of which (701) are present in more than one compartment. Figure 3.5 gives an overview of the different classes of metabolites present in the model according to their MetaCyc classification.

Of all metabolites in the model, 1161 are producible in at least one of the com-

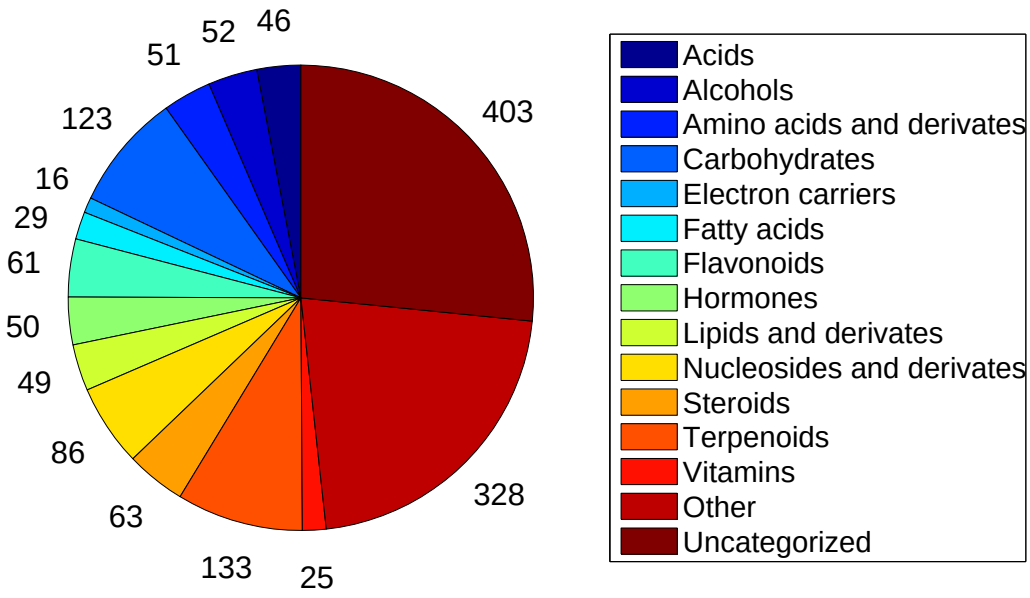


Figure 3.5: Distribution of types of metabolites in the model of *M. truncatula*. The classes are derived from the MetaCyc database. Uncategorised compounds have no categories assigned to them in MetaCyc.

partments. The remaining 209 metabolites fall into three main classes. The first class, containing 121 metabolites, includes those substances which form conserved quantities. The second contains 65 metabolites which are disconnected from the remaining network, while the third class contains the 23 metabolites that are only present in degradation pathways.

Examples for the first class are the acetyl carrier group linked to fatty acids, macromolecular redox carriers, such as ferredoxins, and tRNAs, which are necessary for the incorporation of amino acids into proteins. Most of these are small proteins and RNA molecules implicitly contained in the biomass, which includes the protein and nucleic acid composition of a grown cell. Since the model can produce biomass, biosynthesis is possible for over 86% of the compounds in this class. The remaining are small molecules, which cannot be biosynthesized because they are themselves needed for their

own biosynthesis. An example is quinate, for which no biosynthetic route is present in our database, due to the lack of a clear biosynthetic route within MetaCyc (where it is only considered as a degradation substrate). However it can be used as one of two routes from coumaroyl-CoA to caffeoyl-CoA in which it is needed as an carrier-like compound.

The second class contains many of the biotin biosynthesis pathway intermediates (6 compounds). The reason why this pathway is disconnected is that it is unknown how the pimeloic acid moiety necessary for biotin biosynthesis is synthesised in plants. Most of the other metabolites (37) in this class are produced by specific enzymes, but no way for the production of their precursors is known.

The third class consists of metabolites, which result from the degradation of modified macromolecules, like polyfructofuranose degradation, or which are present only in a degradation pathway for xenobiotics (e. g. cyanate). The phenomenon of metabolites which can only be 'produced' by degradation, was discussed in detail in Section 1.3.3.

Flavonoids are an important class of compounds for the formation of the legume-rhizobia symbiosis and have shown promising properties as antibiotics [205]. Legumes produce a large range of these compounds. Thus, legumes could become important sources for these types of compounds if yields can be increased. Even though the total number of flavonoids in the present model is relatively small (only 61 according to MetaCyc categorisation) they form one of the larger groups of secondary metabolites. Many biosynthesis pathways were extracted from the MedicCyc database which used information from Farag et al. [183] to validate the presence of these compounds. With the importance of these compounds for symbiosis it was imperative to ascertain their producibility. This was achieved in the current model by including hypothetical biosynthesis pathways from literature sources [206, 207, 208].

3.1.3 Discussion

The present model covers most aspects of metabolism and assigns the site of many biosynthetic processes to specific compartments. It allows growth on either sucrose (for pure sink tissues like roots during night and day), starch (during the night) or light and carbon dioxide (in leaves during the day). It covers areas of secondary metabolism important for the formation of the rhizobial symbiosis, like the flavonoid pathways, allowing investigation of gene deletion effects on flavonoid production and prediction of possible fix$^-$ or nod$^-$ mutants. With the exception of some metabolites involved in disconnected reactions the model is able to synthesise a majority of 92.3% of all present compounds. Even though many of these metabolites are not included in the biomass it is essential that they can be produced, as there would otherwise be a dilution of intermediates during growth (which would contradict the steady state assumption). The importance of the dilution effect during growth was discussed in detail by Kruse and Ebenhöh [136] and Benyamini et al. [137].

The scale of the present model is comparable with other genome-scale plant network reconstructions [120, 121, 171]. Compartmentalisation was achieved with a novel approach using the Network Extension method to bridge the gaps resulting from assignment by gene homology. The high involvement of the plastid in biomass production, in particular in amino acid biosynthesis, is in good agreement with current knowledge. In plants, many amino acids are almost exclusively produced in the plastid. Even for amino acids that are producible in the cytosol via transaminases, the necessary nitrogen is initially fixed in the plastid [209]. The model is the first charge balanced genome-scale reconstruction of a plant, and the first metabolic reconstruction of *M. truncatula*. It can provide valuable insight into the metabolism of this model plant and allow

inference of properties of legumes.

3.2 Biomass composition

Knowing the biomass composition of the plant is critical for any further analysis, since it is one of the most important constraints for the linear program. It is essential that the metabolic network can provide the organism with energy for its maintenance requirements and to supply the metabolites necessary for growth. The latter is the most important constraint for metabolic modelling, as it is the main source of restriction for the flux space.

3.2.1 Day and night differences

Biomass composition of *M. truncatula* was established for three major tissues (leaves, roots and stem) at end of day (ED) and end of night (EN), see Table 3.1. There is a significant increase in the percentage of cell wall material between end of day and end of night measurements in both stem and root tissues ($p < 0.05$). This is most likely due to a decrease in the fraction of soluble compounds in both of these tissues, even though this is only significant for stems ($p < 0.01$) while it is not significantly different in roots (with p = 0.10). The drop in the amount of soluble compounds is likely due to a slow depletion of sucrose and other currency metabolites which are replenished during the day and consumed during the night.

The second change is in the amounts of starch present at end of day and end of night. This change reflects the plant's effort to store carbon as starch during the day (high end of day value) and consume this starch during the night (low end of night value). As the change in starch content is rather large it is even more important than the change in other currency metabolites. While roots only showed minor changes

in starch content between ED and EN ((0.88 ± 0.02) and $(0.98 \pm 0.05)\,\mathrm{mg/g}$ respectively), the changes in stem and particularly leaf are much higher. For leaves, the change between ED and EN was $(19.62 \pm 1.66)\,\mathrm{mg/g}$, almost 2% of the total leaf biomass, while it was $(4.25 \pm 0.27)\,\mathrm{mg/g}$ in stems. This has to be put into perspective with growth rates, being a substantial fraction of each days growth. Relative growth rates of plants range from $0.100\,\mathrm{g\,g^{-1}\,d^{-1}}$ to $0.120\,\mathrm{g\,g^{-1}\,d^{-1}}$ for *M. Sativa* [210] and $0.12\,\mathrm{g\,g^{-1}\,d^{-1}}$ to $0.25\,\mathrm{g\,g^{-1}\,d^{-1}}$ for *A. thaliana* [19]. While other metabolites' relative amounts need to be produced with respect to the relative growth rates, starch has to be produced on a daily basis. Thus, when considering a relative growth rate of $100\,\mathrm{mg\,g^{-1}\,d^{-1}}$, in leaves a total of $19.62\,\mathrm{mg/g}$ has to be starch. Under the assumption that the majority of growth occurs during the day, this would infer that instead of 2% of the produced biomass (which would be the appropriate relative growth rate), about 20% of the biomass has to be starch as it has to be resynthesized from 0 every day. However, starch does not contribute to the structural biomass (i.e. the biomass that becomes a permanent part of the plant) and only imposes an increase in photosynthesis activity during the day. We therefore excluded starch from the formulation of our biomass composition to reflect this transient nature of starch.

Table 3.1: Biomass composition of *M. trunctula*

Component	Dry weight composition in % (g/gDW)						
	End of day			End of night			Av.
	Leaf	Root	Stem	Leaf	Root	Stem	
Cell wall	45.81 ± 1.11	50.93 ± 2.54	46.96 ± 0.29	46.51 ± 6.05	57.30 ± 1.73	55.42 ± 4.34	50.49
Protein	23.59 ± 3.25	4.15 ± 0.98	5.60 ± 0.36	20.70 ± 2.49	7.12 ± 3.49	4.80 ± 0.46	10.99
Total soluble	32.65 ± 0.79	32.23 ± 3.22	36.10 ± 2.00	32.77 ± 4.89	25.72 ± 2.33	25.94 ± 4.00	30.90
Starch	2.01 ± 0.16	0.09 ± 0.01	0.49 ± 0.02	0.04 ± 0.01	0.10 ± 0.05	0.06 ± 0.01	0.86(a) and 0.07(b)
Soluble fraction parts	% of dry weight						% of Soluble
	End of day			End of night			
	Leaf	Root	Stem	Leaf	Root	Stem	
RNA	0.86 ± 0.35	0.32 ± 0.31	0.61 ± 0.16	0.52 ± 0.06	0.42 ± 0.18	0.39 ± 0.05	1.68
DNA	1.55 ± 0.37	1.54 ± 0.73	0.93 ± 0.32	1.36 ± 0.53	1.66 ± 0.46	0.94 ± 0.47	4.30
Lipids	7.52 ± 0.11	2.01 ± 0.58	2.79 ± 0.30	6.65 ± 0.86	2.34 ± 0.21	2.35 ± 0.24	12.77
Chlorophylls and Carotenoids	0.99 ± 0.15	-	0.27 ± 0.02	0.70 ± 0.10	-	0.22 ± 0.02	1.173
MetOH Soluble composition	Soluble weight composition in % of measured soluble extract						
	End of day			End of night			Av.
	Leaf	Root	Stem	Leaf	Root	Stem	
Arabinose	3.65 ± 0.73	6.38 ± 0.72	3.14 ± 0.27	4.04 ± 0.47	6.96 ± 1.65	3.41 ± 0.43	4.60
Asparagine	49.09 ± 2.50	28.98 ± 3.62	66.69 ± 2.59	60.57 ± 2.49	29.38 ± 0.86	68.08 ± 3.84	50.46
Aspartate	0.18 ± 0.03	0.32 ± 0.09	0.17 ± 0.00	0.25 ± 0.00	0.21 ± 0.05	0.16 ± 0.02	0.21
Citrate	0.85 ± 0.12	1.22 ± 0.17	0.53 ± 0.04	1.03 ± 0.05	1.20 ± 0.32	0.63 ± 0.10	0.91

Table 3.1: (continued)

Component	End of day			End of night			Av.
	Leaf	Root	Stem	Leaf	Root	Stem	
Fructose	3.64 ± 0.44	9.04 ± 1.12	5.58 ± 1.00	3.19 ± 0.23	9.26 ± 5.32	5.10 ± 0.46	5.97
Glucose	3.69 ± 0.47	13.28 ± 2.80	2.72 ± 0.18	4.66 ± 0.33	14.41 ± 1.65	3.06 ± 0.27	6.97
Glutamate	7.17 ± 0.93	6.78 ± 0.75	3.69 ± 0.24	7.37 ± 0.36	6.97 ± 1.14	3.64 ± 0.48	5.94
Glycine	0.46 ± 0.08	0.76 ± 0.08	0.35 ± 0.03	0.50 ± 0.06	0.81 ± 0.16	0.38 ± 0.06	0.54
Isoleucine	0.15 ± 0.01	0.20 ± 0.01	0.14 ± 0.01	0.17 ± 0.08	0.18 ± 0.02	0.15 ± 0.01	0.16
Malate	1.26 ± 0.17	2.42 ± 0.38	0.73 ± 0.06	1.01 ± 0.09	1.99 ± 0.73	0.78 ± 0.13	1.37
Myo-inositol	2.45 ± 0.20	1.93 ± 0.21	1.14 ± 0.09	1.98 ± 0.10	2.10 ± 0.50	1.05 ± 0.15	1.78
Pyroglutamic acid	0.32 ± 0.02	0.37 ± 0.18	0.42 ± 0.09	0.19 ± 0.04	0.05 ± 0.05	0.35 ± 0.1	0.28
Palmitic acid	0.56 ± 0.31	0.66 ± 0.14	0.31 ± 0.24	0.60 ± 0.24	0.66 ± 0.06	0.33 ± 0.07	0.52
Phenylalanine	0.16 ± 0.27	0.02 ± 0.03	0.49 ± 0.04	0.98 ± 0.06	0.17 ± 0.30	0.59 ± 0.37	0.40
Pinitol	1.75 ± 0.05	0.51 ± 0.26	1.59 ± 0.08	2.10 ± 0.04	0.64 ± 0.36	1.43 ± 0.18	1.34
Proline	0.21 ± 0.03	0.15 ± 0.02	0.08 ± 0.01	0.20 ± 0.01	0.16 ± 0.02	0.08 ± 0.01	0.15
Serine	0.72 ± 0.08	0.00 ± 0.00	0.19 ± 0.06	0.73 ± 0.07	0.00 ± 0.00	0.06 ± 0.04	0.28
Stearic acid	1.14 ± 0.64	1.26 ± 0.19	0.72 ± 0.55	1.35 ± 0.57	1.30 ± 0.14	0.73 ± 0.17	1.08
Succinate	0.66 ± 0.10	1.15 ± 0.14	0.41 ± 0.04	0.60 ± 0.12	1.23 ± 0.26	0.44 ± 0.07	0.75
Sucrose	21.17 ± 0.40	23.66 ± 5.38	10.38 ± 0.69	7.46 ± 0.43	21.37 ± 11.44	8.93 ± 1.30	15.49
Threonine	0.26 ± 0.01	0.34 ± 0.04	0.17 ± 0.01	0.34 ± 0.01	0.34 ± 0.03	0.18 ± 0.00	0.27
Valine	0.45 ± 0.03	0.57 ± 0.05	0.37 ± 0.02	0.66 ± 0.03	0.60 ± 0.08	0.42 ± 0.02	0.51
Protein composition	Soluble weight composition in % of protein						
	End of day			End of night			Av.
	Leaf	Root	Stem	Leaf	Root	Stem	
Alanine	7.51 ± 0.36	7.27 ± 0.38	7.06 ± 0.25	7.52 ± 0.07	7.71 ± 0.05	8.02 ± 0.16	7.52
Arginine	2.60 ± 0.71	2.97 ± 0.41	3.26 ± 0.18	3.66 ± 0.28	4.17 ± 0.07	4.11 ± 0.03	3.46

Table 3.1: (continued)

Component	End of day			End of night			Av.
	Leaf	Root	Stem	Leaf	Root	Stem	
Asparagine and Aspartate*	14.76 ± 1.31	13.85 ± 1.63	12.93 ± 0.26	13.14 ± 0.81	11.99 ± 0.40	12.51 ± 0.19	13.20
Cysteine *2	2.02 ± 0.29	2.16 ± 0.13	2.15 ± 0.05	2.45 ± 0.10	2.39 ± 0.06	2.38 ± 0.05	2.26
Glutamine and Glutamate*	13.01 ± 0.84	12.50 ± 1.19	12.10 ± 0.34	12.43 ± 0.48	12.10 ± 0.24	12.43 ± 0.05	12.43
Glycine	9.91 ± 2.85	8.46 ± 1.92	7.25 ± 0.38	7.85 ± 0.56	8.08 ± 0.26	8.72 ± 0.19	8.38
Histidine	2.20 ± 1.32	2.88 ± 0.94	3.19 ± 0.12	3.42 ± 0.38	3.48 ± 0.25	2.92 ± 0.34	3.02
Isoleucine	0.54 ± 0.15	0.44 ± 0.16	0.33 ± 0.09	0.53 ± 0.07	0.38 ± 0.03	0.55 ± 0.05	0.46
Leucine	4.77 ± 0.48	5.15 ± 0.12	4.74 ± 0.04	4.29 ± 0.30	4.57 ± 0.16	4.14 ± 0.43	4.61
Lysine	6.56 ± 0.83	7.18 ± 0.60	6.92 ± 0.23	7.43 ± 0.36	6.78 ± 0.16	6.45 ± 0.17	6.89
Methionine	2.02 ± 0.29	2.16 ± 0.13	2.15 ± 0.05	2.45 ± 0.10	2.39 ± 0.06	2.38 ± 0.05	2.26
Phenylalanine	6.05 ± 0.15	6.15 ± 0.06	5.70 ± 0.03	5.68 ± 0.04	6.51 ± 0.06	6.48 ± 0.11	6.10
Proline	2.03 ± 0.67	2.77 ± 1.31	2.99 ± 0.52	2.09 ± 0.39	3.64 ± 0.30	3.04 ± 0.07	2.76
Serine	5.96 ± 0.65	5.32 ± 0.64	4.87 ± 0.29	5.24 ± 0.24	4.27 ± 0.07	4.41 ± 0.27	5.01
Threonine	4.52 ± 2.61	4.79 ± 2.37	8.26 ± 1.61	6.09 ± 1.29	4.26 ± 0.54	4.26 ± 0.38	5.36
Tryptophan*2	6.26 ± 0.12	6.29 ± 0.12	6.08 ± 0.05	6.09 ± 0.13	6.76 ± 0.08	6.80 ± 0.10	6.38
Tyrosine	6.47 ± 0.10	6.42 ± 0.22	6.47 ± 0.07	6.50 ± 0.22	7.00 ± 0.18	7.11 ± 0.08	6.66
Valine	2.80 ± 0.21	3.21 ± 0.36	3.57 ± 0.23	3.15 ± 0.26	3.50 ± 0.14	3.28 ± 0.45	3.25

Table 3.1: Biomass composition of *M. truncatula*. Measurements were taken at end of day and end of night. For the biomass composition used in the analysis the average values were used and pyroglutamate was assumed to have spontaneously cyclized from glutamate. The MetOH soluble amount for the calculations is assumed to be the remaining fraction of the total soluble amount. Errors are standard deviation of three technical replicates.

* Glutamate/glutamine and aspartate/asparagine are assumed to be half the peak of aspartate and glutamate respectively as the acid hydrolysis leaves only glutamate/aspartate.

*2 Cysteine and tryptophan could not be directly measured. Cysteine was assumed to be at equal concentration as methionine, while tryptophan was assumed to be the average of phenylalanine and tyrosine.

3.2.2 General observations and additional data

While data for protein composition, soluble metabolites composition, lipid fraction, starch fraction, cell wall fraction, chlorophyll concentration and nucleotide concentration were determined experimentally (see Section 2.3), other data was obtained from the literature, or assumed to be similar to data for *M. sativa*. The root:stem:leaf ratio was based on Mettupalli [53] and Yousfi et al. [211] and assumed to be 1:1:1. Lipid composition data (based on fatty acid distributions) was taken from Bakoglu et al. [212]. The cell wall composition, including pectins, lignin, cellulose and hemicellulose, was derived from data of Jung and Engels [213], Nakashima et al. [214] and Johnson et al. [215] and assumed to be similar in *M. truncatula* as in *M. sativa*. GC content was assumed, based on measurements by Blondon et al. [216]. An overview of this data is provided in Table 3.2.

The general biomass reaction contains the relative amounts listed in the average column of Table 3.1 and accounts for the energy (ATP) necessary to polymerise proteins (amino acid loading on tRNAs and polymerisation) and nucleic acids (RNA and DNA). It does not, however, account for maintenance energy requirement.

The distribution of compounds found in the samples is similar to those found in other plants [120]. In addition to most commonly detected metabolites the GC-MS showed a peak which proved to be pinitol. Pinitol is a myo-inositol derivative which is commonly associated with osmotic regulation. It is assumed that pinitol is present due to the growth conditions in a hydroponic solution. The large amount of asparagine detected stems from the composition of the growth solution. As it was necessary to inhibit any symbiosis the amount of ammonium nitrate used was selected at a high level, since high levels of available nitrogen reduce the likelihood of symbiosis. This

Component	Metabolite	Fraction	Source
Lipid	Palmitic acid	10.79	[212]
	Stearic acid	1.88	[212]
	Oleic acid	12.10	[212]
	Linoleic acid	41.95	[212]
	Linolenic acid	28.53	[212]
Cell wall	Hemicellulose	28.3	[215]
	Cellulose	44.5	[215]
	Lignin	13.0	[215]
	Pectin	14.2	[215]
Hemicellulose	Xylose	69-73.1	[213]
	Mannose	24.4-27.8	[213]
	Fucose	2.5-3.2	[213]
Lignin	S-Lignin	12.3	[214]
	G-Lignin	81.3	[214]
	H-Lignin	6.4	[214]
Pectin	Uronic acids (Galacturonate)	64.3-65.9	[213]
	Arabinose	14.2-14.8	[213]
	Galactose	15.8-16.5	[213]
	Rhamnose	4.0-4.6	[213]

Table 3.2: Data from external sources used for the biomass formulation. All data is for *M. sativa* and was assumed to be similar in *M. truncatula*

large amount of available nitrogen led to an accumulation of nitrogen within the plant which was mainly stored as asparagine. This has to be kept in mind when simulating symbiotic nitrogen fixation. The asparagine accumulation is only due to the excessive amounts of available nitrogen in the nutrient solution, whereas nitrogen availability is the limiting factor during symbiotic nitrogen fixation.

Chapter 4

Analysis of energy and nitrogen metabolism of *M. truncatula*

The present model is a good source to estimate energetic costs of biomolecules, and to this end an in-depth analysis was performed. As *M. truncatula* is a model for the legume-rhizobia symbiosis the nitrogen metabolism is also of high importance when analysing the model. This chapter gives an overview of simulated reactions of the nitrogen metabolism to different nitrogen sources. In addition, the effect of proton gradients, reflecting soil pH, was investigated, and is presented.

4.1 Energy requirements in photosynthetic organisms

There are some general observations when investigating energetic costs for different metabolites. One important aspect is the inability of metabolism to convert ATP into reductant, which makes it important to split the costs between reductant and energy

costs. While reductant can easily be employed to regenerate ATP by transport to the mitochondrion and subsequent oxidative phosphorylation, or even by direct cyclic electron transfer in the thylakoid membrane, the opposite is not possible under normal conditions. The inability to convert ATP into reductant needs to be kept in mind for the evaluation of any energy requirement study. In theory, the mitochondrial ATPase and the whole electron transfer chain could be forced to work in reverse; however, the conditions necessary for this reverse function are far from the conditions found in a living organism. When using a carbon source as an energy source, this effect is not as pronounced as during photosynthetic growth, since the pentose phosphate pathway can be employed to produce reductant without regenerating ATP. There are also bypasses of the ATP-producing succinate synthase in the TCA, allowing the exclusive production of reductants without any ATP regeneration. In contrast, photosynthesis produces a minimal amount of ATP per reductant. Thus, during the day, an increased requirement for reductant necessarily induces an increased amount of ATP production. The amount of ATP per reductant produced by photosynthesis is about 1.286 ATP per NADH. Since the Calvin cycle (see Figure 1.8) needs 3 ATP and 2 NADPH to fix one carbon, there is a lack of about half an ATP per fixed carbon in the cycle.

This discrepancy is dependent on the type of compound fixed, as more oxidized molecules can themselves provide some reductant. If the system is optimised to minimise light usage, with a constraint to fix one CO_2, the mitochondrial electron transfer chain is used to regenerate ATP with the available reductant, as it is more efficient than the photosynthetic electron transfer chain. Even though the photosynthetic electron transfer chain is in theory able to cycle electrons in order to alter the NADH to ATP ratio, this cycling is less (light-)efficient than using the electrons to produce reductants and transport them to the mitochondrion. One electron cycle in the thy-

lakoid membrane has a theoretical yield of only 0.86 ATP. If the same amount of electrons were transported to the mitochondrion as NADH, the yield would increase to 2.5 ATP. However, the cyclic electron flow is more flux efficient than transport to the mitochondrion, a property that can make it preferable if sufficient light is available. A similar example of a more flux-efficient but less energy-efficient pathway is the pentose phosphate pathway (PPP). While the PPP does not yield the same amount of ATP as the combination of glycolysis and TCA, it is more flux-efficient when higher levels of reductant are required.

4.1.1 Energy and reductant requirements for different biomolecules

The energetic requirements for multiple biomolecules were investigated to get an impression of the general requirement for energy by the plant. An overview of energy requirements for some groups of molecules is provided in Figure 4.1.

The highest reductant costs for biosynthesis can be found in fatty acids, as their saturation requires high amounts of reductant. Lignins and carotenoids also rank high in the reductant requirements while organic acids and nucleotides have the lowest demand. This is expected of organic acids, as these are the products of reductant yielding pathways like the TCA but unexpected for nucleotides.

ATP requirements in general are rather similar for all compound classes, with the highest requirements per carbon being found in fatty acids. The analysis suggests that larger or more complex molecules (fatty acids, chlorophylls, vitamins/cofactors) have higher ATP requirements on a per carbon basis than smaller molecules, like amino acids, organic acids, or sugars. The highest reductant to ATP requirement ratios can be found in nucleotides.

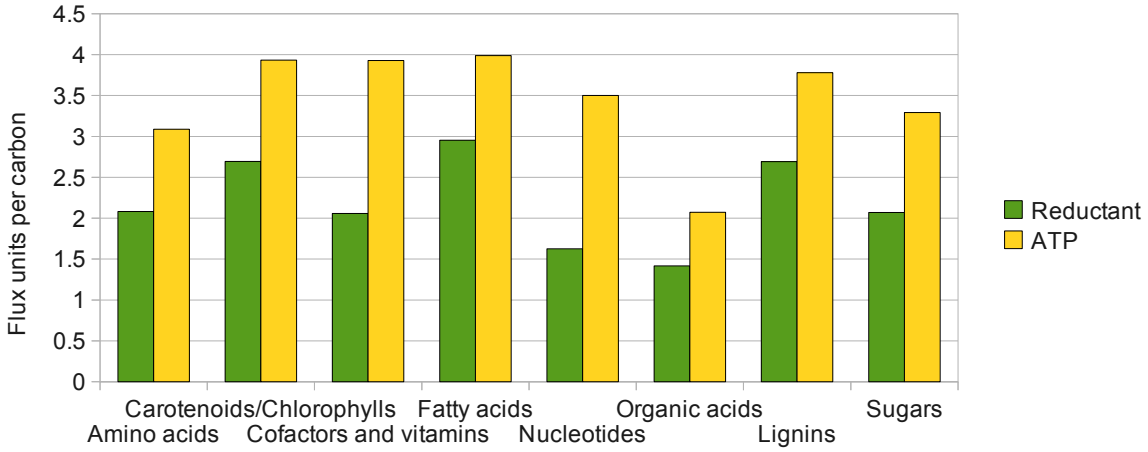


Figure 4.1: Simulated reductant and ATP costs of different groups of substances per carbon. The simulations were performed using only CO_2 as carbon source and providing all reductant and ATP by the DEHOG and ATPase reaction, respectively. The displayed costs are the respective fluxes representing the consumption of ATP or reductant during biosynthesis as unit of flux through ATPase and DEHOG respectively.

4.1.2 Energy consumption

After determining the energetic requirements of single metabolites, the energetic requirements during growth were investigated. To this end, reaction scanning was used to establish the effect of different energy requirements on the system. For night conditions the theoretical optimal starch to biomass conversion rate (considering carbon) is 92.1% which is 0.1% higher than the conversion rate in flux minimisation. This translates into a minimum of respired starch for biomass biosynthesis of 0.688 mg/g or a minimum requirement equal to 0.13 mmolATP/g per night. Maintenance energy in plants however seems to vary substantially, with whole plant maintenance measurements between 7 mg and 47 mg glucose per gram dry weight (gDW) per day according to De Vries [217]. For *Trifolium repens* (white clover - a relative to *M. truncatula*) a consumption of 15 mg/g of glucose per day was measured. Assuming an equal distri-

bution of this maintenance over the day and an 8h night period, a nightly requirement for glucose of 5 mg/g can be assumed. This is roughly 56% of all the starch build up during the day.

Scanning was performed as described in Chapter 2. The biomass composition was set as constraint for the model simulating the changing demands with respect to growth. The biomass formulation for the day conditions took starch biosynthesis (7.9% of the total daily produced biomass) into account, while night conditions considered starch as source. The investigation was performed for both day and night conditions, where either light and carbon dioxide (as sources for energy and carbon, respectively) or starch (as the source for both energy and carbon) was used.

4.1.3 Day conditions

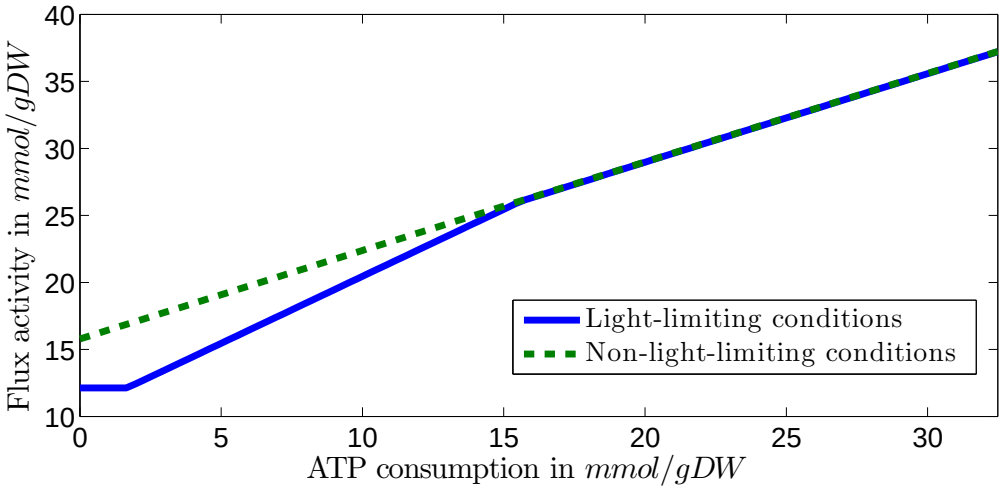
Using photosynthesis as the source of energy, an additional ATP requirement of 23.87 mmol/gDW of produced biomass was found. This requirement was in addition to the minimum feasible ATP produced by photosynthesis of 100.09 mmol/gDW (which results from requirements for reductants - see Section 4.1). The measurements were obtained by first employing a constraint to produce one flux unit of biomass (equivalent to 92.4×10^3 g). The ATPase reaction was allowed to regenerate ATP while DEHOG was turned off. This forces the model to use the light reactions for reductant production while ATP can also be produced by the ATPase. The system was minimised for light consumption and the resulting light consumption fixed. Finally, the ATPase flux was minimised. This forces the system to use a minimal amount of ATP, and allows the distinction between automatic photosynthetic ATP regeneration and additional requirement. An investigation showed that there is a demand for ATP for biosynthesis, which is higher than the amount automatically produced when regenerating reduc-

tant. When scanning for different ATP consumption rates, two different situations are therefore considered.

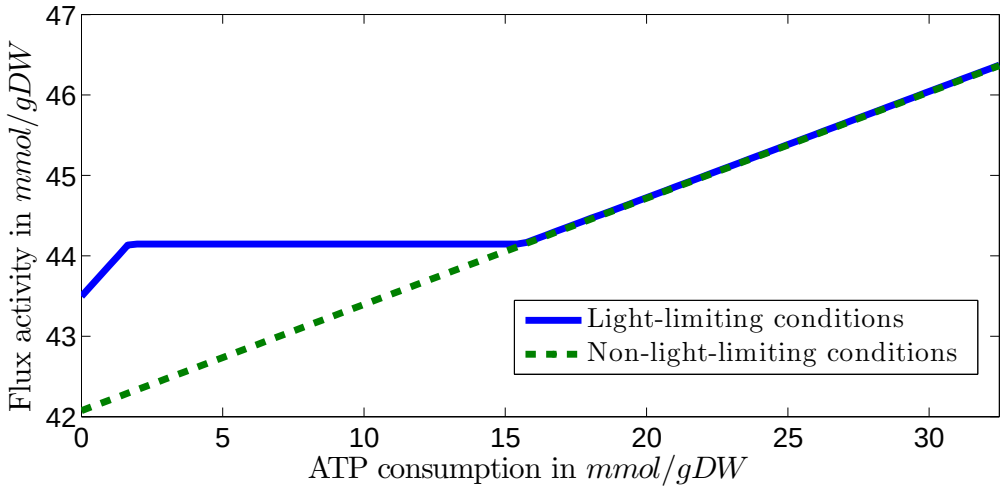
In the first situation, there is plenty of light, and the system does not starve for any nutrient. In this situation, the effects described in Section 4.1 can again be observed. Enough light is taken up to produce a surplus of ATP to minimise the overall flux. This surplus is about 3.65 mmol/gDW ATP which translates to about 0.1 ATP per fixed biomass carbon. As oxidative stresses which have to be tackled by reductants are not considered in the analysis, the surplus can be expected to be higher in reality. This surplus can only be observed, if the system is optimised to yield a solution with minimal total flux.

If however, light is considered to be limited, i.e. the system is minimised for light uptake before an overall flux minimisation, this initial surplus is not present. The light-limiting conditions showed a higher activity in the mitochondrial electron transfer chain and oxidative phosphorylation. When no additional ATP is required, the ATP production in the mitochondria is initially 0.73 ATP per biomass carbon. This is considerably lower in the non-limiting conditions, where about 0.56 ATP per fixed biomass carbon are being produced in the mitochondria. The remaining ATP demand is covered by regeneration resulting from higher photosynthetic activity. The objective differences between both solutions however are rather small. The light-limiting condition shows only a slightly smaller (3.2%) light uptake. In turn the flux-minimal solution is only about 0.01% better than the light-limiting solution, making both solutions generally very similar. In fact, the changes get even smaller with increasing ATP demand.

While in the light-limiting solution light demand constantly rises and the mitochondrial electron transfer chain is increasing accordingly, only the latter is true in the non-light-limiting conditions (see Figure 4.2). At low ATP demand the surplus light



(a) Mitochondrial ATPSynthase activity



(b) Photosynthesis activity

Figure 4.2: Differences between flux minimal and light optimal solutions. In the non-light-limiting condition, additional photosynthesis was employed to provide ATP (wasting reductant) in the plastid which could have been produced more efficiently by the electron transfer chain. This leads to a flat transition phase where the reductant is shuttled to the mitochondria to produce additional ATP.

in the non-light-limiting condition allows the usage of less energy-efficient but more flux-efficient routes. With increasing energy demand however, the additional photosynthetic activity is needed to fulfil this demand. The surplus of energy decreases with

increasing energy demand and the solutions become ever more similar. The differences between both optimisation procedures diminishes to a 0.01% higher light consumption for flux minimisation and 0.001% higher total flux in the light-limiting condition. These adaptations occur at ATP demands which are potentially much lower than the actual maintenance energy demand. This indicates that the plant has evolved to optimise its enzymatic efficiency simultaneously to a most efficient energy production.

4.1.4 Night conditions

Similar to the day conditions the system was initially tested for additional ATP requirements. The additional ATP required for biomass production from starch is 10.83 mmol/gDW or 0.28 ATP per carbon. As expected, the additional ATP requirement is lower than during the day as the available carbon is already fixed. Some biosynthetic processes even provide small amounts of ATP when starch is used as the carbon source (a detailed analysis is provided in Section 4.2.1). This is partially due to the generation of ATP by the TCA when higher amounts of reductant are required. To test how much ATP automatically gets produced during the night due to reductant requirements, DEHOG was allowed to act in reverse and carbon dioxide was restricted to being exported.

The amount of additional reductant required is 3.09 mmol/gDW. It has to be kept in mind that this requirement infers that all starch carbon gets converted into biomass. This leads to an increased ATP requirement of 15.31 mmol/gDW. When minimising for starch usage without additional energy or reductant sources, a respiratory loss of 7.8% is observed. Reductant necessity accounts for roughly 49% of this loss while ATP demand accounts for the other 51%. In addition, it is noteworthy that the obtained solution is most carbon efficient but does retain some energy, which only becomes

available if more carbon is available. During biomass production, energy is lost due to

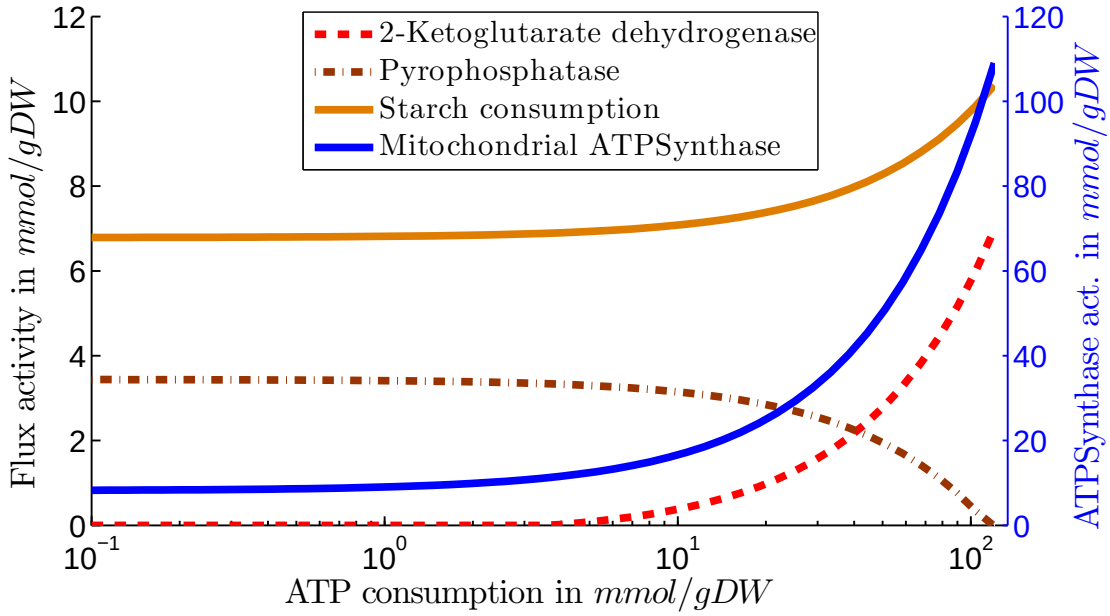


Figure 4.3: Reactions of the metabolic network to increased ATP demand during the night. The TCA does not increase activity at low ATP demands but only reacts to increased demand to supply reductant to the electron transfer chain. With increasing ATP demand, additional starch has to be degraded. This allows the use of pyrophosphate in the fructose bisphosphatase step of glycolysis (not shown). The pyrophosphate surplus (pyrophosphatase activity) from biosynthesis is therefore reduced with increasing starch consumption.

an excess in pyrophosphate, which is represented by a high activity of the inorganic pyrophosphatase reaction. The surplus of PPI originates from polymerisation reactions and loading of carriers for polymerisation. This pyrophosphatase enzyme catalyses the hydrolysis of a phosphate bond without any coupled metabolic activity. The reason for this excess is due to the limitation of fructose 6-phosphate (F6P) in comparison to pyrophosphate (PPI). When additional starch becomes available, more F6P can be formed and the surplus PPI from biosynthesis can be used. Figure 4.3 gives an overview of these observations. By recovering the energy from the phosphate bond in

the fructose bisphosphatase step, this energy can be recovered.

There is an interesting observation, with respect to flux alternate optima in flux minimisation in the biosynthesis route of serine (see Figure 4.4), that was made during this analysis. Two different routes exist for this biosynthesis and it is yet unknown which of them is employed [218]. Interestingly both routes do not only require the same amount of energy, but also need precisely the same amount of flux.

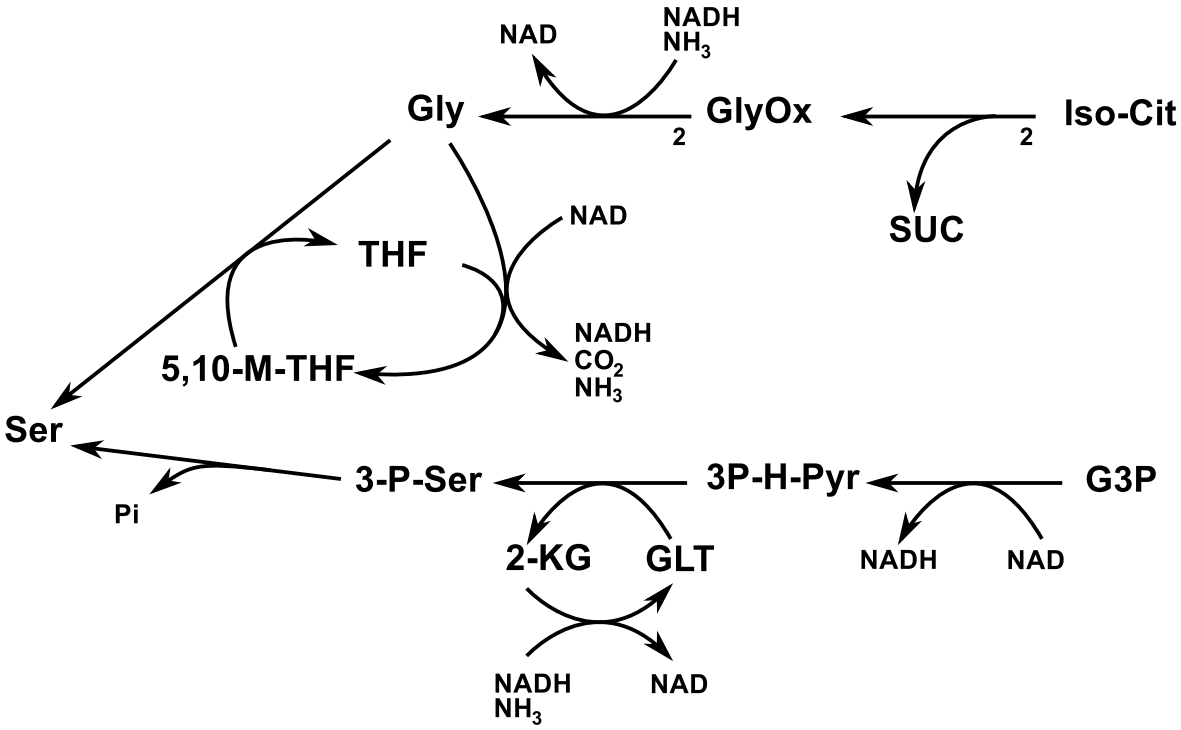


Figure 4.4: Alternative routes for serine biosynthesis. This is a rare example where two different pathways are both equally energy-efficient and, involving additional reactions from the network, also equally flux-efficient.

4.1.5 Discussion

The limit for growth during the night is the amount of available starch and other currency metabolites – with starch being the dominant factor. The obtained biomass data for *M. truncatula* shows a total starch accumulation of about 7.9 mg/g. With

a maximal determined conversion rate of starch to biomass of 92% the theoretical maximal biomass accumulation during the night would reach 7.2 mg/g, only about 7% of the assumed daily growth rate of 100 mg/g. These theoretical maxima do however not take maintenance energy into account, which plays an important role in plants. Comparing the theoretical growth maxima with experimentally observed data of either respiration [19] or substrate consumption [217] implies that calculations, at least during the night, need to take extensive maintenance energy into account. Analysing the cost for different biomolecules and the total energy for biosynthesis of a fixed amount of biomass, it can readily be assumed that the majority of maintenance energy is not growth related.

The ATP to reductant ratio, mentioned above, also plays an important role. This is particularly striking when investigating night conditions, where it becomes obvious that the average biomass molecule is more reduced than starch. This can easily be seen as freely available ATP does not allow a full conversion from starch carbons to biomass carbons. It is however interesting that the biomass production process in itself provides some additional energy which would essentially be lost if not for the non-growth-related energy requirements. Pyrophosphate in particular is produced in many polymerisation reactions (and thereby probably even more abundant from non-growth-related protein and RNA polymerisation). This energy would essentially be lost without the pyrophosphate dependent phosphofructokinase, which can recycle this energy by reintroducing it to the glycolytic pathway.

The alternate optima in serine biosynthesis elucidate possible reasons for the regulation of this pathway. As both pathways are similarly energy- and flux-efficient, none of the pathways is in itself preferable. However, the isocitrate cleavage pathway employs several enzymes that are also involved in other pathways, while the normal serine

Species	Maintenance ($mmol\ gDW^{-1}\ h^{-1}$)
Aerobacter aerogenes	2.3 − 6.8
Escherichia coli	18.9
Lactobacillus casei	1.5
Lactococcus cremoris	1.4 − 18
Saccharomyces cerevisiae	< 1

Table 4.1: Maintenance energies for several microorganisms. Based on Poolman et al. [120] with data from Stephanopoulos et al. [219].

pathway is comparably isolated. This could indicate that the normal serine biosynthesis pathway is preferred if the isocitrate cleavage enzymes are not expressed, while the isocitrate pathway is preferred if the enzymes are available, depending on other parts of the metabolism.

There is another interesting observation that can be made when looking at non-growth-related maintenance requirements. During the night, a plant is very limited with respect to the amount of available nutrients, as starch is the only available carbon source. The numbers of De Vries [217] indicate that the total consumption of starch over a day is in the range between 0.04 mmol/gDW and 0.26 mmol/gDW. Given the observed ATP regeneration by glucose respiration of 32.4 : 1, this leads to an hourly maintenance from $0.053\,\mathrm{mmol\,g^{-1}\,h^{-1}}$ to $0.353\,\mathrm{mmol\,g^{-1}\,h^{-1}}$. This, however, is far from maintenance energies in prokaryotic systems, which are about 5- to 100-fold larger (see Table 4.1). Interestingly it is rather close to the estimations for *S. cerevisiae*, which indicates that eukaryotic organisms require less maintenance or evolved a more efficient maintenance system than prokaryotes. In contrast, when heterotrophic growth of *A. thaliana* was investigated by Poolman et al. [120], they found a maintenance flux of $7.1\,\mathrm{mmol\,g^{-1}\,h^{-1}}$, which is closer to the range of the listed prokaryotic systems.

In planta these values, however, would mean that per hour an equivalent of about $39.4\,\mathrm{mg\,g^{-1}\,h^{-1}}$ of glucose would be needed for maintenance. In an eight hour night, this would amount to 315 mg/g or 30% of the total plant dry weight.

With maximal starch amounts of about 80 mg/g [220] (assuming a FW/DW ratio of 10:1) this would pose a serious problem for the plant. Extensive data from Pyl et al. [19] shows that *A. thaliana* respiration ranges from 18 mg to 37 mg carbon per gram DW or 1.1 mg/h to 2.3 mg/h during a 16h night. With a sugar mass composition of about 2:3 (C:O+H) this means a consumption of about 2.75 mg to 5.75 mg of sugars per hour. This would suggest an ATP maintenance from $0.5\,\mathrm{mmol\,g^{-1}\,h^{-1}}$ to $1\,\mathrm{mmol\,g^{-1}\,h^{-1}}$, which is still much lower than the estimated maintenance for heterotrophic growth. With the present data estimates would be even lower, as there is much less sugar, which could supply the required energy. With only about 10 mg/g and eight hour nights the maximal amount of available ATP maintenance energy is $0.22\,\mathrm{mmol\,g^{-1}\,h^{-1}}$.

With these observations it seems necessary to evaluate the difference in maintenance costs between heterotrophic and phototrophic growth, and indeed between whole plant systems and plant cell cultures. This is particularly important as current data shows a strong discrepancy between these costs indicating that cell culture maintenance costs are closer to bacterial maintenance costs than to plant maintenance.

4.2 Nitrogen metabolism analysis

Nitrogen in the biosphere is mainly supplied as either nitrate or ammonium. While nitrate is highly oxidized, ammonium is highly reduced. In fact, from a metabolic point of view the only difference between nitrate and ammonium is the amount of reductant required for its incorporation. However, as mentioned earlier, there is no way that ATP can be converted into reductant and therefore, under physiologic conditions, the

effect of changing the nitrogen source is not obvious. In all of the above simulations the model used ammonium, while not using nitrate at all. Since nitrogen metabolism is highly important for symbiotic nitrogen fixation an analysis of the different costs for ammonium and nitrate was performed (presented in Section 4.2.1). Furthermore, effects of increasing costs of ammonium and different availability of ammonium were assessed (see Section 4.2.2)

4.2.1 Biosynthesis costs of nitrate vs ammonium

Two approaches were taken to assess cost differences of biomolecules using either ammonium and nitrate as source of nitrogen. First, cost differences under night conditions were assessed using starch as the carbon source. Interestingly a set of compounds was found which can be produced for "free", i.e. without loss of CO_2 directly converting starch carbon into biomass carbon. Table 4.2 gives an overview of different compounds and the changing costs using nitrate or ammonium. The costs were determined by evaluating the minimal amount of starch required for biosynthesis, and providing additional energy and reductant from the ATPase and DEHOG reactions as required.

Unfortunately, this method was not applicable for the comparison of starch and the combination of light and CO_2 as the sources of energy and carbon. Since light only provides reductant and ATP, the minimal amount of light required would always be zero if DEHOG and ATPase were able to provide reductant and ATP. To determine the differences between light and starch conditions, fluxes for both reactions were set to zero and the system was investigated with the goal of minimising either light uptake (i.e. photosynthetic activity) or starch consumption.

Carbon fixation lacks small amounts of ATP for each fixed carbon. However, when additional reductant is needed, this lack turns into a surplus, as seen in the day con-

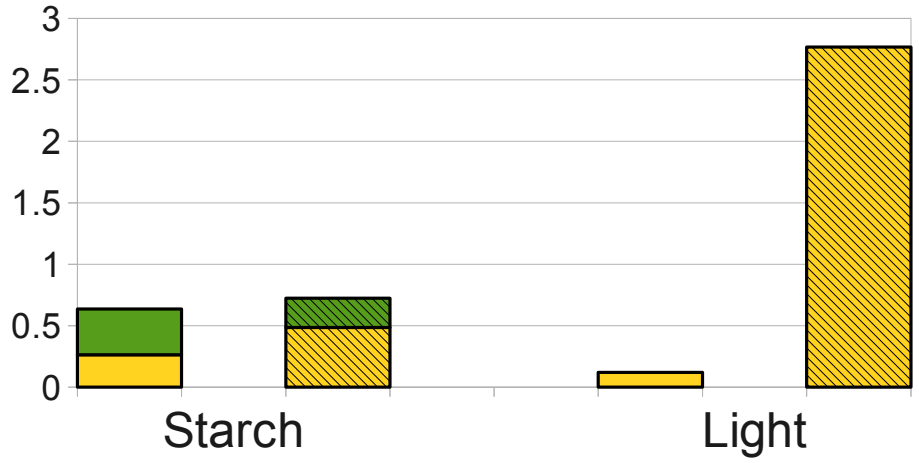


Figure 4.5: Average surplus of reductant (green) and ATP (yellow) using ammonium (clear) or nitrate (hatched) as nitrogen source over all exportable biomolecules (units in flux/molecule). ATP and reductant were calculated using either light or starch as the sole source of energy. The values are obtained as fluxes through DEHOG and ATPase respectively.

ditions in Figure 4.5. Therefore it can be assumed that there is surplus ATP during the day, while no reductant surplus will be observed. With the exception of a few molecules (marked as 'free' in Table 4.2), no surplus is expected for starch. Figure 4.5 depicts the strong increase in ATP surplus on photosynthesis when nitrate is used as nitrogen source. There is also a small increase in surplus ATP for biosynthesis using starch. This is contrasted by a decrease in surplus reductant. However, additional starch needs to be respired to satisfy the reductant demands. Since starch respiration can also provide additional ATP, a small increase in available ATP is expected for those compounds for which biosynthesis already provides some ATP. It has to be noted that the decrease in available reductant corresponds to an ATP amount that is larger than the increase in available ATP. Therefore, even though more starch has been used, the amount of free energy from biosynthesis has decreased.

Table 4.2: Costs in ATP and reductant for the biomass components of *M. trunctula*

Compound class		Growth on ammonium			Growth on nitrate		
		Reductant	ATP	Free	Reductant	ATP	Free
Amino acids	Alanine	0	-1.21	✓	4	-1.21	
	Arginine	0	1.07		15	3.57	
	Asparagine	-1.71	0	✓	6	0.71	
	Aspartate	-2	-0.62	✓	2	-0.62	
	Cysteine	3	3.29		7	3.29	
	Glutamine	-0.34	0	✓	7	1.64	
	Glutamate	-0.74	0	✓	3	0.64	
	Glycine	-0.52	0	✓	3	1.19	
	Histidine	0	1.57		10	6.57	
	Isoleucine	3	0.57		7	0.57	
	Leucine	3	3.57		7	3.57	
	Lysine	2	-0.43		10	-0.43	
	Methionine	5	6.14		9	6.14	
	Phenylalanine	2	3.36		6	3.36	
	Proline	1	0.64		5	0.64	
	Serine	-1	-0.21	✓	3	-0.21	
	Threonine	0	1.38		4	1.38	
	Tryptophan	1	8.05		9	8.05	
	Tyrosine	1	3.36		5	3.36	
	Valine	2	-0.36		6	-0.36	
Carotenoids and chlorophylls	Chlorophyll a	37	60.24		53	60.24	
	Chlorophyll b	39	60.24		55	60.24	
	all-trans-lycopene	28	35.81		28	35.81	

Table 4.2: (continued)

Compound class	Common name	Growth on ammonium			Growth on nitrate		
		Reductant	ATP	Free	Reductant	ATP	Free
Cofactor and vitamins	S-adenosylhomocysteine	0	12.83		23	15.33	
	Coenzyme-A	2	21		30	21	
	Flavin adenine dinucleotide	0	9.07		29	26.57	
	Heme	24	50.67		40	50.67	
	Pyridoxal phosphate	0	5.31		23	17.81	
	Tetrahydrofolate	5	31.14		21	31.14	
	Thiamine diphosphate	0	2.43		4	2.43	
Fatty acids	Linoleic acid	18	20.71		18	20.71	
	Linolenic acid	19	20.71		19	20.71	
	Oleic acid	17	18.71		17	18.71	
	Palmitic acid	14	16.52		14	16.52	
	Stearic acid	16	18.71		16	18.71	
Sugars	Fructose	0	0.57		0	0.57	
	Cellulose*	0	1.57		0	1.57	
	Pinitol	1	4.67		1	4.67	
	Glucose	0	0	✓	0	0	✓
	Arabinose	0	4.48		0	4.48	
	Fucose	1	1.57		1	1.57	
	Galactose	0	1.57		0	1.57	
	Mannose	0	1.57		0	1.57	
	Myo-inositol	0	0.57		0	0.57	
	Rhamnose	1	1.57		1	1.57	
	Sucrose	0	2.14		0	2.14	
	Xylose	0	4.48		0	4.48	

Table 4.2: (continued)

Compound class	Common name	Growth on ammonium			Growth on nitrate		
		Reductant	ATP	Free	Reductant	ATP	Free
Nucleotides	Adenosine triphosphate	-1.02	0	✓	15	9.95	
	Cytidine monophosphate*2	-0.52	0	✓	9	6.19	
	Deoxy-adenosine monophosphate*2	-0.02	0	✓	16	9.95	
	Deoxy-cytidine monophosphate*2	0	1.19		10	6.19	
	Deoxy-guanosine monophosphate*2	-0.22	0	✓	15	11.95	
	Guanosine monophosphate*2	-1.22	0	✓	14	11.95	
	Thymidine monophosphate*2	0	5.45		7	7.95	
	Uridine monophosphate*2	-0.52	0	✓	5	6.19	
Pectins	Coniferyl alcohol	7	8.95		7	8.95	
	Coumaryl alcohol	5	4.86		5	4.86	
	Sinapyl alcohol	9	13.05		9	13.05	
Organic acids	Ascorbic acid	-0.97	0	✓	-0.97	0	✓
	Citric acid	-3	-1.43	✓	-3	-1.43	✓
	Malic acid	-2	-0.62	✓	-2	-0.62	✓
	Oxalacetic acid	-3	-0.62	✓	-3	-0.62	✓
	Succinic acid	-1	-1.62	✓	-1	-1.62	✓
	Galacturonic acid	-1.37	0	✓	-1.37	0	✓

Table 4.2: Costs of metabolites based on starch as carbon source. The costs are split between reductant costs (NADPH to NADP) and ATP costs (ATP to ADP + Pi). Costs were determined using a 1:1 starch carbon to metabolite carbon conversion. Positive values indicate an energy requirement during biosynthesis while negative numbers represent a surplus being available.

* Values provided are for one six-carbon monomer of a cellulose polymer.

*2 Nucleotides are produced as triphosphates, but are exported as monophosphates since they are incorporated into DNA and RNA as monophosphates

4.2.2 Effects of increasing ammonia cost and changes in the abundance of ammonium and nitrate

Ammonium is the more cost-efficient option when incorporating nitrogen into the biomass. As expected, ammonium is therefore preferred over nitrate irrespective of the optimisation method used, as it is both more flux-efficient and energetically cheaper. However, in nature there is only a limited supply of both sources, and it is possible for the plant to run out of either. Alternatively, it can become very difficult for the plant to reach ammonium. With flux minimisation aiming at optimising metabolic efficiency this change in bioavailability can directly be taken into account by increasing the cost of a flux as described in Section 2.2.2. This simulation reflects the situation that, while ammonium is in theory available, it becomes more and more costly to acquire. A second approach is to limit the actual flux through the ammonium uptake, which would reflect the situation that there is only a restricted total amount of ammonium available. There is however a strong distinction to be made; in the first approach, a necessary assumption is that the growth is not limited by the amount of nutrients available, but rather by a general maximal uptake or processing of metabolites. This is necessary to allow an interpretation of the simulations with respect to the changing environment. In the second approach a restriction on the total amount of ammonia as well as starch is assumed. This reflects the assumption that enough energy for maximal growth on ammonium is available, but the increased cost of nitrate cannot be satisfied. During the day a difference in growth can only be observed if photosynthesis cannot provide enough reductant. However during the night, different aspects of the metabolism are likely to influence the potential growth.

4.2.2.1 Scanning the effects of increasing ammonium costs

When performing a weight scan on the ammonium uptake flux the network responds to increased weight in multiple ways (displayed in Figure 4.6). The pentose phosphate pathway is activated, as more nitrate needs to be reduced. A decline in mitochondrial ATP production by a reduction of the electron transfer chain activity can be observed. In addition, the system switches from exporting positively charged protons when taking up ammonium to exporting negatively charged bicarbonate when importing nitrate. The proton export is in good correspondence with the general observation that ammonium nutrition leads to a decrease in soil pH [221]. The bicarbonate export is in agreement with a slight increase of pH on nitrate nutrition [221], with bicarbonate commonly acting as mild base. This export and activity change is accompanied by a drop in the initial TCA steps in the mitochondrion. However, this decrease is balanced by an increase in the activity of plastidal isozymes to the corresponding mitochondrial TCA enzymes. Overall there is an increase in TCA activity tightly coupled to the increase in nitrate usage. This increase is supplemented by an enhanced glycolytic activity (see Figure 4.7).

While low nitrate usage leads to increased activity in all glycolytic enzymes, higher nitrate usages only affects the glycolytic reactions following the triose-isomerase step. The increased demand for reductant leads to a diversion of glucose from glycolysis into the pentose phosphate pathway (PPP), as it is a faster route for reductant regeneration. This redirection leads to a decrease in the initial reactions of glycolysis from glucose-6-phosphate to the triosephophates, as glucose is diverted towards the PPP. In contrast, the ATP and reductant producing steps of glycolysis are increasing their activity. They are driven by the product from the PPP, glyceraldehyde-3-phosphate

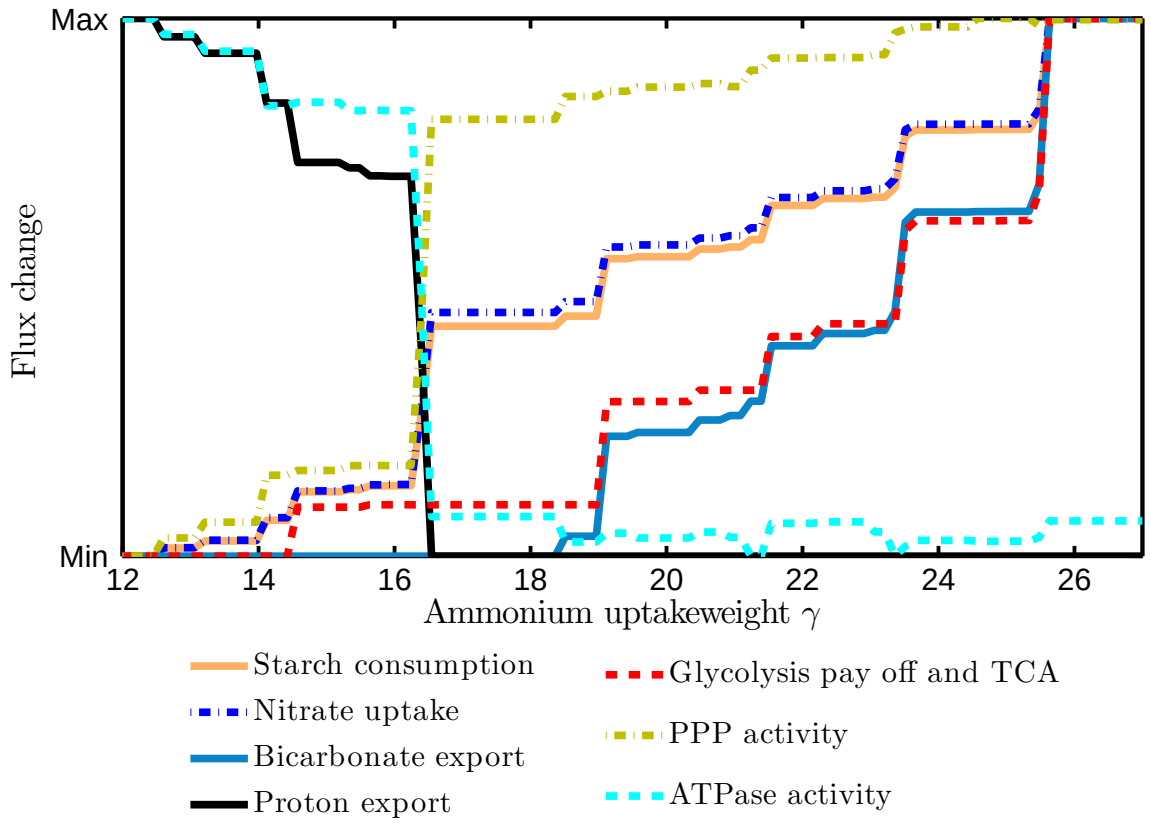


Figure 4.6: Changes of different subsystems with respect to a changing ammonium uptake weight. The objective is flux minimisation. Changes are displayed as relative changes between minimum (0) and maximum (1). The increasing starch uptake allows for a reduced mitochondrial ATP production (Complex V) due to additional ATP from the TCA and glycolysis. Additional reductant necessary for nitrate reduction is provided by the pentose phosphate pathway (PPP) as this is the faster route for reductant generation.

(GAP) which is metabolised to pyruvate. However, the pyruvate is not transferred to the mitochondrion. Instead it is directly relocated to the plastid and there used to obtain additional reductant for nitrate reduction, due to the increased demand for reductant.

While in theory, the PPP is unable to provide more reductant than glycolysis and

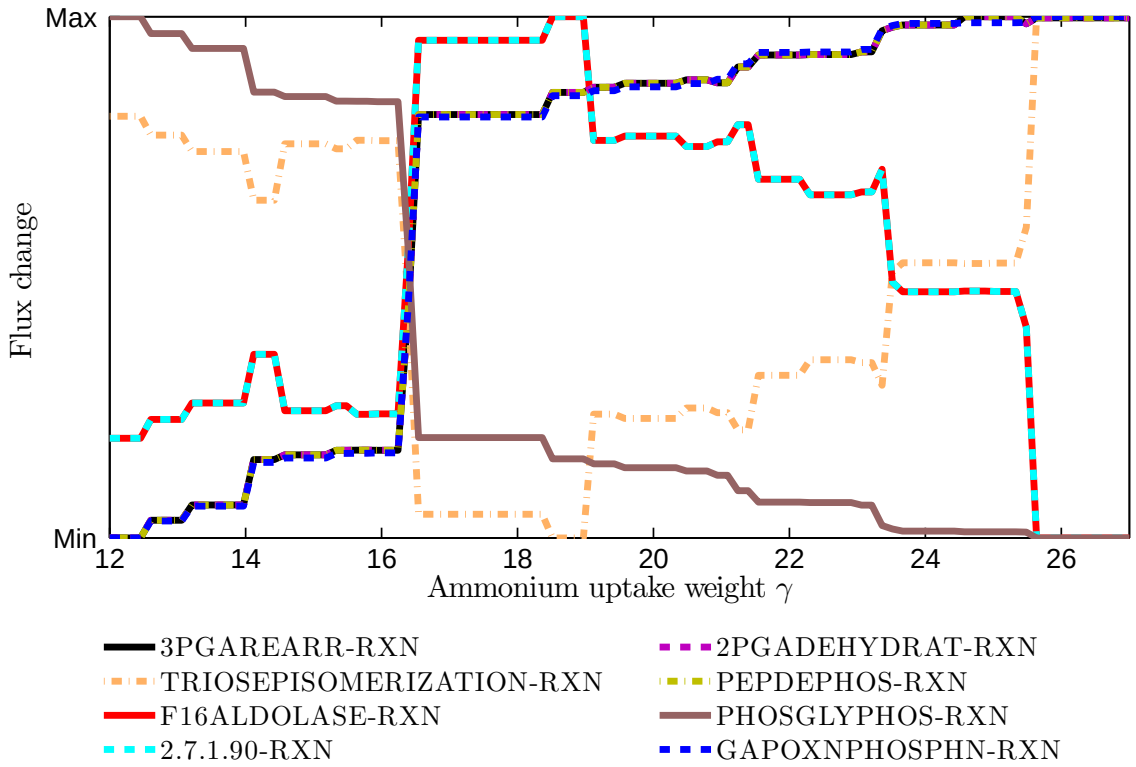


Figure 4.7: Change in the activity of glycolytic reactions with changing weight for ammonium import. Initially the starting reactions of glycolysis show increased activity which is soon reaching a maximum. With more nitrate being used (see Figure 4.6 this increase stops and the necessary glucose 6-phosphate is diverted to the PPP. Only the reactions downstream of glyceraldehyde 3-phosphate keep increasing, producing both reductant for nitrate assimilation and ATP. Labels are the MetaCyc identifiers for the respective reactions.

TCA, it is able to provide the respective reductant using fewer reactions, but at the same time causes a loss of energy which could otherwise be used for ATP biosynthesis. In both PPP and the combination of TCA and glycolysis, six reductant equivalents are extracted but the PPP provides those reductants in fewer steps. It is therefore more readily available and, adding to this, provides the reductant directly in the plastid where it is required for nitrate reduction. The latter is also the reason for the relocation of the initial TCA reactions, which allow a direct usage of reductant in the plastid for

nitrite reduction. The downside of this faster delivery is that the PPP does not provide any ATP; however using the aim of flux minimisation, the enzymatic usage is more efficient, and could finally save energy that would otherwise be necessary to synthesize additional enzymes.

4.2.2.2 Restricted ammonium uptake

The ammonium weight scan reflects the situation that ammonium is becoming more difficult to reach, and allows the system to distribute the fluxes. To investigate the effects of restricted ammonium a flux scan of the ammonium importer was performed. After initial optimisation for flux minimisation, and no restriction on starch, it was obvious that the same overall effects as for the weight scan were found again:

- Increased PPP activity when less ammonium was available;
- Higher starch requirements when more nitrate had to be assimilated; and
- A slow turn down of the mitochondrial electron transfer chain, in response to the increased availability of ATP from starch degradation.

To further investigate possible effects of nitrate the system was studied under the assumption of a restricted available amount of starch and a changing amount of available ammonium. This changes the assumption of readily available nutrients to a situation where starch is becoming limiting and ammonium is depleting. The first observation is that, as expected, the maximal biomass amount producible on nitrate is only 85% of the amount producible using ammonium. This change in biomass leads to closely linked changes in multiple biosynthetic pathways. Investigating these changes is futile as they are directly associated with the change in biomass and do not provide any additional information.

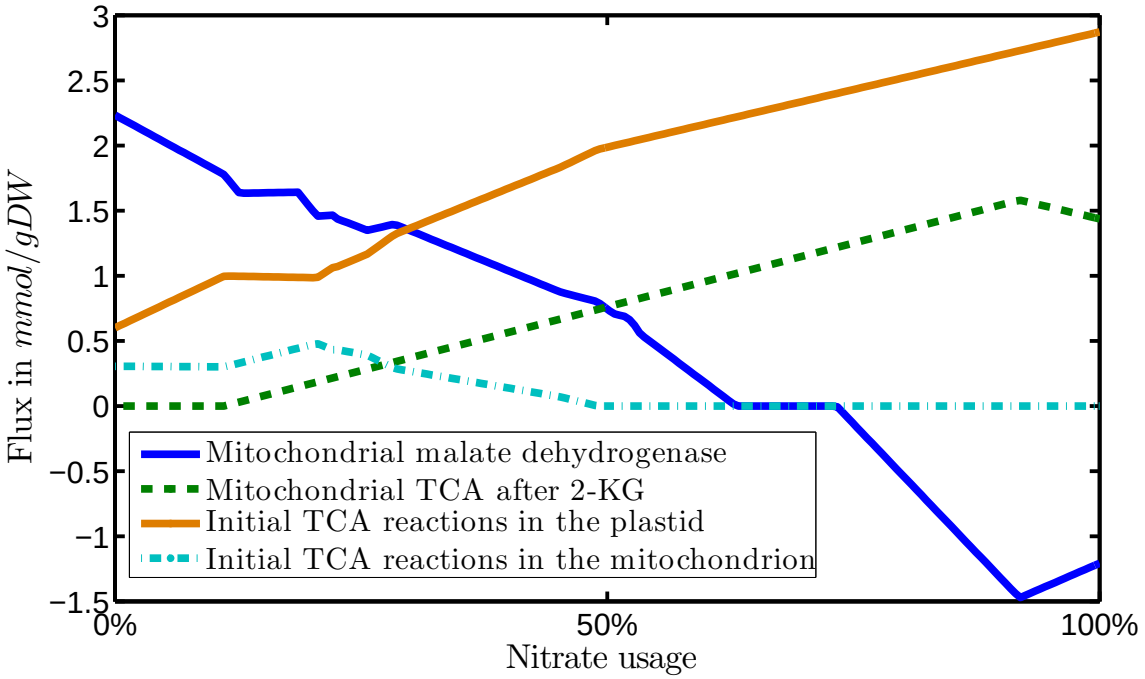


Figure 4.8: Reaction of the TCA to restricted ammonium supply under growth limited conditions. Data is shown with respect to the percentage of nitrogen from nitrate. When plenty of ammonium is available, only the initial TCA reactions in the mitochondria are active for precursor production and malate dehydrogenase is used to provide reductant for the electron transfer chain. The increasing reductant demand in the plastid leads to a redistribution of flux from the mitochondrial TCA to the plastid. Only the TCA reactions that are exclusively present in the mitochondrion carry increased flux. The reductant generated by these reactions is shuttled to the plastid.

There is an effect with respect to flux redistribution of the TCA from the mitochondrion to the plastids, which shows similarities to the observations in the ammonium weight scan. However, in contrast to the flux limiting situation in the flux weight scan, the PPP is not used as this would lead to a waste of ATP and thereby carbon. The TCA in the mitochondrion is split into two parts: one from oxalacetic acid to 2-ketoglutarate (2KG) and a second from 2KG to fumarate (see Figure 4.8). This is mainly due to the fact that several initial TCA reactions are also present in the plastid, and are active there to provide the additional reductant necessary for nitrate

assimilation. Overall, the TCA activity is continually increasing, due to the increasing demand for reductant. Simultaneously, ATP regeneration from the electron transfer chain in the mitochondrion is reduced, freeing more reductant for nitrate assimilation; a situation similar to that under non-carbon-limited conditions.

4.2.3 Discussion

Nitrate and ammonium place different energetic requirements on the metabolism. While ammonium can readily be incorporated in many reactions, nitrate has to be reduced before it can be used by the plant. This makes synthesis of metabolites more expensive when nitrate is used as a substrate, but it may also function as a sink for reductants if such a sink is necessary.

The scans presented in this section consider two distinct situations. In the first scan the underlying assumption was that the plant is not working under growth optimising conditions but rather tries to quickly adapt to a lack of ammonium. The second approach was more related to a situation in which the carbon source, as opposed to the nitrogen source, is the growth-limiting factor. The main difference between these two conditions is apparent in the activity of the PPP. Using the PPP, the system can achieve a fast adaptation to the change in ammonium availability. Only a few fluxes need to be activated to obtain more reductant at the expense of a little more starch. In the present simulations, the model predicts an exclusive uptake of ammonium if both ammonium and nitrate are abundant, due to the considerably lower amount of reductants needed when comparing ammonium incorporation to nitrate incorporation. This result is in agreement with experimental observations that nitrate uptake can be inhibited when ammonium [222] is available. Thus, in an evolutionary context, the model suggests that the reason for this inhibition is an increased energetic efficiency.

Simultaneously, proton export is observed, which is in accordance with soil acidification when ammonium fertilizers are used [223]. If ammonium is limited, the modelled fluxes change gradually to an increased uptake of nitrate, which are accompanied by a higher energy and reductant demand. This demand is met by increased respiration during the night. The dual uptake of nitrate and ammonium can be beneficial to keep the growth more constant and would allow the system to maintain a more stable internal redox state. If however the plant is strongly limited in carbon (due to short days, or very low light during the day), it might be desirable to use all available energy. In nature, an intermediate state is likely to be observed. While it is unlikely that there is a rapid switch in the amount of available ammonium or nitrate, the system could have to adapt to more available nitrate (when fertilizer is applied) making it necessary to provide additional reductant. In these instances an increase in PPP activity would be expected, while it would not be expected during normal growth. Since light is freely available during the day, a sufficient amount of reductant can easily be produced in the plastid. While this can be observed during the day, it is likely that the plant will initialise a shuttling of reductants within the cells to distribute the restricted amount of reductants during the night. As metabolite shuttling is commonly associated with some energetic cost, it can be beneficial to use the PPP instead of trying to shuttle the reductants produced elsewhere.

4.3 Different nutrient solution properties

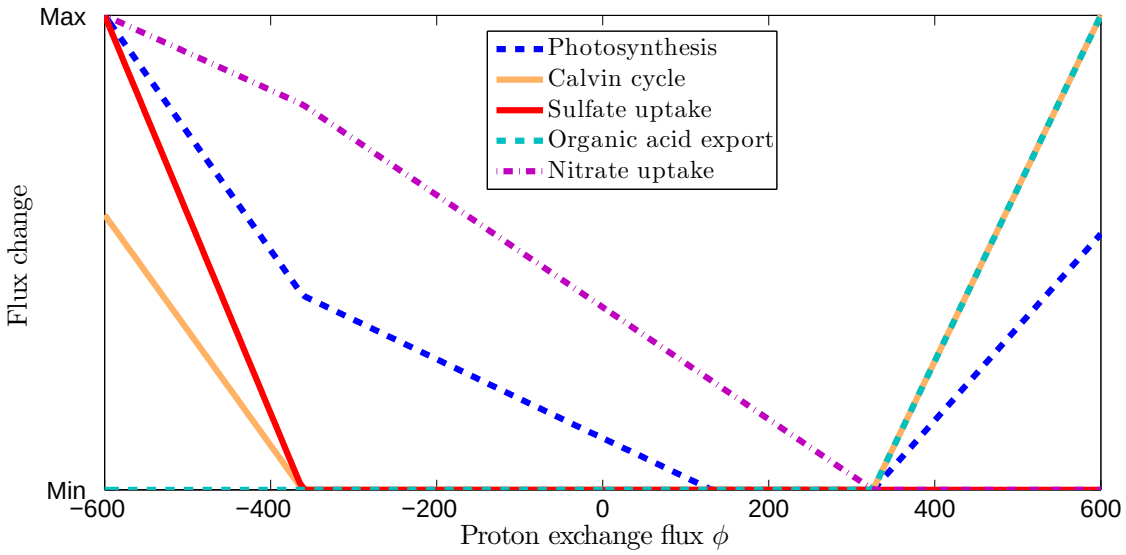
As shown in the ammonium uptake scans, the plant is able to modify its surrounding pH by the choice of nutrients used during biosynthesis. However, in these instances it was assumed that the plant is free to exchange protons, and thereby charges with its environment. In nature however, the plant commonly has to adapt to different

environments. The following describes an investigation of changing proton pressure and draft, which represents low or high pH conditions respectively.

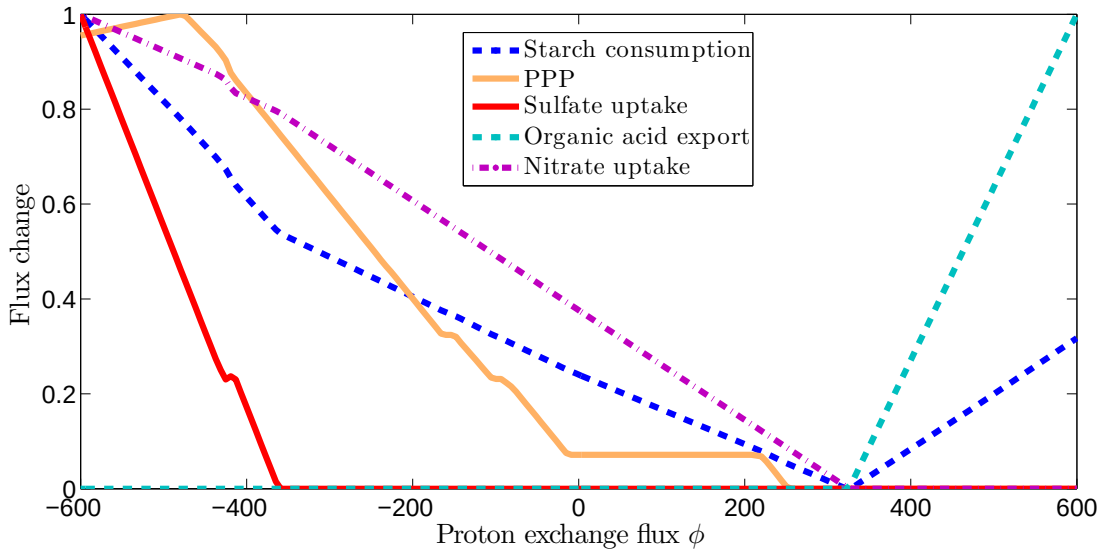
4.3.1 Soil acidity

As mentioned in Section 2.1.2.1, the model was curated to achieve a complete internal atomic and charge balance. This is crucial for the following investigations, as it allows an investigation of the effects of charge pressure, which were employed to mimic different environmental pH levels. To this end a flux scan of the proton exchange transporter was employed, using ϕ as scan parameter, while the bicarbonate exporter was turned off. Negative values of ϕ represent a push of protons into the system, thus simulating the effects of an acidic environment. Correspondingly, a positive value represents a pull of protons out of the system over the cell membrane, mimicking an alkaline environment. Flux minimisation was used as objective for the analysis, and flux distributions that achieve a net charge balance of zero were determined. The analysis was performed for phototrophic growth (day conditions) and for growth on starch (night conditions).

The resulting flux responses to changes in the proton flux ϕ are depicted in Figure 4.9. Under both conditions, the response to changes in acidity displays three characteristic phases. For low pH conditions (large negative values of ϕ, $\phi < -358$), the forced influx of protons is extremely large and cannot be compensated by a simple exchange of nutrients. These values reflect a situation in which establishing an electrochemical gradient across the membrane to protect the cell is not a viable option. To balance the influx of positive charges, the system imports additional negative charges in the form of nitrate (pink dash-dotted lines) and sulphate (red solid lines). However, this excess material has to be converted into some neutral substance which can be ex-



(a) Day conditions



(b) Night conditions

Figure 4.9: Adaptation of metabolic fluxes to a simulated change in ambient acidity. On the x-axis, the value ϕ is varied and represents the fixed influx (negative) or outflux (positive) of protons. Fluxes on the y-axis are normalised as in Figure 4.6. The top panel presents simulation results for phototrophic growth, the bottom panel presents changing fluxes under night conditions (growth on starch).

ported or stored. Under the employed optimality condition, cysteine (and under night conditions a small amount of glycine) is produced and exported, because it can be produced with minimal effort (minimal overall flux). This wasteful production of cysteine, necessary to avoid accumulation of positive charges, requires additional carbon. A consequence is an increase in photosynthetic activity and concomitant increase of carbon fixation (day, blue dashed and solid salmon lines in Figure 4.9a, respectively) or starch consumption (night, blue dashed line in Figure 4.9b). In the second phase, moderate charge input or output ($-358 < \phi < +325$), accumulation of charges is avoided by adapting the ratio of nitrate versus ammonium import. A lower net import of positive charges is balanced in this phase by a reduced import of nitrate (pink dash-dotted) and a simultaneous increased import of ammonium (not shown). The model suggested that it should be possible for the plant to react to a changed proton abundance by using different nitrogen sources in this phase.

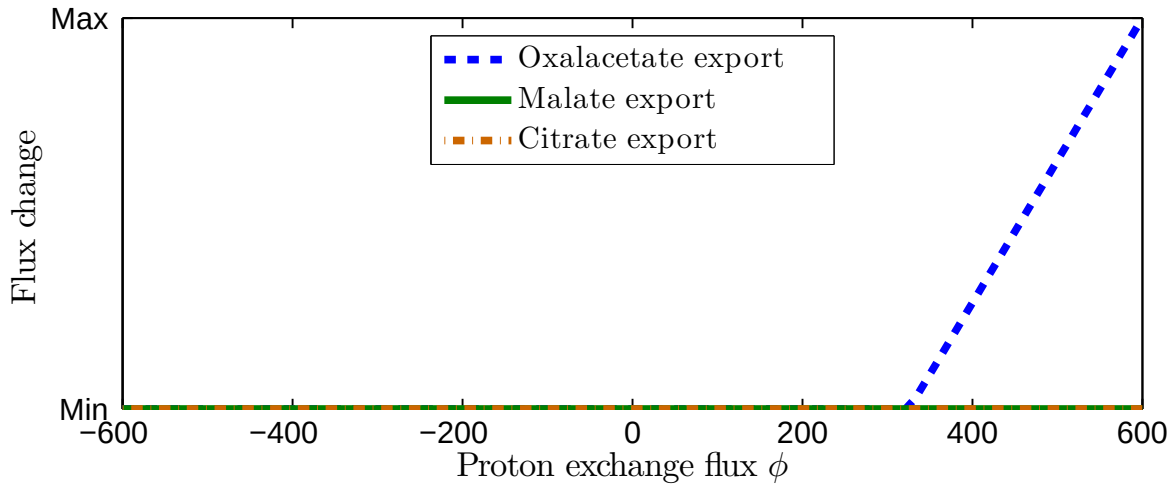
In conditions forcing protons into the cell ($\phi < 0$), nitric acid is imported (HNO_3, in the model as combination of H^+ and NO_3^-) and thereby the environment is de-acidified to some extent. Ammonium is imported as NH_4^+ and the overall biomass is charge balanced. Therefore, with each ammonium ion a surplus of positive charge is imported. As a consequence, nitrate is imported for conditions forcing protons out of the cell ($\phi > 0$). An influx of nitrate is observed for values $\phi < +325$, because approximately 325 nitrogen atoms are required to build one unit of biomass. The observation that in day conditions the photosynthetic flux remains constant for $+125 < \phi < +325$ is an artefact resulting from the fact that flux minimisation tends to overestimate the energy requirements (see Section 4.1.2). In this phase, the surplus energy is used for shorter, but less efficient routes. This artefact can be avoided by first minimising the flux through photosynthesis and subsequently minimising the overall flux. The overall

picture remains the same, but the photosynthetic flux reaches a clear minimum at $\phi = +325$.

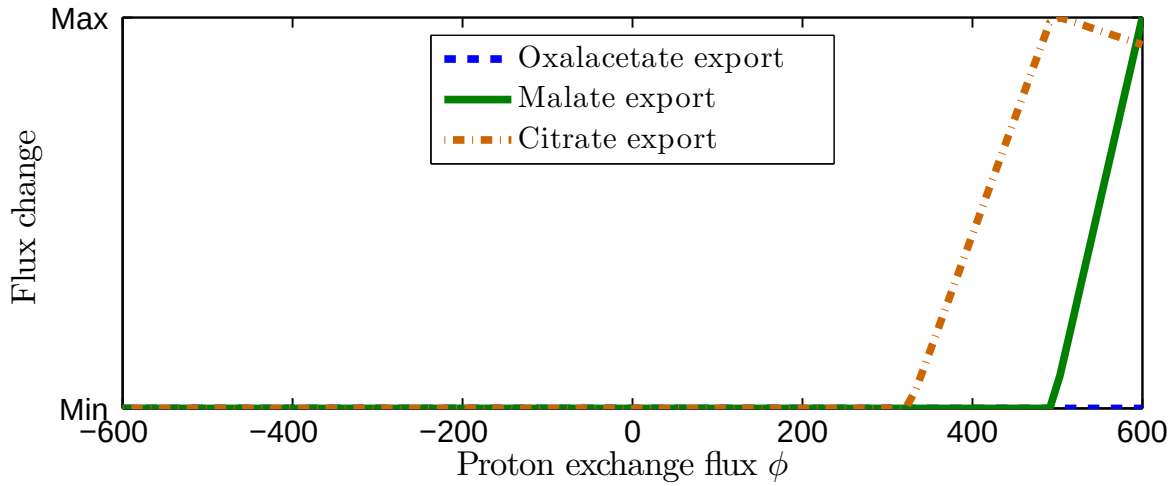
Finally, if the efflux of protons is increasingly large ($\phi > +325$), the loss of positive and concomitant accumulation of negative charge is countered by the production and excretion of negatively charged metabolites. Under the optimality criteria applied here, carboxylic acids are the least costly (in terms of flux) method to achieve this (dashed cyan lines). As for the export of cysteine for extreme values of proton influx, this production is accompanied by an increased carbon demand. During the day, this carbon is provided by an increased photosynthetic rate (red in Figure 4.9a); during the night by a higher consumption of starch (red in Figure 4.9b).

Overall, the responses are similar during the day and the night. However, two interesting differences can be observed. During the night, proton import leads to a higher activity of the pentose phosphate pathway (blue line in Figure 4.9b). This pathway delivers the reductants that are required because of the higher nitrate to ammonium uptake ratio. However, for extreme values ($\phi < -425$), the mechanism is different. Here, additional sulphate uptake is accompanied by high starch consumption for the production of cysteine. In contrast, during the day the additionally required reduction is derived directly from photosynthesis (larger slopes of the red line in Figure 4.9a for the extreme values $\phi < -358$ and $\phi > +325$). In these extreme regimes, the Calvin cycle displays increased activity to provide the carbons which have to be exported to avoid charge accumulation.

The second difference is revealed by inspecting the particular carboxylic acids which are produced and exported (see Figure 4.10). Whereas during the night citrate and malate are the compounds produced with minimal flux cost, oxalacetic acid is less costly during the day. During both day and night, the exported compounds are derived from



(a) Day conditions



(b) Night conditions

Figure 4.10: Export of organic acids during proton scanning. During the day (a) the only exported acid is oxalacetic acid, which is the organic acid with the least remaining energy per charge ratio. During the night (b), as carbon is not assumed to be limiting, citrate and malate are exported, which could still provide some energy. This energy however is more flux expensive than uptake of additional starch.

triose phosphates through the reverse PEP carboxylase. The main reason for this difference is based in flux minimisation. During the night, with fixed carbon in the form of starch, it is cheaper to export malate and citrate. However, during the day

oxalacetic acid is exported, as oxalacetic acid is the most oxidised form of dicarboxylic acids and the reduced amount of reductants for the same amount of charge is preferable.

4.3.2 Discussion

Plants are adaptable to a large range of pH conditions and the present simulation highlights potential mechanisms for this adaptability. While proton export was observed for ammonium uptake, a forced efflux of protons induces the uptake of ammonium. The high level of curation, which ensured that all reactions are precisely charge balanced, allowed the study of metabolic flux rearrangement to counter an accumulation of charged particles. Simulating different soil acidities by proton pressure indicated a possible cause for altering the plants uptake ratio of ammonium versus nitrate uptake beside availability. Depending on the environmental pH, proton gradients will lead to diffusion of positive charges into or out of the cell. The plant has to compensate the charge imbalance imposed by the proton transporter and the model suggests that one way could be the uptake of charged inorganic nutrients (like sulfate or nitrate).

For a wide range of conditions, a simple solution is provided by correspondingly adapting the ratio of assimilated nitrate (negative) versus ammonium (positive). This observation can explain why the presence of ammonium in many plants [224] inhibits nitrate uptake but does not turn it off completely, while from an energetic consideration, nitrate should not be used at all. It is interesting to note that the simulated conditions fit well to the observation that nitrate nutrition leads to higher growth rates in low pH media than ammonium nutrition [225].

At the same time, limitations of FBA when interpreting computational results simulating changes in environmental pH have to be considered. In particular, ionic barriers commonly employed by plants cannot be modelled in FBA. Clearly, the model

predictions should be interpreted with care, in particular regarding the balancing of charged particles. The model does not include osmosis and the import and export of inorganic ions, such as potassium, calcium or chloride. It is natural to assume that, to a limited extent, import and export of inorganic ions will counter the electrochemical gradient over the cellular membrane, and thus avoid a continuous flux of charges. However, the capacity to accumulate or release inorganic ions is clearly limited, and diffusion in the ambient environment makes it unlikely that a flux of these counter ions can be the sole mechanism to adapt to a prolonged exposure to extreme pH values.

The model results, while not necessarily exactly reflecting the true conditions, provide interesting hypothetical solutions on how to cope with externally forced fluxes of charges across the cellular membrane in a sustained manner. The model therefore predicts the metabolic reaction to these conditions, while it does not incorporate other physiological processes. Thus, while charge gradients are an important protection mechanism of the plant, the metabolic processes provide a way to actively influence the proton pressure.

Chapter 5

Metabolic modelling of the legume-rhizobia symbiosis

After construction and analysis of the *M. truncatula* model, a reconstruction of the symbiont is necessary to investigate the symbiotic behaviour. The plant model however needs to be adapted simultaneously, as the nodule is a highly specialised tissue. To this end, a nodule submodel was created and combined with either of two rhizobial models. Both the rhizobial models and the nodule model will be presented in the following section. This chapter will provide results of the analysis of the interaction between either rhizobial model with *M. truncatula* and will illustrate potential reasons for experimental observations.

5.1 Models for *S. meliloti* and the nodule of *M. truncatula*

As the symbiosis between legumes and rhizobia occurs in highly specialised tissues, the nodules, it is not a loose connection between two distinct organisms. Instead, the plant cells are highly specialised, and metabolite transport between bacteroid and plant nodule cells is tightly regulated. This section describes the specialised networks used to model the symbiosis, and provides information about the exchange between both partners.

5.1.1 A nodule model for *M. truncatula*

The nodules of a legume are highly specialised for symbiotic nitrogen fixation. The metabolism of this tissue is adapted for nitrogen assimilation, and the nodules provide a highly specialised environment for the rhizobia. To reflect this in the simulation of symbiotic nitrogen fixation, a submodel of the genome-scale model was created using established methods for tissue specific model generation [198]. The compounds forming the biomass for the nodule were assumed to be the same as the root fraction in the biomass determination experiments. Furthermore, additional exporters were included to allow the production of all compounds. This addition was necessary as the method used would otherwise remove all reactions not able to carry flux. The addition of exporters circumvents this for reactions which are otherwise only inhibited due to a lack of final metabolite export. Furthermore, constraints were included to prohibit the use of light and force the consumption of sucrose as the main carbon source. Expression data of Benedito et al. [49] was used to obtain gene activities.

The resulting submodel contained 848 conversions, 8 importers, 27 exporters, and

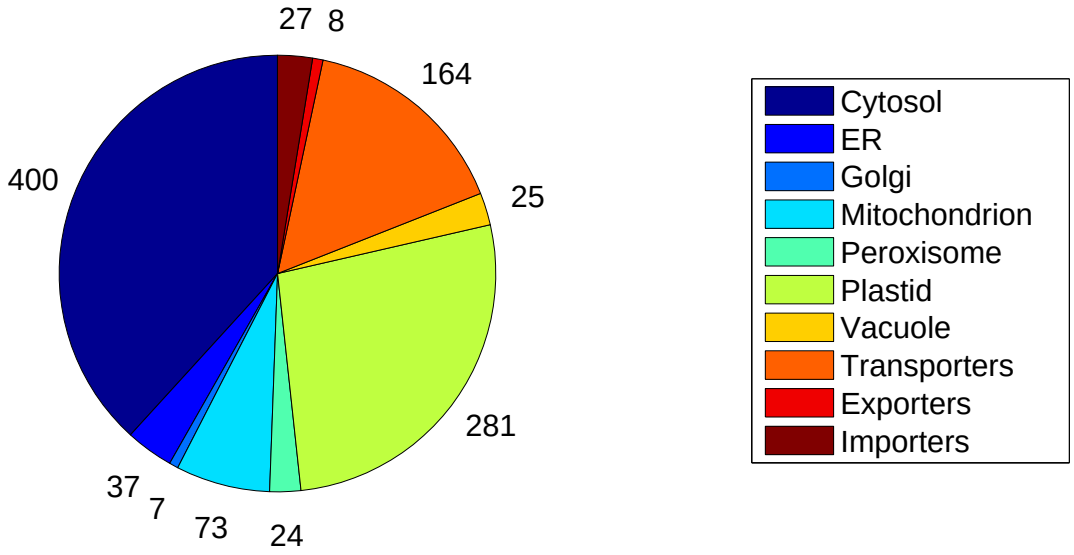


Figure 5.1: Distribution of reactions in the nodule submodel. The relative compartment sizes are similar to the genome-scale model, while the fraction of transport reaction has increased. Many reactions not directly necessary for biomass production were removed by the method of Shlomi et al. [198] when gene expression for the coding enzyme was insufficient.

164 transporters. All compartments were still present in the submodel, with the reaction distribution is detailed in Figure 5.1. While the relative compartment sizes are comparable to the genome-scale model, the relative amount of transporters has increased. This is partially due to the restriction of redundant reactions to one compartment in the tissue model generation. While the algorithm tries to activate as many reactions as possible, it punishes the inclusion of additional reactions, thus leading to the removal of replicate reactions in different compartments. The fraction of reactions present in multiple compartments decreased from 35% in the genome-scale model to 21% in the nodule model. However, this removal of redundancy makes it necessary to shuttle more compounds. The submodel is able to produce biomass on nitrate and ammonium and can provide dicarboxylic acids to a potential symbiont.

E. coli Id	Description	Reason
M_ca2_c, M_cl_c, M_cu2_c, M_cobalt2_c, M_fe2_c, M_fe3_c, M_k_c, M_mg2_c, M_mn2_c, M_mobd_c, M_nh4_c, M_so4_c, M_zn2_c	Ca^{2+}, Cl^{-}, Co^{2+}, Cu^{2+}, Fe^{2+},Fe^{3+}, K^{+}, Mg^{2+}, Mn^{2+}, MoO_4^{-2}, NH_4^{+}, SO_4^{2-}, Zn^{2+}	Inorganic ions would not contribute to fluxes, as they would just be taken up.
M_murein5px4p_p	Peptidoglycans	Polymer biosynthesis was removed
M_pe160_p, M_pe160_c, M_pe161_p, M_pe161_c	Phosphatidylethanolamines	No corresponding compounds present in the MetaCyc database
M_thmpp_c	Thiamine pyrophosphate	Unclear steps in the biosynthesis pathway

Table 5.1: Compounds removed from the biomass formulation of *E. coli* to adjust it to the generated rhizobial model.

5.1.2 Two models for *S. meliloti*

Two models for the rhizobial partner were investigated. The first model was taken directly from the MetaCyc database for *S. meliloti* 1021 (MC-Model). The model was curated to be able to produce a biomass composition, which was assumed to be similar to that of *E. coli* (model iJR904) [118]. The changes in the biomass definition are listed in Table 5.1. The most significant change is the omission of ions within the biomass. Even though ions are an important part of the biomass, the main effect in an FBA simulation is their uptake, as they are mostly disconnected from any other reaction. Thus they do not contribute to other parts of metabolism in an FBA simulation. They are essentially only taken up and deposited without further contributing to other fluxes. Their inclusion would therefore not help in validating any biosynthetic pathway. With the exception of those listed in Table 5.1, the MC-model could produce all other biomass compounds.

The second model was taken from Zhao et al. [68] (SNF-Model) and is based on the

R. etli model by Resendis-Antonio et al. [101]. It was adapted to *S. meliloti* by investigating homologies and excluding reactions unlikely to be present in *S. meliloti*, while adding some *S. meliloti* specific reactions. The model is restricted to the symbiotic nitrogen fixation stage of the rhizobium, and does therefore not contain the capability to produce biomass.

For the present work, both models were adjusted to obtain atomic and charge balance where necessary. Transport to and from both models was taken from the literature and is displayed in Figure 1.12.

5.1.3 Maximal fluxes between plant and bacteroid

Symbiotic nitrogen fixation is performed under very particular conditions. First, the oxygen uptake rate is highly controlled by leghemoglobin and the nodule environment. Second, the dicarboxylic acid transporter can only translocate a limited quantity of acids. While these fluxes have been investigated [226, 227], they are based on pure bacteroids. It is therefore necessary to convert the measurements to fluxes relative to the growth rate of *M. truncatula*. As detailed in Section 3.2 the growth rate of *M. truncatula* ranges from $0.10\,\mathrm{g\,g^{-1}\,d^{-1}}$ to $0.12\,\mathrm{g\,g^{-1}\,d^{-1}}$. For the following calculations, the growth rate is assumed to be at the lower end of these values, as $0.1\,\mathrm{g\,g^{-1}\,d^{-1}}$. This growth rate will be referred to as the normal growth rate. The ratio between nodule dry weight and plant dry weight is in the range between 0.01 g/g [228] and 0.035 g/g [229]. To put the measurements for different uptake rates of bacteroids into perspective, a bacteroid count of $8 \cdot 10^{10}$ bacteroids per gram nodule [230] and a bacteroid weight of $3 \cdot 10^{-12}$ g per bacteroid [227] is assumed. Thus the total amount of bacteroids per gram of plant dry weight is in between 2.40 mg/g and 8.40 mg/g.

The measured uptake rates for dicarboxylic acids are $1.326\,\mathrm{\mu mol\,mg^{-1}\,h^{-1}}$ and

Flux	Experimental value $nmol\ mg^{-1}\ min^{-1}$ (based on bacteroid weight)	Model constraint $\mu mol\ g^{-1}\ h^{-1}$ (based on plant weight)	Source
Succinate uptake	22.1	0 - 9.4248	[226]
Malate uptake	18.7	0 - 11.1384	[226]
Total dicarboxylate uptake	-	0-11.1384	
Aspartate uptake	75.8	0 - 37.99*	[226]
Oxygen uptake	14.1 - 39.2	16.8*2	[227]

Table 5.2: Flux constraints for metabolite exchange between rhizobium and plant. For conversion between bacteroid and plant weight a ratio of 8.4 mg/g was assumed.
*: This constraint was used to limit the total amino acid exchange (uptake and export).
*2: The listed oxygen uptake is in the upper range of possible uptake rates. When alternate values are used this is explicitly mentioned.

1.122 µmol mg^{-1} h^{-1} for succinate and malate, respectively [226]. These measurements are implemented in the analysis of both models by restricting the succinate and malate uptake rates accordingly. Furthermore, an additional constraint to restrict the combined uptake of succinate and malate to 1.326 µmol mg^{-1} h^{-1} was introduced. This restriction is set as both dicarboxylic acids are taken up by the same dicarboxylic acid transport system (DCT). The maximal oxygen uptake was determined with a 95% confidence interval in the range from 0.846 µmol mg^{-1} h^{-1} to 2.352 µmol mg^{-1} h^{-1} [227]. In addition a maximal uptake and release of amino acids was incorporated based on the data for aspartate uptake of 4.548 µmol mg^{-1} h^{-1} by McRae et al. [226]. This constraint was added such that the combined flux of amino acids into and out of the rhizobium cannot exceed this value. Table 5.2 gives an overview of flux constraints in the model with respect to the fraction of bacteroids per plant.

Using these figures, the maximal nitrogen fixation per gram plant dry weight in the SNF-model is 1.28 µmol/h to 12.3 µmol/h. This would support a plant growth rate (us-

ing our biomass formulation) of $0.34\,\mathrm{mg\,g^{-1}\,h^{-1}}$ to $3.29\,\mathrm{mg\,g^{-1}\,h^{-1}}$ which is about 8% to 79% of the expected growth rate based on the experiments of Lötscher et al. [210]. This growth rate, however, is under the assumption that there is no additional energy needed within the rhizobium. Furthermore, the growth conditions used for biomass determination provided a large surplus of nitrogen sources. As detailed, this has undoubtedly led to an accumulation of asparagine to very high levels, which are unlikely to be found in symbiotic growing plants, for which nitrogen availability is limiting. The symbiotic biomass composition was therefore altered to reflect this by reducing the amount of non-protein-bound asparagine to only 5% of the amount accumulated by the plants growing in hydroponic solution, which is much closer to the accumulation of other amino acids. By this change the maximum growth rates in the model doubled to 16 - 153% of the expected growth rate. For further studies, the maximal oxygen uptake rate was set to $16.8\,\mathrm{\mu mol\,mg^{-1}\,h^{-1}}$ based on plant dry weight, which is in the upper range of possibilities and allows for an additional requirement of maintenance energy. This allows a maximal growth rate of 123.5% of the normal growth rate on the SNF-model.

5.2 Energy requirements of the symbiont

Plant growth under symbiotic nitrogen fixation is lower than growth using sufficient fertilizer. It was shown that the growth rate of plants relying on symbiotic nitrogen fixation is approximately between 60% and 90% of the growth of plants supplied with sufficient amounts of fertilizer [231]. As the maximal growth rate determined above is assuming no maintenance on the rhizobial part, the effect of a range of maintenance energies on the maximal growth rate was investigated. The effect of increasing ATP demand on the SNF-model is displayed in Figure 5.2.

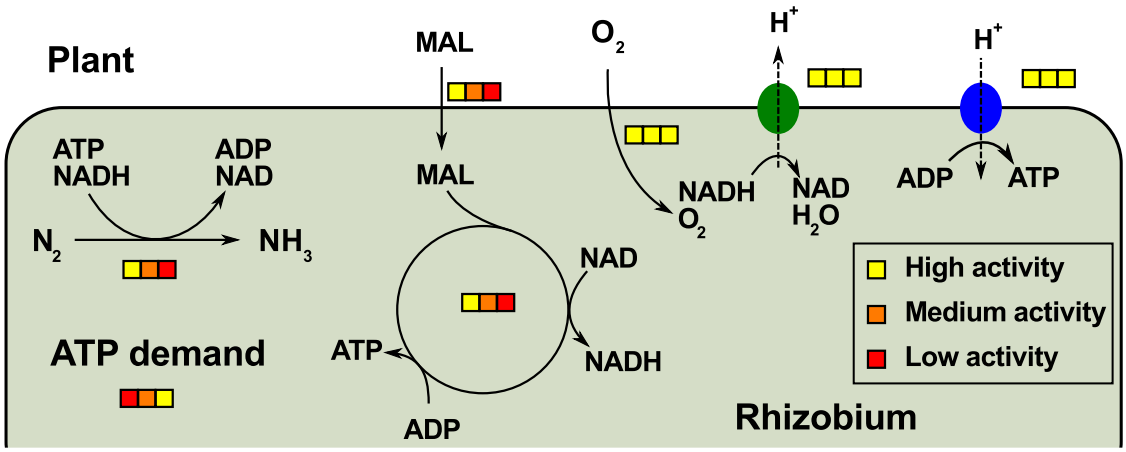


Figure 5.2: An increase in rhizobial ATP demand leads to a decreased import of malate and decreased TCA activity in the rhizobium. The restricting flux is the oxygen uptake flux, which limits the possible amount of respiration. When optimising for maximal nitrogen fixation or maximal plant growth, the oxygen uptake is always at its upper bound. While malate can easily be used to provide additional reductants without requirement of additional oxygen, ATP demand has to be satisfied by the electron transport chain (green) and the ATPSynthase (blue) as surplus reductant is not allowed to accumulate. The boxes indicate the activity at different ATP requirements with yellow indicating maximal activity (over different ATP demands) and red indicating low activity. The first box represents no ATP requirement while the last represents an ATP requirement equal to 65.98 µmol g^{-1} h^{-1}. The analysis was performed on the SNF-model.

The optimisation used was plant biomass maximisation, which is equivalent to maximal nitrogen fixation for the rhizobial system. At all tested ATP demands, the restricting flux for growth was the oxygen uptake flux. This observation was confirmed by the MC-model, which shows a similar response. Oxygen is needed to drive the electron transfer chain and to allow the surplus reductant from the TCA to regenerate ATP from ADP. While there is a small amount of ATP produced by the TCA, the nitrogenase needs a ratio of eight ATP to three reductant. The TCA instead provides a ratio of one to four, a large surplus of reductant. As in anaerobic bacteria this leads

to a necessity to dispose of reductants and the only way to do so is by the electron transfer chain, thus leading to the limitation by oxygen. It follows that, with increasing ATP demand, less and less malate is needed by the rhizobium, as less reductant can be used for nitrogen fixation.

A maintenance requirement of $7.6\,\mathrm{mmol\,g^{-1}\,h^{-1}}$ of ATP is commonly assumed as non-growth-related maintenance energy in *E.coli* [117]. Assuming a similar maintenance energy for *S. meliloti* translates to $63.47\,\mathrm{\mu mol\,g^{-1}\,h^{-1}}$ based on plant weight. The amount of fixed nitrogen with this ATP demand in the SNF-model only allows for a maximal production of biomass of 41% of the experimentally determined growth rate. This is much lower than the 60-90% observed in experiments for growth during symbiotic nitrogen fixation. When assuming 90% growth, the maximal ATP available is $25.96\,\mathrm{\mu mol\,g^{-1}\,h^{-1}}$, which translates to about 40% of the non-growth-related maintenance of *E.coli*.

These deviations clearly show that the rhizobia work at a maximal capacity and suggest that there are likely some differences to the measurements *in vivo*. Either more oxygen is available, the maintenance energy is significantly lower than that in *E. coli* , or there are some alternative routes within the metabolic network, which are not present in the SNF-model. There are three observations from this initial study that point to the latter. First, amino acid cycling has been proposed as beneficial to nitrogen fixation[100] and was not observed in the SNF-model. Second, higher oxygen uptake rates are rather unlikely, as the oxygen supply is tightly regulated and its uptake rates were measured using leghemoglobin as transporter. Higher concentrations could in fact lead to an inactivation of nitrogenase. Finally, even though maintenance energy measurements can be quite diverse [232] the lower bounds seem to be somewhere near $5\,\mathrm{mmol\,g^{-1}\,h^{-1}}$ for bacteria, which is still higher than the $3.09\,\mathrm{mmol\,g^{-1}\,h^{-1}}$ for 90%

growth (based on bacteroid weight). Thus, the restrictions in the creation of the SNF-model might have been too tight, and it might be missing some reactions.

5.3 ATP generation in the rhizobial model

An investigation of the MC-model showed that in this model the potential growth is higher with a maximal growth rate (without maintenance energy) of 208% of the normal growth. With a maintenance of 63.47 µmol g^{-1} h^{-1} (similar to the maintenance in *E. coli*) the maximal growth rate is still at 104% of normal growth which is above the expected range for symbiotic nitrogen fixation and allows for some additional flexibility. This drastic increase in comparison with the SNF-model stems from several reactions present in the MetaCyc database which are not present in the SNF-model, and from a difference in one reversibility constraint. The differences are displayed in Figure 5.3.

One of the involved reactions not present in the SNF-model is a citrate (*pro-3S*)-lyase which catalyses the reversible lysis of citrate directly to oxalacetic acid and acetic acid. The lyase allows pyruvate degradation to acetate, which yields one ATP per pyruvate, thus increasing the ATP per reductant ratio. The presence of this pathway however is dubious, considering that knock-out studies have found that citrate synthase is essential for nitrogen fixation in *S. meliloti* [233]. It is likely that this specific citrate lyase was only present in the MetaCyc database due to a misinterpretation during the automated construction.

However, when restricting the flux through citrate-lyase to zero, the maximal fixation still allows for a 196% growth without maintenance and 96% with maintenance, which is close to the experimentally observed growth rates during symbiotic nitrogen fixation. Thus the main difference is not based in the citrate lyase reaction but in the reversibility of the pyruvate carboxylase (PYC) reaction (displayed in Figure 5.4).

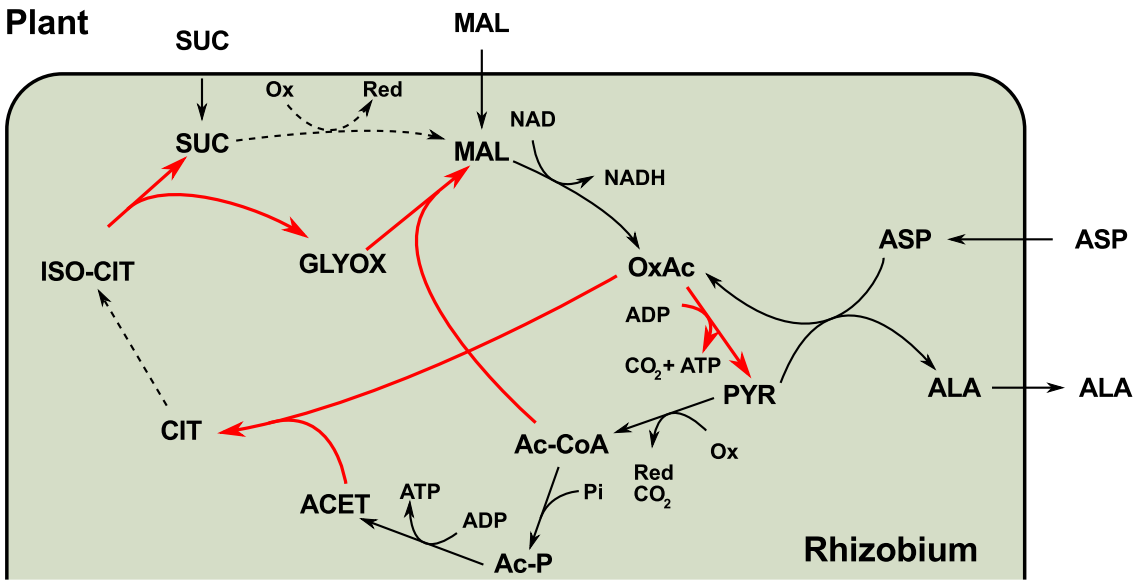


Figure 5.3: Some differences of the SNF-model and the MC-model. Reactions only present in the MC-model are marked in red. Citrate lyase allows the acetate bypass for citrate synthesis to produce additional ATP while the glyoxylate shunt reduces the amount of reductant per ATP. The most important difference however is the reversibility of pyruvate carboxylase, which allows the ATP producing carboxylation of oxalacetic acid to pyruvate. Dashed arrows indicate several individual steps which are combined into one reaction.

This reaction also allows an efficient usage of the glyoxylate shunt, which allows the rhizobium to avoid most of the reductant generating steps in the TCA. PYC allows the formation of pyruvate from oxalacetic acid by decarboxylation while simultaneously regenerating one ATP from ADP. The ΔG^0 of PYC is 8.5 kJ/mol (according to eQuilibrator [197]). However due to the comparatively large amounts of dicarboxylic acids available, and the high demand for ATP, this equilibrium can conceivably be shifted towards pyruvate. Another effect of this redirection of oxalacetic acid is that the cycling of amino acids can be observed in the MC-model. The simulation suggests that more than 40% of the nitrogen is exported as alanine and glutamate is used as an

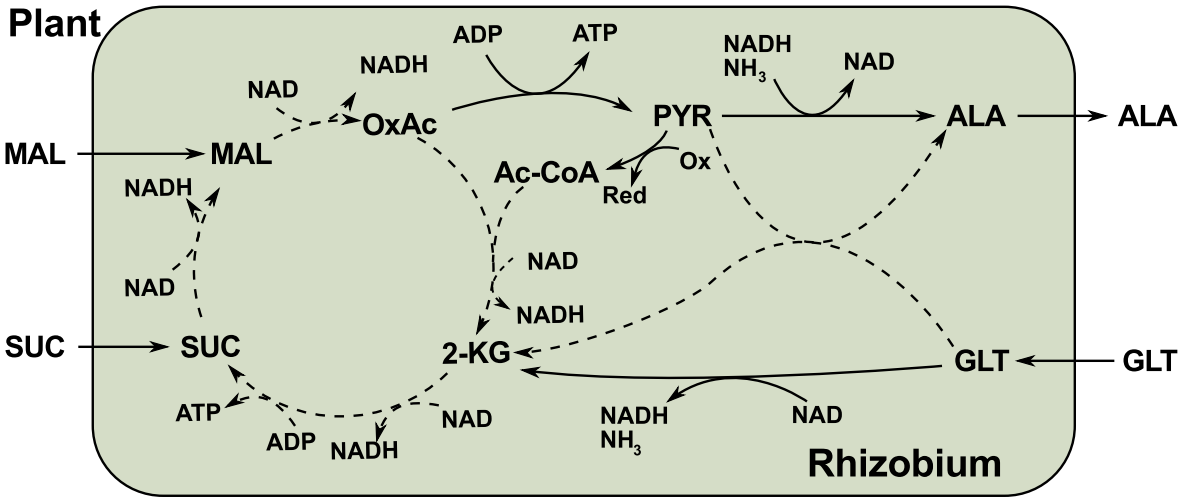


Figure 5.4: Rhizobial activity with blocked citrate lyase in the MC-Model. The PYC activity is coupled with an increase in alanine dehydrogenase forming alanine from pyruvate, and consuming reductant. Simultaneously glutamate is taken up as an additional substrate to drive the TCA, and provide substrate oxalacetic acid (via the TCA) for alanine biosynthesis.

additional carbon source when optimising for growth.

Another difference between the MC- and the SNF-model is in the electron carrier capacity. In the MC-model, flavoproteins can transfer their electrons to NADP, and from there to NAD which is not possible in the SNF-model. This model also encompasses a succinate dehydrogenase, which links the oxidation of succinate to the reduction of flavoproteins. Thus there is a route that can reduce NAD, where commonly quinone is reduced. This leads to a higher rate of proton transfer over the membrane per succinate oxidation. This reaction can potentially occur if oxygen is extremely limiting, but is generally only a halfreaction of complex II in the electron transfer chain. While the representation of such reactions within MetaCyc does reflect current knowledge about the mechanisms, the reactions are unlikely to happen independently. Thus they have to be considered with care. Removing this possibility, and linking succinate oxidation directly to quinone reduction, reduces the maximal growth rate to 173% without and

78% with the assumed maintenance energy, which leads to a decrease and figures close to the expected growth rates. Under these conditions malate becomes limiting, as it is the most efficient reductant source (since it allows a higher NADH to quinol ratio than succinate), and aspartate is imported (instead of exported). This is contrary to the common assumption that aspartate is exported [100]. When, however, aspartate uptake is removed, the maximal growth rates decrease to 154% of normal growth without and 59% with maintenance, which is below the expected range. Thus, either reactions are still missing from the model, or the energy requirement is distinctly lower.

5.4 Effects of oxygen limitation

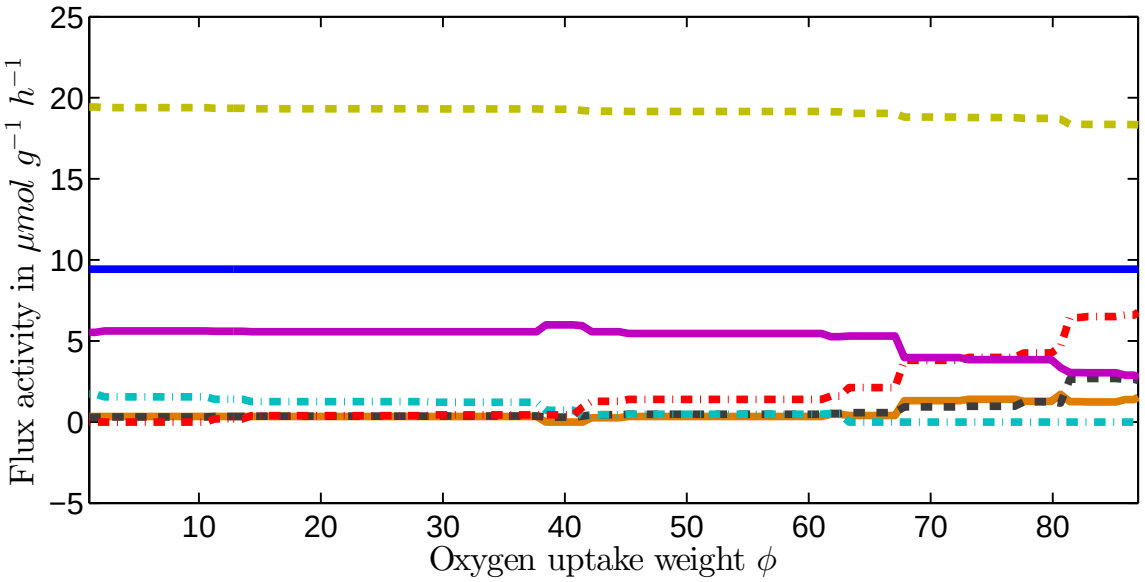
While the above studies focused on the maximal nitrogen fixation rate, it cannot be verified, whether the rhizobial system is optimised for this goal. As demonstrated, oxygen supply is a limiting factor for nitrogen fixation in the models. It is however interesting to observe what kind of effect a limitation of oxygen has on the fluxes in the MC-model before it becomes limiting. The MC-model was chosen for this investigation, as it provides a higher flexibility than the SNF-model. For the following analysis, the maximal oxygen uptake of the rhizobium is set to the upper limit ($19.76\,\mu mol\,g^{-1}\,h^{-1}$) of the measurements by Bergersen [227]. Citrate-lyase was only allowed to carry flux in the citrate lysing direction, and no other reactions were restricted. The uptake limits for amino acids and dicarboxylic acids were set as above. The biomass amount produced was set to 80% of the normal growth rate, reflecting the commonly slower growth rate during symbiotic nitrogen fixation. Finally, the maintenance energy was assumed to be $63.47\,\mu mol\,g^{-1}\,h^{-1}$ for the rhizobium. The system was optimised for minimal flux, and a weight scan (with weight ϕ) was performed for the oxygen importer.

The first observation is that the maximal available oxygen consumption rate of

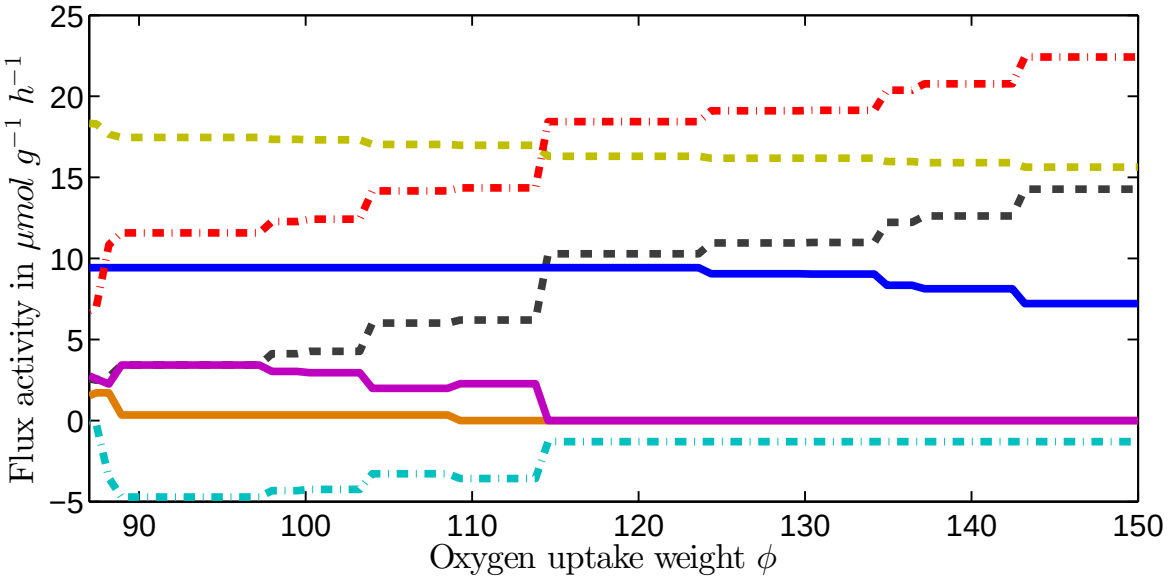
$19.76\,\mu mol\,g^{-1}\,h^{-1}$ was never fully used. The applied constraints therefore allow some variability, and no flux is at a limit. The importers and exporters that show a response to changing oxygen weight are displayed in Figure 5.5. Two main phases can be distinguished. In the first phase ($1 \leq \phi \leq 87$), ammonia is the main source of nitrogen provided to the plant (see Figure 5.5a). However, some nitrogen is fixed in the rhizobium, and exported as either aspartate or alanine.

With higher ϕ more alanine is exported, and ammonia export starts decreasing (at $\phi = 67$). At this weight, there is a short increase in succinate import. Succinate is converted to oxaloacetate in two oxidising steps, and further metabolised to pyruvate by action of PYC, which acts as substrate for alanine dehydrogenase. The necessity for this increase is due to the already maximal malate influx. The second phase starts at $\phi = 87$ (see Figure 5.5b). Aspartate export has stopped and aspartate is now imported as a carbon source.

The imports and exports are in agreement with the suggested scheme by Lodwig et al. [100], but instead of cycling glutamate for aspartate, as was suggested in *R. leguminosarum*, an exchange of glutamate for alanine is observed. The alanine export is in agreement with observations by Waters et al. [107], where alanine was suggested to be the only exported nitrogen source. Waters et al. also observed that other amino acids were similarly exported, if additional oxygen was available to the rhizobium [107]. In the present model, the reason for the export of amino acids is based on the redox balance. While cheap oxygen allows the electron transfer chain and ATPase to run at the necessary capacity, the activity of the electron transfer chain becomes more restricted with increasing oxygen cost. As the cytochrome *c* oxidase is the only active rhizobial enzyme requiring oxygen, the ATPase activity is directly linked to the uptake of oxygen, and therefore reduced when oxygen uptake becomes more expensive.



(a) First phase



(b) Second phase

Figure 5.5: Scan of the oxygen import cost ϕ representing the weight for the rhizobial oxygen import flux. At low ϕ nitrogen is exported largely as ammonia (a). At higher oxygen cost the nitrogen is exported as alanine (b), while glutamate and aspartate are used as additional carbon sources.

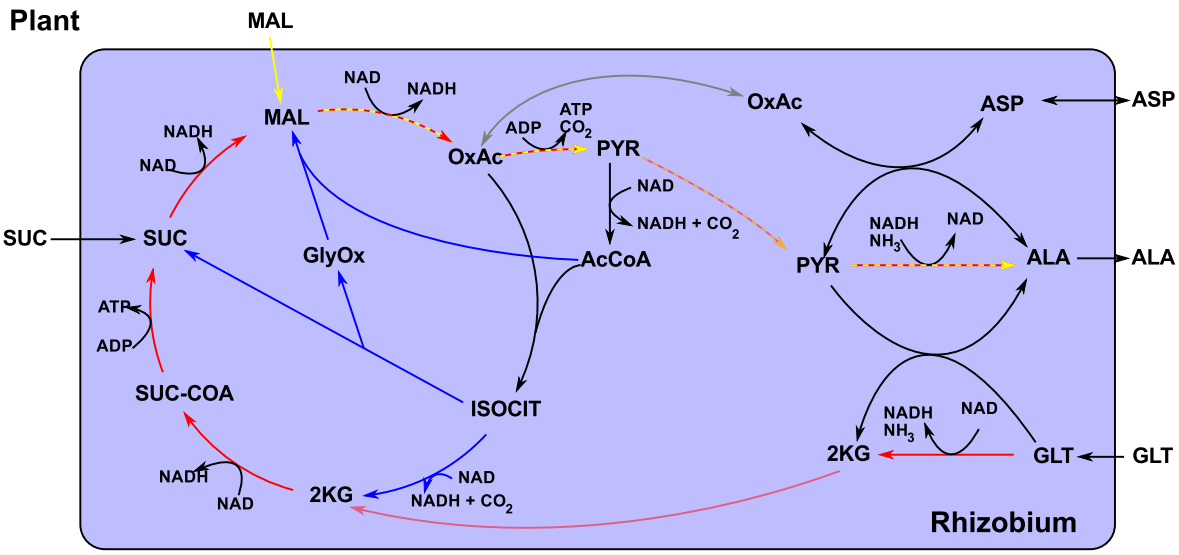


Figure 5.6: Alternative routes to adapt to oxygen limitation. While the glyoxylate shunt (blue) and the normal TCA provide three reductants per ATP, cycling glutamate for alanine yields 1 ATP per 1.5 reductants (red). A direct malate to alanine conversion is even more efficient yielding ATP without reductant (yellow). The malate needed however is fast becoming restricting, and no oxidative phosphorylation can be driven by this bypass.

This leads to a necessity for other reactions to replace the regeneration of ATP. However, these alternate pathways have to produce less reductant. Figure 5.6 shows two alternate routes by which this is achieved, and the flux distribution change of the involved reactions. While the glyoxylate shunt and the normal TCA (blue) yield the same amount of ATP per reductant, glutamate to alanine conversion and malate to alanine conversion have a higher ATP per reductant ratio.

Alanine dehydrogenase plays an important role in all these simulations. By removing reductant to fix ammonia to pyruvate, alanine dehydrogenase enables additional cycling of e.g. glutamate to generate ATP in the TCA. Figure 5.7 shows the change in metabolites produced by the pathways displayed in Figure 5.6.

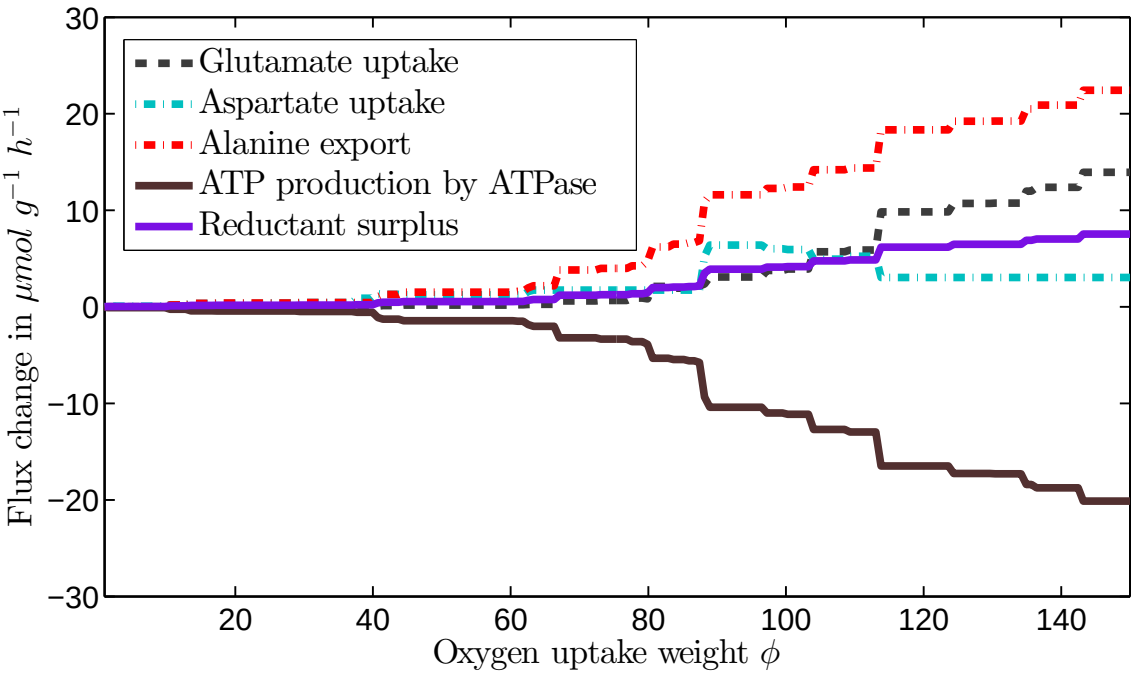


Figure 5.7: Alteration in the metabolite uptake and export from the rhizobium in response to increasing oxygen cost in the network displayed in Figure 5.6. Reductant from the decreasing activity of the ATPase is freed, and needs to be consumed. This is achieved by the alanine dehydrogenase where it is used to fix the produced ammonia. Simultaneously, the ATP demand is matched by an increase in pyruvate carboxylase and succinyl-CoA synthase. The latter is driven by the import of glutamate, and transfer of the amino group from glutamate to oxaloacetate yielding aspartate and 2-KG. 2KG can then be used to drive the succinyl-CoA synthase with a minimum of reductant production inside the rhizobium.

5.4.1 Other factors influencing amino acid cycling

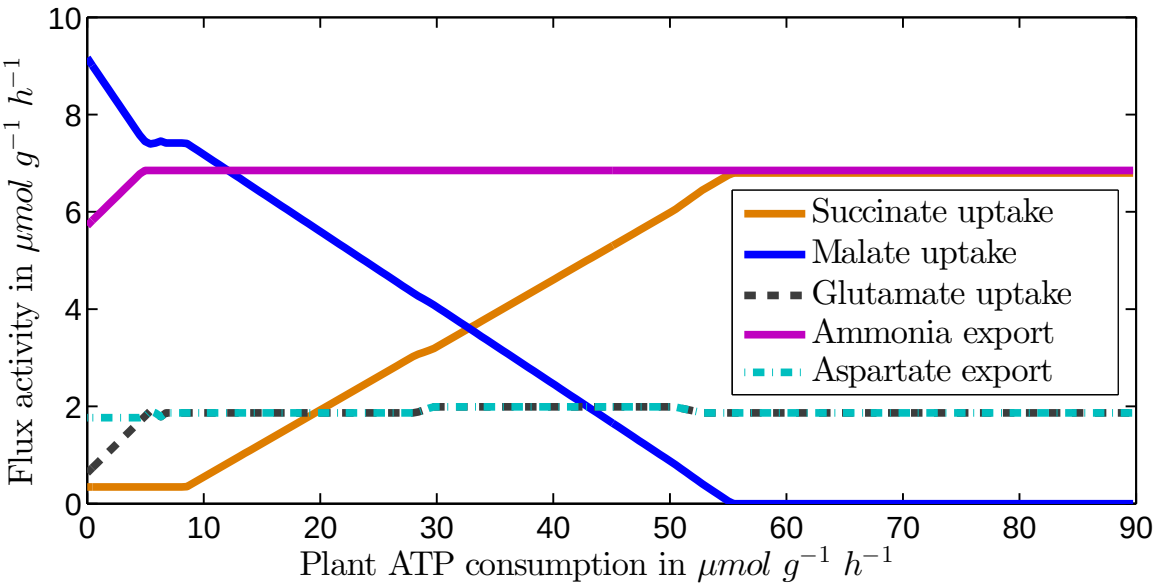
The energy requirement of the plant is another factor that influences the amount of amino acid cycling. To illustrate this, a scan of increasing ATP demand was performed. The settings were the same as for the oxygen limitation scan above, with the exception that the maintenance ATP for the SNF-model was reduced to 50%. This was necessary, as the SNF-model would otherwise not be able to sustain the defined growth. When no additional ATP requirement beside growth-related ATP requirements is assumed, the

quantity of amino acids cycled is depending mainly on the restriction of oxygen. When additional maintenance is required an increase in amino acid cycling is visible (shown in Figure 5.8), which fits well to the suggested exchange of glutamate and aspartate.

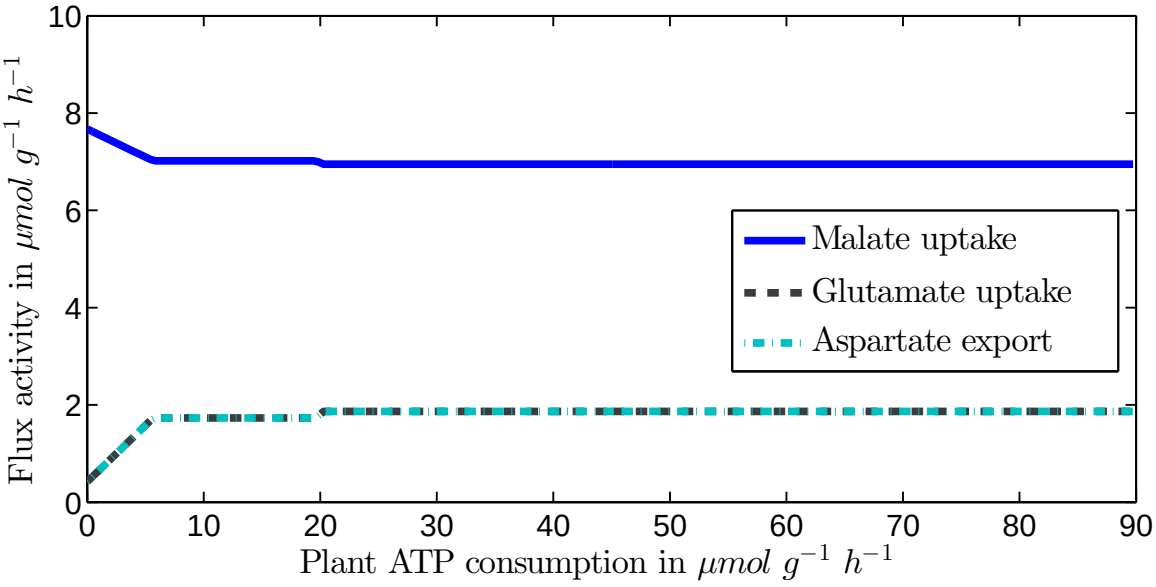
This change is observable in both the SNF- and MC-model. The switch is induced by the increased requirement for the plant mitochondrion to produce ATP, and it is cheaper (flux wise), to let the rhizobium transfer the amino group from glutamate to oxalacetic acid. It also provides the rhizobium with additional energy for nitrogen fixation which would otherwise have to be supplied as malate or succinate. In the MC-model there is also a gradual switch from malate supply to succinate supply. This is accompanied by an increase in oxygen uptake (not shown), as more ATP is synthesised by oxidative phosphorylation. The higher oxygen uptake rate saves some carbon which would otherwise have been respired by the pyruvate carboxylase. In addition to this, the MC-model also exhibits an increase in ammonia export (purple), as the aspartate is no longer derived from pyruvate but from the imported glutamate.

5.5 Discussion

The presented analysis shows that oxygen is a limiting factor for symbiotic nitrogen fixation. In both the MC- and the SNF-model additional carbon can not be employed without further oxygen. Thus the optimisation of nitrogen fixation leads to the selection of ATP regenerating pathways that avoid the consumption of ATP. The inflexible highly focused nature of the SNF-model indicates that this model is missing some pathways, as it was not able to provide sufficient nitrogen for the expected growth during nitrogen fixing conditions, when even small amounts of rhizobial maintenance were assumed. In contrast the MC-model provided some potential targets for optimisation of nitrogen fixation, and indicated potential pathways missing in the SNF-model. One possible



(a) MC-model



(b) SNF-model

Figure 5.8: Response to increased energy demand in the plant system. The rhizobial transaminases are increasingly used to produce the required aspartate for growth. In both the MC- (a) and SNF-model (b) glutamate is taken up, and aspartate is produced. In the SNF-model this allows a reduction in required malate. In the MC-model the increased demand for ATP biosynthesis is accompanied by a switch from malate to succinate, and a higher oxidative phosphorylation in the rhizobium.

route that could allow sufficient nitrogen to be fixed in the SNF-model is by allowing pyruvate carboxylase reaction to function in reverse. A reversible pyruvate carboxylase can also explain observations of increased alanine export [107]. One promising option to optimise nitrogen fixation is the insertion of a (*pro-3S*)-citrate lyase enzyme, for which simulations suggested an improvement in the symbiotic nitrogen fixation capacity of *S. meliloti*.

The investigation also elucidated potential reasons for the observation of amino acid cycling. By circumventing parts of the TCA, ATP can be obtained, while less reductant is produced. This would allow the symbiont to fix more ammonia, as the requirement of a two to one, ATP to reductant, ratio of the nitrogenase enzyme leads to the limitation of the process by oxygen.

The requirement for a microaerobic environment by nitrogenase induces a situation similar to the situation in anaerobic bacteria. While excessive reductant is available, it cannot be employed to drive ATP regeneration as it is restricted by the lack of oxygen and becomes inhibiting due to the redox balance. In some anaerobic bacteria, this surplus is released by using hydrogen as a final electron acceptor in hydrogenase enzymes [234]. An approach to increase the nitrogen fixation yield could therefore be to introduce a hydrogenase enzyme. The enzyme could use protons as final electron acceptor, producing hydrogen, and would lower the amount of surplus reductant. This would allow an increased ATP generation from the TCA, as surplus reductant could efficiently be disposed of.

However, there are suggestions that oxygen is not the main limiting factor for nitrogen fixation, and instead additional reductant could be employed to obtain more ATP from oxidative phosphorylation. This is particularly interesting, as there are some rhizobial species which already exhibit some hydrogenase activity [235]. However,

the measured hydrogenase activity is commonly associated with an uptake hydrogenase [236, 237] and not with an increased release of hydrogen, as would be suggested if the environment of the rhizobium is limited similar to an anaerobic situation. In *S. meliloti* no hydrogenase activity has been observed. Therefore a genetic modification introducing a hydrogenase into *S. meliloti* could increase the efficiency of oxygen usage. Rhizobial hydrogenases seem to resemble a complex similar to complex I [238]. Therefore the improvement would stem from a change in balance between entry of electrons in the electron transfer chain from complex II to a complex similar to complex I, if a rhizobial hydrogenase was used.

Chapter 6

Conclusion and outlook

In the present thesis a genome-scale metabolic reconstruction for the model plant *Medicago truncatula* was presented. An overview of the properties of the reconstruction was provided, and it was shown that the scope of the model is comparable to other plant reconstructions [120, 121, 171]. To my knowledge, the presented model is the first fully charge-balanced plant model. It thus provides a valuable resource for the investigation of metabolic responses to external events involving the exchange of charges. For model curation and analysis, a detailed biomass composition of *M. truncatula* was obtained using established methods. The composition was determined for three different types of tissues (roots, stems and leaves), and can be used for a more detailed analysis of those tissues. The extensive nature of this compilation of data provides a foundation for future work.

While part of this data is directly applicable other parts have to be considered with more care. Recent findings in *A. thaliana* have, for example, shown a change in biomass composition in response to changing nitrogen sources [203]. While e.g.protein and cell wall composition data is likely to be unaffected by these kinds of changes,

the total fractions of lipid, protein and cell wall are likely to be altered. In particular, the soluble fraction is very likely to show an alternate composition in response to an altered environment. This was considered in the adaptation performed for the symbiotic nitrogen fixation simulations. Still, the current set of data allows a thorough investigation of the plants' metabolism.

The presented analysis indicates a preference of ammonium over nitrate as substrate, as the latter is more expensive when used as nutrient for biomass production, both in terms of flux and in terms of reductant. For legumes, this reduction in energy demand is particularly important when considering growth during symbiotic nitrogen fixation. Symbiotic nitrogen fixation is a very expensive task, with respect to energy and reductant, and avoiding additional energy requirements is important. Many plants however tend to grow better on nitrate nutrition than on ammonium nutrition [239]. They accumulate higher total amounts of nitrogen when nitrate is used as nitrogen source, but the relative nitrogen fraction of the biomass is smaller than in plants growing on ammonium or ammonium-nitrate mixtures [203, 239]. This is particularly striking in *A. thaliana* which, at least in cell cultures, cannot grow on ammonium as the sole nitrogen source and eventually dies [203, Fig 1a].

Ammonium can pose several different kinds of stress to the plant. When ammonium is incorporated into the biomass this is commonly associated with a release of protons. Thus ammonium can both influence the internal pH of the plant, or, if the protons get excreted, can lead to soil acidification [240]. In addition, it has been discovered that plant maintenance costs increase when ammonium is the sole nitrogen source [203, 241]. The assumed reason for this increase is the active export of ammonium from the plant cytosol to maintain a low internal ammonium concentration. This in turn leads to a high energy requirement to export ammonium against a high concentration

gradient, wasting significant amounts of energy [203]. However, it is not yet fully understood why many plants exhibit this futile cycling, while other plants are much less affected. *M. truncatula* in particular is only slightly affected by ammonium nutrition and copes well with ammonium as the sole nitrogen source [239]. This is expected, as a significant fraction of nitrogen in symbiotic nitrogen fixation is supplied in the form of ammonium [108, 109]. An export of the fixed ammonium would further hamper plant growth and make the symbiosis much less efficient.

In many plants, futile cycling of ammonium is likely to be the most important contributor to ammonium toxicity. In ammonium tolerant plants, like *M. truncatula*, the performed simulations suggest that the soil acidifying effect plays a more prominent role. Thus, while not applicable to ammonium intolerant plants, the simulated model reactions to a change in nitrogen supply elucidate the energetic effects of this change in *M. truncatula*. The soil acidification hypothesis was further supported by the investigation of the effects of changing proton pressure. While the analysis can only reflect the metabolic response, and ignores the accumulation of positive charges as a charge barrier, it still showed good agreement with biological observations (see Section 4.3.2). This indicates that the presented model is suitable for the investigation of changes to the external environment, and that its analysis is able to indicate potential reasons for observed changes.

The presented genome-scale model is an important tool for the analysis of general properties of *M. truncatula*. However, the investigation of symbiotic nitrogen fixation required a more focused model, as the nodule is a highly specialised tissue. To this end a nodule specific submodel was created, which was used in interaction with two rhizobial models. Analysis of both combined models suggested that oxygen supply is the main limiting factor for nitrogen fixation. The investigation also showed, that either the

amount of maintenance required by the bacteroids is less than that in other microbes, or that the model based on the *R. etli* reconstruction can only supply insufficient energy. This indicated that the model based on MetaCyc (MC-model) provides a better approximation on the actual state of the bacteroids, due to its higher flexibility. Both models showed that the cycling of amino acids can be beneficial with respect to efficient use of energy. While this finding supports the general observation of amino acid cycling [100], it cannot explain the essentiality of branched chain amino acid supply in some rhizobial species [110]. However, *S. meliloti* does not show this essentiality and this suggests that the affected rhizobia have repressed biosynthesis systems for these amino acids during symbiosis.

The analysis of the MC-model further suggests that the introduction of a (*pro-3S*)-citrate lyase into the rhizobial partner could lead to an increase in nitrogen fixing activity. Exploring this potential, and investigating the effects of different pyruvate carboxylases, could lead to an improved nitrogen fixation. While this is promising, it is possible that there are other enzymes in rhizobia involved in the regeneration of pyruvate that are yet undetected. Pyruvate carboxylase (PYC) has been investigated by Dunn et al. and a knock out was not found to alter nitrogen fixation [242]. However activity of PYC was detected in rhizobia grown on succinate, suggesting that the enzyme is acting in reverse or that it is increasing oxalacetic acid concentrations at the expense of energy. Finally, the observation that an increased energy demand in the host plant leads to an increased cycling of amino acids indicates that the plant can use the rhizobium to convert amino acids for its own biosynthetic activity.

Further research should be directed at improving the accuracy of the model and incorporating additional information into the reconstruction. A recent suggestion by Cheung et al. [243] could improve the accuracy of charge translocation over membranes.

The current model uses uncharged transport over membranes and only allows charged translocation in the electron transfer chains. While this is accurate for some symport and antiport processes, other transport processes need charge gradients, which have to be established by active proton translocation. Cheung et al. incorporated data on multiple known transporters and thus achieved an explicit inclusion of transport energy costs [243]. This approach improves the ability of a network to estimate maintenance energy costs. It also alters the energetic costs for biomolecules in a compartmentalised network, if the biosynthetic process is distributed between multiple compartments. The main limitation however is in the available information on transport systems. While some more important transporters are well characterised, there is a lack of data on others. When this data becomes available, incorporating it in a future iteration of the model will lead to an improved starch to biomass conversion estimation. Since the main effect of charged transport is the increase in active proton gradient formation, the main impact of this inclusion is expected to be on the energy metabolism.

Another improvement that can be achieved is the incorporation of different protonation states for different compartments. The current model assumes the protonation states as deposited in MetaCyc. This means that a pH of 7.3 is assumed and the metabolite protonation state is that of the major state at this pH. Adjusting the protonation states for different compartments allows a better estimation of energetic costs. The difference in pH between the mitochondrial matrix and the cytosol is about 0.75 pH units [1, p. 847]. This can lead to changes in protonation states and thereby the additional release of protons in the mitochondrial matrix. Simultaneously, this release could lower the proton gradient used for oxidative phosphorylation and alter the potential energy gain by the ATPase.

From an experimental perspective, flux measurements, in particular uptake of nu-

trients and release of secondary metabolites, could greatly improve the potential of the model. Data on photosynthetic activity during the day, and respiratory activity during the night, would allow an estimation of the maintenance energy of the plant. Biomass data for growth on different nitrogen sources, and during symbiotic nitrogen fixation, would further allow a more focused study of particular aspects of the metabolic processes involved during these conditions. This data could also be most valuable in determining the factors that influence growth on ammonium nutrition in *Medicago truncatula*.

With the large amount of flavonoids present in the model and the extensive flavonoid biosynthesis pathway reconstruction, it could also serve as a potential tool for genetic engineering. Because flavonoids can act as antibiotics [205] they are of high interest for the pharmaceutical industry. With the availability of cell cultures of *M. truncatula* [44], this potential can be exploited by investigating the present model with methods like OptKnock to improve the simulated production of flavonoids. The indication of potential modifications to the metabolism that divert additional flux towards the biosynthesis of flavonoids, gives this system a great potential in the bioproduction of flavonoids.

Finally the presented model is expected to serve as a tool for modellers and biologists alike to help improving research into legumes. The presented analysis shows the potential of the model and prompts hypotheses about the effects of ammonium on legumes. The model in its current state can serve as a basis for further research into the improvement of the symbiotic nitrogen fixation process and other processes specific to *M. truncatula*.

Appendices

Appendix A

Comparison with MedicCyc 1.0

Changes and alterations to reactions from MedicCyc incorporated in the model and database.

MedicCyc identifier	Replacement	Comment
1.1.1.39-RXN	MALIC-NAD-RXN	
1.21.3.1-RXN	removed	Dead end and annotation not specific
1.3.3.9-RXN	removed	Annotation changed
1.5.1.34-RXN	removed	Gene no longer found in genome
1.97.1.2-RXN	removed	Annotation change and dead end
2-HYDROXYISOFLAVONE-REDUCTASE-RXN	RXN-3562	
2.3.1.114-RXN	removed	Dead end and annotation not specific
2.3.1.153-RXN	RXN-7945	
2.4.1.113-RXN	removed	No protein modification in the model
2.4.1.150-RXN	removed	Dead end and annotation not specific
2.4.1.65-RXN	removed	Unspecific annotation
2.4.2.38-RXN	removed	Dead end and annotation not specific
2.4.2.39-RXN	RXN-9104	
2.7.1.152-RXN	removed	Unspecific annotation
2.7.1.99-RXN	RXN0-1134	
3-SULFINOALANINE-AMINOTRANSFERASE-RXN	removed	Dead end and annotation not specific
3.1.2.19-RXN	ACYL-COA-HYDROLASE-RXN	
3.1.3.46-RXN	6-PHOSPHOFRUCTO-2-KINASE-RXN	
3.2.1.149-RXN	removed	Dead end and annotation not specific
3.2.1.2-RXN	removed	Included in starch degradation
3.5.4.29-RXN	GTP-CYCLOHYDRO-II-RXN	
3.6.1.41-RXN	removed	Annotation changes (genes reannotated)
3.6.3.14-RXN	ATPSYN-RXN	

Changes and alterations to reactions from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
4-NITROPHENYLPHOSPHATASE-RXN	removed	Gene reannotated and dead end
4.2.3.12-RXN	removed	Dead end and annotation not specific
6.3.2.10-RXN	removed	Dead end and annotation not specific
6.3.4.10-RXN	removed	No protein modification in the model
ACETYLSYNCLTH-RXN	removed	Unspecific annotation
ADENYLATE-ISOPENTENYLTRANSFERASE-RXN	removed	Dead end and annotation not specific
ALCOHOL-DEHYDROGENASE-ACCEPTOR-RXN	ALCOHOL-DEHYDROG-GENERIC-RXN	
ALKANE-1-MONOOXYGENASE-RXN	removed	Annotation changed and dead end
AMINEOXID-RXN	removed	Unspecific reaction
ARISTOLOCHENE-SYNTHASE-RXN	removed	
ARYLAMINE-N-ACETYLTRANSFERASE-RXN	removed	Dead end and annotation not specific
BISPHOSPHOGLYCERATE-MUTASE-RXN	removed	Annotation changed and dead end
BISPHOSPHOGLYCERATE-MUTASE-RXN	removed	Annotation specific to another reaction and dead end
CAFFEATE-O-METHYLTRANSFERASE-RXN	RXN-1104	
CARBOXYMETHYLENE BUTENOLIDASE-RXN	removed	Dead end and annotation not specific

Changes and alterations to reactions from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
CHOLESTENOL-DELTA-ISOMERASE-RXN	removed	Dead end and annotation not specific
CYSSYNMULTI-RXN	CYSTSYN-PWY	
CYTOSOL-AMINOPEPTIDASE-RXN	removed	Substrates not defined
DARABALDOL-RXN	removed	Dead end and annotation not specific
DECAPCISTRANSFER-RXN	RXN-8999	
DEOXYADENOSINE-KINASE-RXN	removed	Dead end and annotation not specific
DEOXYCYTIDINE-DEAMINASE-RXN	CYTIDEAM-RXN	
DHS-PHOSPHATASE-RXN	removed	Dead end and annotation not specific
DIENELACHYDRO-RXN	removed	The coding genes code for another reaction and this reaction is a dead end
DIHYDROOROTATE-DEHYDROGENASE-RXN	RXN0-6490	
EPOXIDE-HYDROLASE-RXN	removed	Products not specified
FORMATEDEHYDROG-RXN	RXN0-3281	
GALACTOACETYLTRAN-RXN	removed	Annotation changed
GALACTOSIDE-2-L-FUCOSYLTRANSFERASE-RXN	RXN-9463	
GLUCOSAMINE-6-P-DEAMIN-RXN	removed	Annotation is for another reaction (isomerase not deaminase)
GLYCERALDEHYDE-DEHYDRO-RXN	removed	Dead end and annotation not specific
GLYCOPHOSPHORYL-RXN	removed	Included in starch degradation
GLYMALTOPHOSPHORYL-RXN	removed	Included in starch degradation

Changes and alterations to reactions from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
GST-RXN	removed	Unspecific annotation
HYDROXYPYRUVATE-REDUCTASE-RXN	RXN0-300 + GLYCERATE-DEHYDROGENASE-RXN	
ISOVALERYL-COA-DEHYDROGENASE-RX	RXN0-2301	
KETOISOCAPROATE-RXN	removed	Dead end and annotation not specific
KETOLACTOSE-RXN	removed	Dead end and annotation not specific
L-3-CYANOALANINE-SYNTHASE-RXN	removed	Gene annotation no longer present
L-ASCORBATE-OXIDASE-RXN	RXN-3523+ RXN-3541	
L-ASCORBATE-PEROXIDASE-RXN	RXN-12440	
L-ASPARTATE-OXIDASE-RXN	removed	This reaction is performed by a D-aspartate oxidase and not using L-aspartate
L-IDITOL-2-DEHYDROGENASE-RXN	removed	Dead end and annotation not specific
LCYSDESULF-RXN	removed	Annotation specific for another reaction
LIPOXYGENASE-RXN	removed	Dead end and annotation not specific
LONG-CHAIN-ALCOHOL-OXIDASE-RXN	removed	Unspecific reaction and annotation no longer present
LONG-CHAIN-FATTY-ACYL-COA-REDUCTASE-RXN	1.2.1.41-RXN	Changed to instance reaction
MALTACETYLTRAN-RXN	removed	Annotation changed
MYOSIN-HEAVY-CHAIN-KINASE-RXN	2.7.11.7-RXN	
NACGLCTRANS-RXN	removed	Dead end and annotation not specific

Changes and alterations to reactions from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
NAG6PDEACET-RXN	removed	Reaction is a dead end and annotated gene no longer present
PEPTIDOGLYCAN-GLYCOSYLTRANSFERASE-RXN	removed	Dead end and annotation not specific
PHENOL-BETA-GLUCOSYLTRANSFERASE-RXN	removed	Missing information on products and unspecific annotation
PHOSICITDEHASE-RXN	removed	No protein modification in the model
PHOSPHORYL-RXN	RXN0-5184	
PROTEIN-LYSINE-6-OXIDASE-RXN	removed	No protein modification in the model
PROTOCHLOROPHYLLIDE-REDUCTASE-RXN	RXN1F-10	
PSEUDOURIDYLATE-SYNTHASE-RXN	RXN0-5398	
R222-RXN	removed	Dead end and annotation not specific
R345-RXN	removed	Annotation no longer present
R542-RXN	removed	Dead end and annotation not specific
R9-RXN	3-HYDROXYBUTYRYL-COA-DEHYDRATASE-RXN	irreversible
RIBAZOLEPHOSPHAT-RXN	removed	Dead end and annotation not specific
RXN-10	removed	Dead end and annotation not specific
RXN-1161	FERREDOXIN–NITRITE-REDUCTASE-RXN	
RXN-1883	RXNQT-4141	
RXN-22	removed	Unspecific annotation
RXN-2962	removed	Dead end and annotation not specific

Changes and alterations to reactions from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
RXN-3443	1.2.1.9-RXN	Irreversible NADP variant present and only unspecific annotations
RXN-37	removed	Dead end and annotation not specific
RXN-3742	FOLYLPOLYGLUTAMATE SYNTH-RXN	
RXN-4522	removed	Annotation changed
RXN-4543	RXN0-6274	
RXN-5424	removed	Dead end and annotation not specific
RXN-5444	removed	Dead end and annotation not specific
RXN-5603	removed	Dead end and annotation not specific
RXN-6762	removed	Annotation unspecific
RXN-7562	ORNITHINE-GLU-AMINOTRANSFORASE-RXN	
RXN-7632	removed	Dead end and annotation not specific
RXN-7643	removed	Annotation unspecific
RXN-7657	removed	Dead end and annotation not specific
RXN-7667	RXN-7668	
RXN-7692	removed	Dead end and annotation not specific
RXN-7700	removed	Dead end and annotation not specific
RXN-7706	removed	Dead end and annotation not specific
RXN-902	removed	Dead end and annotation not specific
RXN-982	removed	Dead end and annotation not specific
RXN0-1131	removed	Dead end and annotation not specific
RXN0-1132	1.8.1.4-RXN	
RXN0-382	removed	Annotation for non-deoxy version of this reaction and dead end
RXN1F-158	removed	Energetically unplausible reaction

Changes and alterations to reactions from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
RXN2GG-5427	RXN-9439	
RXN3O-504	removed	Dead end and annotation not specific
SUCCGLU-SEMIALDEHYDE-SUCCGLU-RXN	removed	Dead end and annotation not specific
SUCCINATE-DEHYDROGENASE-(UBIQUINONE)-RXN	SUCCINATE-DEHYDROGENASE-UBIQUINONE-RXN	MetaCyc ID used
SUCCORNTRANSAM-RXN	removed	Dead end and annotation not specific
SUCROSE-PHOSPHORYLASE-RXN	removed	Annotation changed
TRNA-ISOPENTENYLTRANSFERASE-RXN	RXN0-6274	
UBIQUITIN–CALMODULIN-LIGASE-RXN	removed	Dead end and annotation not specific
UBIQUITIN-THIOLESTERASE-RXN	removed	No protein modification in the model
UDP-N-ACETYLGLUCOSAMINE-4-EPIMERASE-RXN	removed	Was removed from the MetaCyc pathway from which it was inferred
YEAST-SQUALENE-RXN	2.5.1.21-RXN + RXN66-281	

Table A.1: Differences to MedicCyc(1.0) with respect to reactions. Details are provided to alterations and reasons for the removal of specific reactions from the database. The most common reason for the exclusion of a reaction was a change in the genome annotation or missing evidence in MedicCyc.

Changes and alterations to pathways from MedicCyc incorporated in the model and database.

MedicCyc identifier	Replacement	Comment
ALANINE-VALINESYN-PWY	removed	This is a bacterial alternative and is replaced by alanine and valine biosynthesis pathways
ANAEROFRUCAT-PWY	PWY-1042+PWY-5481	
ASPBIO-PWY	ASPARTATESYN-PWY	Replaced by MetaCyc version (which omits the first step present in other pathways)
BRANCHED-CHAIN-AA+ALA-PWY	Split into its subpathways	
CAROTENOID-PWY	PWY-695 +PWY-5945 +PWY-5944 +PWY-5943 + PWY-6475 + PWY-5946	Split into many subpathways
DAPLYSINESYN-PWY	PWY-5097	Replaced the bacterial synthesis route by the plant version
DENOVOPURINE2-PWY	PWY-841 + PWY-6126	
DHGLUCONATE-PYR-CAT-PWY	removed	Bacterial pathway (with the only evidence from glycolytic reactions)
DISSULFRED-PWY	SO4ASSIM-PWY	
DRIBOPMET-PWY	PWY0-163	
DTDPRHAMSYN-PWY	PWY-3221	
FUCCAT-PWY	removed	Evidence is unspecific and the only specific enzyme is no longer present in the new annotation
GALACTCAT-PWY	removed	Insufficient evidence

Changes and alterations to pathways from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
GDPRHAMSYN-PWY	removed	Removed GDP-rhamnose biosynthesis as there were too many steps in this pathway missing and other rhamnose activating pathways exist
GLUGLNSYN-PWY	GLUTSYN-PWY	
GLUTAMINE-SYN2-PWY	removed	Implicitly included by glutamate conversions
GLYCOLYSIS-TCA-GLYOX-BYPASS	removed	Removed superpathway
HEXITOLDEGSUPER-PWY	removed	Removed as the initial reactions have no evidence and the reactions with evidence are in common with other pathways
MANNIDEG-PWY	PWY-3861	Mannitol degradation has genetic evidence for the latter pathway.
P202-PWY	SO4ASSIM-PWY	
P23-PWY	removed	The reductive TCA is does only exist in bacteria.
P4-PWY	removed	Removed superpathway
P562-PWY	removed	Only the initial reaction is retained as no other reaction has genetic evidence
PHENYLALANINE-DEG1-PWY	removed	Initial step is a dead end
PHOSLIPSYN2-PWY	removed	Unspecific pathway, would not be part of metabolic model
POLYISOPRENSYN-PWY	multi	Replaced by multiple specific polyisoprenoid biosynthesis routes

Changes and alterations to pathways from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
PRPP-PWY	removed	Only retained the subpathways
PWY-1141	PWY-5945	
PWY-2723	PWY0-1182	No evidence for the first steps, so the the alternative pathway with evidence was used
PWY-2724 + FAO-PWY	removed	These pathways are yeast pathways and do not have much evidence
PWY-2761	removed	No evidence for the end products and unspecific reaction evidence
PWY-282	removed	No specific evidence
PWY-2902	PWY-2902-1	The only reaction with evidence in the new annotation has its own pathway
PWY-2921	removed	Unspecific evidence and no information that the target is actually produced
PWY-3097	removed	Removed superpathway
PWY-31	removed	No evidence and canavanine not present
PWY-3121	removed	No specific evidence especially for the compound
PWY-3701	PWY0-1507	
PWY-3742	PWY-6614	
PWY-381	PWY490-3	

Changes and alterations to pathways from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
PWY-3961 + PWY-3921	PANTO-PWY	PWY-3921 is a subpathway of PANTO-PWY and there is no enzymatic evidence for the PWY-3961 variant
PWY-3981	removed	Changed β-alanine biosynthesis to the *A. thaliana* pathway
PWY-4101	removed	The evidence for the single reaction is not present in the new annotation
PWY-4201	removed	The annotation changed for the reactions in this pathway
PWY-4302	PWY-3781	Aerobic respiration as commonly present in plants and with evidence from gene annotations
PWY-4441	removed	Completely unspecific evidence
PWY-4562	PWY-6361	Replaced by alternative route with more evidence
PWY-4681	removed	Removed the kieveton biosynthesis as there is no evidence of this compound being produced and no gene annotation
PWY-4762	PWY-3385	
PWY-5043	removed	Included in other pathways
PWY-5044	PWY-6607 + PWY-6596	The original pathway was a superpathway including several reactions without evidence

Changes and alterations to pathways from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
PWY-5064	removed	Removed as the functionality is covered by normal chlorophyll biosynthesis and there is no evidence for most steps
PWY-541	removed	Removed as the only parts with evidence are in the jasmonic acid biosynthesis pathway
PWY-561	removed	The pathway had unbalanced reactions and was replaced by a mixture of other pathways
PWY-581	PWY-3181	This is a variant of the tryptophan degradation pathway present in the new database
PWY-782	removed	Macromolecule modification
PWY-862	removed	Polymer degradation, and therefore not part of the model
PWY0-1061	removed	Altered route of alanine biosynthesis without sufficient evidence
PWY0-845	removed	Removed superpathway
PWY0-901	removed	No selenocysteine found
PWY2-2602	PWY-2582	The old pathway is part of the new one
PWY2GG-5212	PWY-641	The target compound is a group of compounds without proper atomic composition information. The replacing pathway, though unspecific better fits to the actual action

Changes and alterations to pathways from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
PWY2GG-5213	PWY-5152 + PWY-5153 + PWY1F-823 + PWY-5125	Superpathway replaced by partial pathways
PWY2GG-5216	PWY-3261	
PWY2GG-5221	removed	Removed superpathway
PWY2GG-5317	PWY-4381	
PWY2GG-5379	PWY-702 + HOMOSERSYN-PWY	The original PW was using the bacterial version of the methionine part. As there is evidence for a kinase activity it was removed and replaced by the plant specific version
PWY2GG-5412	PWY-6120	More likely as there is a prephenate aminotransferase
PWY2GG-5434	removed	Removed superpathway
PWY2GG-5457	removed	Replaced by alternate pathways producing mannitol with more evidence
PWY2GG-5462	PWY-695	
PWY2GG-5571	PWY-981 + PWY-6406	Superpathway replaced by subpathways
PWY2GG-5807	removed	Removed superpathway
PWY2GG-5872	removed	Removed superpathway
PWY66-201	removed	Rmoved nicotine from the system as there is no evidence for its biosynthesis in *M. truncatula*
PWY66-221	removed	Nicotine removed
PYRIDNUCSAL-PWY	removed	NAD Salvage II performs similar conversions and there is not enough evidence for this variant

Changes and alterations to pathways from MedicCyc (continued)

MedicCyc identifier	Replacement	Comment
PYRIDOXSYN-PWY	PWY-6466	
SULFATE-CYS-PWY	SO4ASSIM-PWY + CYSTSYN-PWY	This pathway is replaced by 2 of its subpathways. In addition one reaction got removed due to the lack of evidence
THISYN-PWY	THISYNARA-PWY	The new pathway is more likely to exist in plants
THREOCAT-PWY	removed	All reactions with evidence in this pathway are included but a large fraction did not have any evidence
TRYPTOPHAN-DEGRADATION-1	removed	No evidence for any intermediate reaction
URDEGR-PWY	PWY-5698 + PWY-5704	These pathways are a branch of the original pathway with more actual enzymatic evidence

Table A.2: Differences to MedicCyc(1.0) with respect to pathways. Details are provided to alterations and reasons for the removal of specific pathways from the database.

Appendix B

Seeds and targets for compartmentalisation

Table B.1: Seeds and targets for Network Extension in compartmentalisation

Golgi Apparatus

Seeds

Compound	MetaCyc IDs	Source
UDP and GDP sugars	UDP-GLUCOSE, GUANOSINE_DIPHOSPHATE _FUCOSE, UDP-D-XYLOSE, UDP-GALACTOSE, UDP-D-GALACTURONATE	[193]

Targets

Compound	MetaCyc IDs	Source
hemicellulose, xyloglucans, pectins	XLFG-Xyloglucans, Glucomannans, Xylogalacturonan	[244]
UMP, GMP, Pi	UMP, GMP, Pi	[193]

Suggested Extension

1.1.1.271-RXN : NADPH + H^+ + GDP-4-dehydro-6-deoxy-D-mannose $\rightarrow$ GDP-L-fucose + $NADP^+$
CELLULOSE-SYNTHASE-UDP-FORMING-RXN : UDP-D-glucose $\rightarrow$ H^+ + UDP + cellulose
GDPMANDEHYDRA-RXN : GDP-α-D-mannose $\rightarrow$ H_2O + GDP-4-dehydro-6-deoxy-D-mannose
RXN-12418 : UDP-D-galactose + an XXLG xylogulcan $\rightarrow$ H^+ + an XLLG xylogulcan + UDP
RXNDLN-35 : a homogalacturonan $\leftrightarrow$ a pectate
UDP-GLUCURONATE-4-EPIMERASE-RXN : UDP-D-glucuronate $\leftrightarrow$ UDP-D-galacturonate
UGD-RXN : H_2O + UDP-D-glucose + NAD^+ $\rightarrow$ UDP-D-glucuronate + H^+ + NADH

Endoplasmic reticulum

Seeds

Compound	MetaCyc IDs	Source
fatty acids	PALMITOYL-COA, STEAROYL-COA, CDP-CHOLINE, GLYCEROL-3P	[244]
UDP-galactose, ATP	UDP-GALACTOSE, ATP	[193]
allantoin	S-ALLANTOIN	[245]

Targets

ceramides, lipids, modified lipids, cholesterol	1-2-DIPALMITOYLPHOSPHATIDYLCHOLINE, CPD0-1422	[244]
allantoic acid	ALLANTOATE	[245]

Suggested Extension

ALLANTOINASE-RXN : *S*-allantoin + H_2O $\rightarrow$ allantoate + H^+
DIACYLGLYKIN-RXN : 1,2-dipalmitoylglycerol + ATP $\leftrightarrow$ ADP + dipalmitoyl phosphatidate + H^+
RXN-5781 : 1,2-dipalmitoylglycerol + CDP-choline $\rightarrow$ 1,2-dipalmitoyl-phosphatidylcholine + CMP + H^+
RXNDLN-4 : palmitoyl-CoA + *sn*-glycerol-3-phosphate $\leftrightarrow$ 1-palmitoylglycerol 3-phosphate + coenzyme A

Peroxisome

Seeds

Name in Source	MetaCyc IDs	Source
fatty acids, serine, glycolate	SER, GLYCOLLATE, PALMITATE, STEARIC_ACID	[244]
sulfite	SO3	[246]
uric acid	URATE	[245]
glutamate	GLT	[247]
oxaloacetate, monodehydroascorbate, serine, isocitrate, jasmonic acid precursors	OXALACETIC_ACID,CPD-318, SER, THREO-DS-ISO-CITRATE, CPD-730	[248]
ATP	ATP	[193]

Targets

glucose, glycerate, glycine	GLC, GLY, GLYCERATE	[244]
sulfate	SULFATE	[246]
allantoin	S-ALLANTOIN	[245]
oxoglutarate	2-KETOGLUTARATE	[247]
acetyl-CoA, citrate, glycerate, ascorbate, malate, succinate, jasmonic Acid	ACETYL-COA, CIT, GLYCERATE, ASCORBATE, MAL, SUC, CPD-734	[248]
AMP, succinate, citrate	AMP, SUC, CIT	[193]

Suggested Extension

RXN-745 : (+)-7-iso-jasmonate $\leftrightarrow$ (-)-jasmonate
RXN-6201 : 5-hydroxy-2-oxo-4-ureido-2,5-dihydro-1H imidazole-5-carboxylate + H^+ $\rightarrow$ CO_2 + *S*-allantoin
RXN-10695 : 3-oxo-2-(cis-2'-pentenyl)-cyclopentane-1-octanoate + ATP + coenzyme A $\rightarrow$ diphosphate + OPC8-CoA + AMP
RXN-10696 : oxygen + OPC8-CoA $\rightarrow$ hydrogen peroxide + OPC8-trans-2-enoyl-CoA
RXN-10706 : oxygen + OPC6-CoA $\rightarrow$ OPC6-trans-2-enoyl-CoA + hydrogen peroxide
RXN-10707 : oxygen + OPC4-CoA $\rightarrow$ hydrogen peroxide + OPC4-trans-2-enoyl-CoA
RXN-10708 : jasmonoyl-CoA + H_2O $\rightarrow$ (+)-7-iso-jasmonate + H^+ + coenzyme A

Plastid

Seeds

Name in Source	MetaCyc IDs	Source
glycerate	CLYCERATE	[244]
malate, glutamine, xanthine, DHAP	MAL, GLN, XANTHINE, DIHYDROXY-ACETONE-PHOSPHATE	[245]
glucose-6P, xylulose-5P, erythrose-4P, triosephosphates, phosphoenolpyruvate, pyruvate, nitrite, sulfate, tetrahydrofolate, S-adenosyl-methionine	ALPHA-GLC-6-P, GLC-6-P, XYLULOSE-5-P, ERYTHROSE-4-P, DIHYDROXY-ACETONE-PHOSPHATE, PEP, GAP, PYRUVATE, THF, SULFATE, NITRITE, S-ADENOSYLMETHIONINE	[193]
ADP or ATP (depending on type of plastid)	ATP, ADP	[21]
CO_2, light, magnesium, cytosine triphosphate, "lipid carriers", "O_2, water, guanosine triphosphate	CARBON-DIOXIDE, Light, MG+2, GLN, CTP, Lipid, ACP, OXYGEN-MOLECULE, WATER, GTP	

Targets

Name in Source	MetaCyc IDs	Source
glycollate	GLYCOLLATE	[244]
oxaloacetate, glutamate, aspartate, fumarate	OXALACETIC_ACID, GLT, L-ASPARTATE, FUM	[245]
jasmonic acid precursor	CPD-730	[248]
phosphate, 3-phosphoglycerate, malate, glucose, ammonia, sulfite, cysteine, methionine, acetyl-coa, chlorophyll, isopentenyl-diphosphate, d-carotene, b-carotene, zeaxanthin, violaxanthin, capsanthin, capsorubin, GTP, tryptophan	Pi, 3-PG, MAL, GLC, AMMONIA, SO3, CYS, MET, ACETYL-COA, CHLOROPHYLL-A, DELTA3-ISOPENTENYL-PP, CPD1F-129, CPD1F-115, CPD1F-130, CPD1F-133, CAPSANTHIN, CAPSORUBIN, GTP, TRP	[193]

starch, maltose, glucose, chlorophyll, palmitate, stearate, linoleate, linoleic acid, oleate	Starch, MALTOSE, GLC, CHLOROPHYLL-A, PALMITATE, STEARIC_ACID, LINOLENIC_ACID, LINOLEIC_ACID, OLEATE-CPD	[21]
methionine, valine, leucin, isoleucine, histidine	MET, VAL, LEU, ILE, HIS	[249, 250, 251]
lysine	LYS	[252]
phenylalanine, tyrosine, tryptophan	PHE, TYR, TRP	[253]
serine	SER	[254]
methionine, threonine	MET, THR	[255]
cysteine	CYS	[256]

Suggested Extension

12-OXOPHYTODIENOATE-REDUCTASE-RXN : 12-oxo-cis-10,15-phytodienoate + H^+ + NADPH $\rightarrow$ $NADP^+$ + 3-oxo-2-(cis-2'-pentenyl)-cyclopentane-1-octanoate
3.1.2.14-RXN : an oleoyl-[acp] + H_2O $\rightarrow$ a holo-[acp] + oleate + H^+
ACETOOHBUTSYN-RXN : pyruvate + H^+ + 2-oxobutanoate $\rightarrow$ CO_2 + 2-aceto-2-hydroxy-butanoate
ADCLY-RXN : 4-amino-4-deoxychorismate $\rightarrow$ pyruvate + H^+ + 4-aminobenzoate
FPPSYN-RXN : geranyl diphosphate + isopentenyl diphosphate $\leftrightarrow$ diphosphate + (2*E*,6*E*)-farnesyl diphosphate
GPPSYN-RXN : isopentenyl diphosphate + dimethylallyl diphosphate $\rightarrow$ geranyl diphosphate + diphosphate
HISTIDPHOS-RXN : L-histidinol-phosphate + H_2O $\rightarrow$ phosphate + histidinol
HISTPRATPHYD-RXN : phosphoribosyl-ATP + H_2O $\rightarrow$ diphosphate + H^+ + 1-(5-phospho-D-ribosyl)-AMP
IMIDPHOSDEHYD-RXN : D-erythro-imidazole-glycerol-phosphate $\rightarrow$ imidazole acetol-phosphate + H_2O
RIBOFLAVINSYNDEAM-RXN : 2,5-diamino-6-(5-phospho-D-ribosylamino)pyrimidin-4(3H)-one + H_2O $\rightarrow$ 5-amino-6-(5-phospho-D-ribosylamino)uracil + ammonia
RIBOFLAVINSYNREDUC-RXN : 5-amino-6-(5-phospho-D-ribosylamino)uracil + H^+ + NADPH $\rightarrow$ $NADP^+$ + 5-amino-6-(5-phospho-D-ribitylamino)uracil
RIBOPHOSPHAT-RXN : 5-amino-6-(5-phospho-D-ribitylamino)uracil + H_2O $\rightarrow$ phosphate + 5-amino-6-(D-ribitylamino)uracil

RXN-5282 : oxygen + H^+ + magnesium-protoporphyrin IX 13-monomethyl ester + NADPH $\rightarrow$ $NADP^+$ + 13^1-hydroxy-magnesium-protoporphyrin IX 13-monomethyl ester + H_2O
RXN-5283 : oxygen + 13^1-hydroxy-magnesium-protoporphyrin IX 13-monomethyl ester + H^+ + NADPH $\rightarrow$ $NADP^+$ + 13^1-oxo-magnesium-protoporphyrin IX 13-monomethyl ester + H_2O
RXN-5284 : oxygen + 13^1-oxo-magnesium-protoporphyrin IX 13-monomethyl ester + H^+ + NADPH $\rightarrow$ $NADP^+$ + divinyl protochlorophyllide *a* + H_2O
RXN-7800 : (2*S*)-2-isopropyl-3-oxosuccinate + H^+ $\rightarrow$ CO_2 + 4-methyl-2-oxopentanoate
RXN-8025 : NADH + oxygen + H^+ + β-carotene $\rightarrow$ β-cryptoxanthin + NAD^+ + H_2O
RXN-8026 : NADH + oxygen + H^+ + β-cryptoxanthin $\rightarrow$ zeaxanthin + NAD^+ + H_2O
RXN-8042 : prolycopene $\leftrightarrow$ *all-trans*-lycopene
RXN-9548 : a stearoyl-[acp] + H_2O $\rightarrow$ a holo-[acp] + H^+ + stearate
RXN-9549 : a palmitoyl-[acp] + H_2O $\leftrightarrow$ a holo-[acp] + palmitate + H^+
RXN-9667 : a reduced flavodoxin + oxygen + a lipid linoleoyl group $\rightarrow$ an oxidized flavodoxin + a lipid α-linolenoyl group + H_2O
RXN-9670 : a lipid + oleoyl-CoA $\rightarrow$ a lipid oleoyl-group + coenzyme A
RXN-11354 : 9,15,9'-tri-*cis*-ζ-carotene $\leftrightarrow$ 9,9'-di-*cis*-ζ-carotene
RXN-12243 : 15-*cis*-phytoene + a plastoquinone $\rightarrow$ a plastoquinol + 15,9'-di-*cis*-phytofluene
RXN-12244 : 15,9'-di-*cis*-phytofluene + a plastoquinone $\rightarrow$ a plastoquinol + 9,15,9'-tri-*cis*-ζ-carotene
RXN0-882 : a reduced ferredoxin + 2-C-methyl-D-erythritol-2,4-cyclodiphosphate + H^+ $\rightarrow$ 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate + an oxidized ferredoxin + H_2O
RXN1F-72 : divinyl protochlorophyllide *a* + H^+ + NADPH $\rightarrow$ $NADP^+$ + monovinyl protochlorophyllide *a*
RXN1F-147 : *all-trans*-lycopene $\leftrightarrow$ δ-carotene
RXNDLN-17 : a lipid α-linolenoyl group + H_2O $\rightarrow$ a lipid + H^+ + α-linolenate
RXNDLN-18 : a lipid linoleoyl group + H_2O $\rightarrow$ a lipid + H^+ + linoleate
SPONTPRO-RXN : H^+ + 2-aceto-2-hydroxy-butanoate + NADPH $\leftrightarrow$ $NADP^+$ + 2,3-dihydroxy-3-methylvalerate

Mitochondrion

Seeds

Name in Source	MetaCyc IDs	Source
glycine	GLY	[244]
AMP, ADP, pyruvate, oxcaloacetate, succinate, arginine, ornithine, phosphate, lysine, histidine, citrulline	AMP, ADP, PYRUVATE, OXALACETIC_ACID, SUC, ARG, ORNITHINE, Pi, LYS, HIS, CITRULLINE	[193]
methionine, s-adenosyl-methionine	MET, S-ADENOSYL-METHIONINE	[255]
glutamate	GLT	[257]
formate, serine, glycine	FORMATE, SER, GLY, OXALACETIC_ACID	[258]
para-aminobenzoate, HMDHP	P-AMINO-BENZOATE, AMINO-OH-HYDROXYMETHYL-DIHYDROPTERIDINE	[259]

Targets

Name in Source	MetaCyc IDs	Source
serine	SER	[244]
ATP, ornithine, citrate, arginine,	ATP, ORNITHINE, CIT, ARG,	[193]
2-ketoglutarate, malate, fumarate	2-KETOGLUTARATE, MAL, FUM	
short chain fatty acids	Octanoyl-ACP (giving ACP as seed)	[260]
keto-pantoate	L-PANTOATE	[261]
adenosyl-homo-cystenine	ADENOSYL-HOMO-CYS	[255]
isocitrate, malate	THREO-DS-ISO-CITRATE, MAL	[262]
tetra-hydro-folate	THF	[257, 259]

Suggested Extension

2-DEHYDROPANTOATE-REDUCT-RXN : H^+ + NADPH + 2-dehydropantoate $\rightarrow$ $NADP^+$ + (R)-pantoate
4.2.1.58-RXN : an (*R*)-3-hydroxybutanoyl-[acp] $\rightarrow$ a crotonyl-[acp] + H_2O
4.2.1.59-RXN : an (*R*)-3-hydroxyoctanoyl-[acp] $\rightarrow$ a *trans* oct-2-enoyl-[acp] + H_2O

ACETYL-COA-CARBOXYLTRANSFER-RXN : ATP + acetyl-CoA + bicarbonate $\rightarrow$ H^+ + ADP + malonyl-CoA + phosphate
CARBAMATE-KINASE-RXN : H^+ + ADP + carbamoyl-phosphate $\rightarrow$ CO_2 + ATP + ammonia
DIHYDROFOLATEREDUCT-RXN : H^+ + NADPH + 7,8-dihydrofolate monoglutamate $\rightarrow$ tetrahydrofolate + $NADP^+$
DIHYDROFOLATESYNTH-RXN : ATP + L-glutamate + 7,8-dihydropteroate $\rightarrow$ H^+ + ADP + 7,8-dihydrofolate monoglutamate + phosphate
H2PTERIDINEPYROPHOSPHOKIN-RXN : ATP + 6-hydroxymethyl-7,8-dihydropterin $\rightarrow$ H^+ + AMP + 6-hydroxymethyl-dihydropterin diphosphate
H2PTEROATESYNTH-RXN : 4-aminobenzoate + 6-hydroxymethyl-dihydropterin diphosphate $\rightarrow$ diphosphate + 7,8-dihydropteroate
MALONYL-COA-ACP-TRANSACYL-RXN : a holo-[acp] + H^+ + malonyl-CoA $\rightarrow$ a malonyl-[acp] + coenzyme A
MALONYL-ACPDECARBOX-RXN : a malonyl-[acp] $\rightarrow$ an acetyl-[acp] + CO_2
RXN-9514 : an acetoacetyl-[acp] + H^+ + NADPH $\rightarrow$ $NADP^+$ + an (*R*)-3-hydroxybutanoyl-[acp]
RXN-9516 : a butyryl-[acp] + a malonyl-[acp] $\rightarrow$ a holo-[acp] + a 3-oxo-hexanoyl-[acp] + CO_2
RXN-9518 : a 3-oxo-hexanoyl-[acp] + H^+ + NADPH $\rightarrow$ $NADP^+$ + an (*R*)-3-hydroxyhexanoyl-[acp]
RXN-9520 : an (*R*)-3-hydroxyhexanoyl-[acp] $\rightarrow$ a *trans* hex-2-enoyl-[acp] + H_2O
RXN-9523 : a malonyl-[acp] + a hexanoyl-[acp] $\rightarrow$ a holo-[acp] + CO_2 + a 3-oxo-octanoyl-[acp]
RXN-9524 : H^+ + a 3-oxo-octanoyl-[acp] + NADPH $\rightarrow$ $NADP^+$ + an (*R*)-3-hydroxyoctanoyl-[acp]
RXN-9657 : NADH + a crotonyl-[acp] + H^+ $\rightarrow$ a butyryl-[acp] + NAD^+
RXN-9658 : NADH + H^+ + a *trans* hex-2-enoyl-[acp] $\rightarrow$ a hexanoyl-[acp] + NAD^+
RXN-9659 : NADH + a *trans* oct-2-enoyl-[acp] + H^+ $\rightarrow$ an octanoyl-[acp] + NAD^+

Table B.1: Seeds and targets used for the Network Extension algorithm in the compartmentalisation step. The suggested extension is listed with the reaction stoichiometries.

References

[1] Voet D and Voet J. (2010). *Biochemistry*. Wiley, 4th edition. 1, 4, 6, 10, 12, 145

[2] Metzler DE (2003). *Biochemistry*, volume 1+2. Academic Press, 2nd edition. 1, 4, 12, 31

[3] Pidwirny M (2013). Fundamentals of Physical Geography (2nd Edition) (2013). 2

[4] Lavoisier A (1790). *Elements of Chemistry tr. by R. Kerr*. Robert Kerr. 2

[5] Zahran HH. *Rhizobium*-Legume Symbiosis and Nitrogen Fixation under Severe Conditions and in an Arid Climate. *Microbiol Mol Biol Rev* (1999), **63**(4), 968–89, table of contents. 2

[6] Wagner SC. Biological Nitrogen Fixation. *Nature Education Knowledge* (2012), **3(10)**, 15. 2

[7] Kinzig AP and Socolow RH. Human Impacts on the Nitrogen Cycle. *Physics Today* (1994), **47**(11), 24–31. 2

[8] BASF Ludwigsburg (1908). Verfahren zur synthetischen Darstellung von Ammoniak aus den Elementen (1908). 2

[9] Crick FH. On protein synthesis. *Sym Soc Exp Biol* (1958), **12**, 138–163. 4

[10] Crick FH. Central dogma of molecular biology. *Nature* (1970), **227**(5258), 561–563. 4

[11] Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* (2003), **72**, 291–336. 6

[12] Altschul SF, Gish W, Miller W, *et al.* Basic local alignment search tool. *J Mol Biol* (1990), **215**(3), 403–410. 6

[13] Altschul SF, Madden TL, Schäffer AA, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* (1997), **25**(17), 3389–3402. 6

[14] Gajewski E, Steckler DK, and Goldberg RN. Thermodynamics of the hydrolysis of adenosine 5'-triphosphate to adenosine 5'-diphosphate. *J Biol Chem* (1986), **261**(27), 12733–12737. 7

[15] Poole RJ. Energy Coupling for Membrane Transport. *Annu Rev Plant Physiol* (1978), **29**, 437–60. 9

[16] Theg SM, Bauerle C, Olsen LJ, *et al.* Internal ATP is the Only Energy Requirement for the Translocation of Precursor Proteins across Chloroplastic Membranes. *J Biol Chem* (1989), **264**(12), 6730–6736. 9

[17] Shi LX and Theg SM. Energetic cost of protein import across the envelope membranes of chloroplasts. *Proc Natl Acad Sci U S A* (2013), **110**(3), 930–935. 9

[18] Sheetz MP, Chasan R, and Spudich JA. ATP-dependent movement of myosin in vitro: characterization of a quantitative assay. *J Cell Biol* (1984), **99**(5), 1867–1871. 9

[19] Pyl ET, Piques M, Ivakov A, *et al.* Metabolism and Growth in *Arabidopsis* Depend on the Daytime Temperature but Are Temperature-Compensated against Cool Nights. *Plant Cell* (2012), **24**(6), 2443–2469. 9, 81, 97, 99

[20] Williams TCR, Miguet L, Masakapalli SK, *et al.* Metabolic Network Fluxes in Heterotrophic Arabidopsis Cells: Stability of the Flux Distribution under Different Oxygenation Conditions. *Plant Physiol* (2008), **148**(2), 704–718. 9, 69

[21] Taiz L and Zeiger E. (2006). *Plant Physiology.* Sinauer Associated, Inc, 4th edition. 10, 11, 12, 14, 17, 167, 168

[22] Mrabet Y (2011). Glycolysis. https://commons.wikimedia.org/wiki/File:Glycolysis.svg (2011). 11, 19, 21

[23] Stitt M. Fructose-2,6-Bisphosphate as a Regulatory Molecule in Plants. *Ann Rev Plant Phys* (1990), **41**(1), 153–185. 10

[24] Singh BK (1998). *Biosynthesis of Valine, Leucine and Isoleucine*, chapter 7, page 227ff. CRC Press. 11

[25] Mrabet Y (2011). Citric_acid_cycle_with_aconitate_2. https://commons.wikimedia.org/wiki/File:Citric_acid_cycle_with_aconitate_2.svg (2011). 13

[26] Hatefi Y. The mitochondrial electron transport and oxidative phosphorylation system. *Annu Rev Biochem* (1985), **54**, 1015–1069. 12

[27] Gray MW. Mitochondrial evolution. *Cold Spring Harb Perspect Biol* (2012), **4**(9), a011403. 14

[28] Fvasconcellos (2007). Schematic diagram of the mitochondrial electron transport chain. http://commons.wikimedia.org/wiki/File: Mitochondrial_electron_transport_chain-Etc4.svg (2007). 15

[29] Alexandre A, Reynafarje B, and Lehninger AL. Stoichiometry of vectorial H+ movements coupled to electron transport and to ATP synthesis in mitochondria. *Proc Natl Acad Sci U S A* (1978), **75**(11), 5296–5300. 14

[30] Petersen J, Förster K, Turina P, *et al.* Comparison of the H+/ATP ratios of the H+-ATP synthases from yeast and from chloroplast. *Proc Natl Acad Sci U S A* (2012), **109**(28), 11150–11155. 14

[31] Bryant DA and Frigaard NU. Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol* (2006), **14**(11), 488–496. 14

[32] Hohmann-Marriott MF and Blankenship RE. Evolution of photosynthesis. *Annu Rev Plant Biol* (2011), **62**, 515–548. 16

[33] Ellis R. The most abundant protein in the world. *Trends Biochem Sci* (1979), **4**(11), 241–244. 17

[34] Kruger NJ and von Schaewen A. The oxidative pentose phosphate pathway: structure and organisation. *Curr Opin Plant Biol* (2003), **6**(3), 236–246. 20

[35] Cook DR. *Medicago truncatula*–a model in the making! *Curr Opin Plant Biol* (1999), **2**(4), 301–304. 23, 24

[36] Meinke DW, Cherry JM, Dean C, *et al. Arabidopsis thaliana*: a model plant for genome analysis. *Science* (1998), **282**(5389), 662, 679–662, 682. 23

[37] Somerville C and Koornneef M. A fortunate choice: the history of *Arabidopsis* as a model plant. *Nat Rev Genet* (2002), **3**(11), 883–889. 23

[38] Young ND and Udvardi M. Translating *Medicago truncatula* genomics to crop legumes. *Curr Opin Plant Biol* (2009), **12**(2), 193–201. 24

[39] Young ND and Bharti AK. Genome-enabled insights into legume biology. *Annu Rev Plant Biol* (2012), **63**, 283–305. 24

[40] Barker D, Bianchi S, Blondon F, *et al.* *Medicago truncatula*, a model plant for studying the molecular genetics of the Rhizobium–legume symbiosis. *Plant Mol. Biol. Report.* (1990), **8**(1), 40–49–. 24

[41] Jones KM, Sharopova N, Lohar DP, *et al.* Differential response of the plant *Medicago truncatula* to its symbiont *Sinorhizobium meliloti* or an exopolysaccharide-deficient mutant. *Proc Natl Acad Sci U S A* (2008), **105**(2), 704–709. 24, 28

[42] Young ND, Debelle F, Oldroyd GED, *et al.* The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* (2011), **480**(7378), 520–524. 24

[43] Colditz F and Braun HP. *Medicago truncatula* proteomics. *J Proteomics* (2010), **73**(10), 1974–1985. 24

[44] Broeckling CD, Huhman DV, Farag MA, *et al.* Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. *J Exp Bot* (2005), **56**(410), 323–336. 24, 56, 146

[45] Barsch A, Carvalho HG, Cullimore JV, *et al.* GC-MS based metabolite profiling implies three interdependent ways of ammonium assimilation in *Medicago truncatula* root nodules. *J Biotechnol* (2006), **127**(1), 79–83. 24, 56

[46] Farag MA, Huhman DV, Dixon RA, *et al.* Metabolomics reveals novel pathways and differential mechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in *Medicago truncatula* cell cultures. *Plant Physiol* (2008), **146**(2), 387–402. 24

[47] Medicago truncatula Genome Project (2013). http://medicago.jcvi.org/cgi-bin/medicago/annotation.cgi (2013). 25, 52

[48] He J, Benedito VA, Wang M, *et al.* The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics* (2009), **10**, 441. 25

[49] Benedito VA, Torres-Jerez I, Murray JD, *et al.* A gene expression atlas of the model legume *Medicago truncatula*. *Plant J* (2008), **55**(3), 504–513. 25, 121

[50] Urbanczyk-Wochniak E and Sumner LW. MedicCyc: a biochemical pathway database for *Medicago truncatula*. *Bioinformatics* (2007), **23**(11), 1418–1423. 25, 53, 72

[51] Barker DG, Pfaff T, Moreau D, *et al.* (2006). *The Medicago truncatula handbook*, chapter Growing *M. truncatula*: choice of substrates and growth conditions, pages 1–26. 25, 67

[52] Watson BS, Asirvatham VS, Wang L, *et al.* Mapping the proteome of barrel medic (*Medicago truncatula*). *Plant Physiol* (2003), **131**(3), 1104–1123. 25

[53] Mettupalli DR (2011). *Impact of Arbuscular Mycorrhiza symbiosis on photosynthesis in* Medicago truncatula. PhD thesis, Göteborgs Universitet. 25, 85

[54] Schliemann W, Ammer C, and Strack D. Metabolite profiling of mycorrhizal roots of *Medicago truncatula*. *Phytochemistry* (2008), **69**(1), 112–146. 25

[55] Postgate JR (1982). *The fundamentals of nitrogen fixation*. Cambridge University Press. 26

[56] Gallon J. The oxygen sensitivity of nitrogenase: a problem for biochemists and micro-organisms. *Trends Biochem Sci* (1981), **6**(0), 19–23. 26

[57] Batut J and Boistard P. Oxygen control in Rhizobium. *Antonie Van Leeuwenhoek* (1994), **66**(1-3), 129–150. 26

[58] van Rhijn P and Vanderleyden J. The *Rhizobium*-Plant Symbiosis. *Microbiol Rev* (1995), **59**(1), 124–142. 26

[59] Galibert F, Finan TM, Long SR, *et al.* The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* (2001), **293**(5530), 668–672. 26

[60] Ampe F, Kiss E, Sabourdy F, *et al.* Transcriptome analysis of *Sinorhizobium meliloti* during symbiosis. *Genome Biol* (2003), **4**(2), R15. 26

[61] Capela D, Filipe C, Bobik C, *et al. Sinorhizobium meliloti* Differentiation During Symbiosis with Alfalfa: A Transcriptomic Dissection. *Mol Plant-Microbe Interact* (2006), **19**(4), 363–372. 26

[62] Hellweg C, Pühler A, and Weidner S. The time course of the transcriptomic response of *Sinorhizobium meliloti* 1021 following a shift to acidic pH. *BMC Microbiol* (2009), **9**, 37. 26

[63] Rüberg S, Tian ZX, Krol E, *et al.* Construction and validation of a *Sinorhizobium meliloti* whole genome DNA microarray: genome-wide profiling of osmoadaptive gene expression. *J Biotechnol* (2003), **106**(2-3), 255–268. 26

[64] Djordjevic MA. *Sinorhizobium meliloti* metabolism in the root nodule: A proteomic perspective. *Proteomics* (2004), **4**(7), 1859–1872. 26

[65] Guerreiro N, Djordjevic MA, and Rolfe BG. Proteome analysis of the model microsymbiont *Sinorhizobium meliloti* : Isolation and characterisation of novel proteins. *Electrophoresis* (1999), **20**(4-5), 818–825. 26

[66] Barsch A, Patschkowski T, and Niehaus K. Comprehensive metabolite profiling of *Sinorhizobium meliloti* using gas chromatography-mass spectrometry. *Funct Integr Genomic* (2004), **4**, 219–230. 10.1007/s10142-004-0117-y. 26

[67] Keum YS, Seo JS, Li QX, *et al.* Comparative metabolomic analysis of *Sinorhizobium* sp. C4 during the degradation of phenanthrene. *Appl Microbiol Biotechnol* (2008), **80**(5), 863–872. 26

[68] Zhao H, Li M, Fang K, *et al.* In silico insights into the symbiotic nitrogen fixation in *Sinorhizobium meliloti* via metabolic reconstruction. *PLoS One* (2012), **7**(2), e31287. 26, 123

[69] Caspi R, Altman T, Dale JM, *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* (2010), **38**(Database issue), D473–D479. 26, 36, 52

[70] Ludwig RA. *Rhizobium* free-living nitrogen fixation occurs in specialized non-growing cells. *Proc Natl Acad Sci U S A* (1984), **81**(5), 1566–1569. 27

[71] Quinto C, Vega HDL, Flores M, *et al.* Nitrogenase reductase: A functional multigene family in *Rhizobium phaseoli*. *Proc Natl Acad Sci U S A* (1985), **82**(4), 1170–1174. 27

[72] Fischer HM. Genetic regulation of nitrogen fixation in rhizobia. *Microbiol Rev* (1994), **58**(3), 352–386. 27

[73] Hellriegel H, Wilfarth H, Roemer H, *et al.* (1888). *Untersuchungen über die Stickstoffnahrung- der Gramineen und Leguminosen*. Berlin, Buchdruckerei der "Post" Kayssler. 28

[74] Crespi and Gálvez. Molecular Mechanisms in Root Nodule Development. *J Plant Growth Regul* (2000), **19**(2), 155–166. 28

[75] Monahan-Giovanelli H, Pinedo CA, and Gage DJ. Architecture of infection thread networks in developing root nodules induced by the symbiotic bacterium *Sinorhizobium meliloti* on *Medicago truncatula*. *Plant Physiol* (2006), **140**(2), 661–670. 28

[76] Ferguson BJ, Indrasumunar A, Hayashi S, *et al.* Molecular Analysis of Legume Nodule Development and Autoregulation. *J Integr Plant Biol* (2010), **52**(1), 61–76. 28, 29, 30

[77] Pueppke SG and Broughton WJ. *Rhizobium* sp. Strain NGR234 and *R. fredii* USDA257 Share Exceptionally Broad, Nested Host Ranges. *Mol Plant Microbe Interact* (1999), **12**(4), 293–318. 28

[78] Phillips DA and Tsai SM. Flavonoids as plant signals to rhizosphere microbes. *Mycorrhiza* (1992), **1**, 55–58. 10.1007/BF00206136. 28

[79] Caetano-Anollés G, Crist-Estes DK, and Bauer WD. Chemotaxis of *Rhizobium meliloti* to the plant flavone luteolin requires functional nodulation genes. *J Bacteriol* (1988), **170**(7), 3164–3169. 28

[80] Spaink HP. Root nodulation and infection factors produced by rhizobial bacteria. *Annu Rev Microbiol* (2000), **54**, 257–288. 30

[81] J. Dénarié J, Debellé F, and Promé JC. *Rhizobium* lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annu Rev Biochem* (1996), **65**, 503–535. 30

[82] Esseling JJ, Lhuissier FGP, and Emons AMC. Nod Factor-Induced Root Hair Curling: Continuous Polar Growth towards the Point of Nod Factor Application. *Plant Physiol* (2003), **132**(4), 1982–1988. 30

[83] Oldroyd GED and Downie JA. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu Rev Plant Biol* (2008), **59**, 519–546. 30

[84] Fournier J, Timmers ACJ, Sieberer BJ, *et al.* Mechanism of infection thread elongation in root hairs of *Medicago truncatula* and dynamic interplay with associated rhizobial colonization. *Plant Physiol* (2008), **148**(4), 1985–1995. 30

[85] Brewin N. Plant Cell Wall Remodelling in the RhizobiumLegume Symbiosis. *Crit Rev Plant Sci* (2004), **23**(4), 293–316. 30

[86] Kereszt A, Mergaert P, and Kondorosi E. Bacteroid development in legume nodules: evolution of mutual benefit or of sacrificial victims? *Mol Plant Microbe Interact* (2011), **24**(11), 1300–1309. 30

[87] Roth LE and Stacey G. Bacterium release into host cells of nitrogen-fixing soybean nodules: the symbiosome membrane comes from three sources. *Eur J Cell Biol* (1989), **49**(1), 13–23. 30

[88] Mergaert P, Uchiumi T, Alunni B, *et al.* Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc Natl Acad Sci U S A* (2006), **103**(13), 5230–5235. 30

[89] Gage DJ. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol Mol Biol Rev* (2004), **68**(2), 280–300. 31

[90] Oke V and Long SR. Bacteroid formation in the *Rhizobium*-legume symbiosis. *Curr Opin Microbiol* (1999), **2**(6), 641–646. 31

[91] Haag AF, Arnold MFF, Myka KK, *et al.* Molecular insights into bacteroid development during *Rhizobium*-legume symbiosis. *FEMS Microbiol Rev* (2012). 31

[92] Udvardi MK and Day DA. Metabolite Transport Across Symbiotic Membranes of Legume Nodules. *Annu Rev Plant Physiol Plant Mol Biol* (1997), **48**, 493–523. 31, 32, 33

[93] Udvardi M and Poole PS. Transport and metabolism in legume-rhizobia symbioses. *Annu Rev Plant Biol* (2013), **64**, 781–805. 31, 32, 33

[94] Appleby CA. Leghemoglobin and *Rhizobium* Respiration. *Annu. Rev. Plant. Physiol.* (1984), **35**(1), 443–478. 31

[95] Ott T, van Dongen JT, Günther C, *et al.* Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Curr Biol* (2005), **15**(6), 531–535. 31

[96] Prell J and Poole P. Metabolic changes of rhizobia in legume nodules. *Trends Microbiol* (2006), **14**(4), 161–168. 32

[97] Bolton E, Higgisson B, Harrington A, *et al.* Dicarboxylic acid transport in *Rhizobium meliloti* isolation of mutants and cloning of dicarboxylic acid transport genes. *Arch Microbiol* (1986), **144**(2), 142–146. 32

[98] Finan TM, Wood JM, and Jordan DC. Symbiotic properties of C4-dicarboxylic acid transport mutants of *Rhizobium leguminosarum*. *J Bacteriol* (1983), **154**(3), 1403–1413. 32

[99] Salminen SO and Streeter JG. Involvement of glutamate in the respiratory metabolism of *Bradyrhizobium japonicum* bacteroids. *J Bacteriol* (1987), **169**(2), 495–499. 33

[100] Lodwig EM, Hosie AHF, Bourds A, *et al.* Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. *Nature* (2003), **422**(6933), 722–726. 33, 128, 132, 133, 144

[101] Resendis-Antonio O, Reed JL, Encarnación S, *et al.* Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. *PLoS Comput Biol* (2007), **3**(10), 1887–1895. 33, 124

[102] Encarnación S, Dunn M, Willms K, *et al.* Fermentative and aerobic metabolism in *Rhizobium etli*. *J Bacteriol* (1995), **177**(11), 3058–3066. 33

[103] Day DA, Poole PS, Tyerman SD, *et al.* Ammonia and amino acid transport across symbiotic membranes in nitrogen-fixing legume nodules. *Cell Mol Life Sci* (2001), **58**(1), 61–71. 33

[104] Brown CM and Dilworth MJ. Ammonia Assimilation by *Rhizobium* Cultures and Bacteroids. *J Gen Microbiol* (1975), **86**(1), 39–48. 33

[105] Castillo A, Taboada H, Mendoza A, *et al.* Role of GOGAT in carbon and nitrogen partitioning in *Rhizobium etli*. *Microbiology* (2000), **146**(7), 1627–1637. 33

[106] Poole P and Allaway D. Carbon and nitrogen metabolism in *Rhizobium*. *Adv Microb Physiol* (2000), **43**, 117–163. 33

[107] Waters JK, Hughes BL, Purcell LC, *et al.* Alanine, not ammonia, is excreted from N2-fixing soybean nodule bacteroids. *Proc Natl Acad Sci U S A* (1998), **95**(20), 12038–12042. 33, 133, 139

[108] Allaway D, Lodwig EM, Crompton LA, *et al.* Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Mol Microbiol* (2000), **36**(2), 508–515. 33, 143

[109] Kumar S, Bourds A, and Poole P. *De novo* alanine synthesis by bacteroids of *Mesorhizobium loti* is not required for nitrogen transfer in the determinate nodules of *Lotus corniculatus*. *J Bacteriol* (2005), **187**(15), 5493–5495. 33, 143

[110] Prell J, White JP, Bourdes A, *et al.* Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proc Natl Acad Sci U S A* (2009), **106**(30), 12477–12482. 33, 144

[111] Prell J, Bourdès A, Kumar S, *et al.* Role of symbiotic auxotrophy in the *Rhizobium*-legume symbioses. *PLoS One* (2010), **5**(11), e13933. 33

[112] de las Nieves Peltzer M, Roques N, Poinsot V, *et al.* Auxotrophy accounts for nodulation defect of most *Sinorhizobium meliloti* mutants in the branched-chain amino acid biosynthesis pathway. *Mol Plant Microbe Interact* (2008), **21**(9), 1232–1241. 34

[113] Ge H, Walhout AJM, and Vidal M. Integrating 'omic' information: a bridge between genomics and systems biology. *Trends Genet* (2003), **19**(10), 551–560. 35

[114] Pfau T, Christian N, and Ebenhöh O. Systems approaches to modelling pathways and networks. *Brief Funct Genomics* (2011), **10**(5), 266–279. 35, 43

[115] Palsson BØ. The challenges of *in silico* biology. *Nat Biotechnol* (2000), **18**(11), 1147–1150. 35

[116] Edwards JS and Palsson BØ. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A* (2000), **97**(10), 5528–5533. 35, 46

[117] Reed JL, Vo TD, Schilling CH, *et al.* An expanded genome-scale model of *Escherichia coli* K-12 (*i*JR904 GSM/GPR). *Genome Biol* (2003), **4**(9), R54. 35, 128

[118] Feist AM, Henry CS, Reed JL, *et al.* A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* (2007), **3**, 121. 35, 123

[119] Archer CT, Kim JF, Jeong H, *et al.* The genome sequence of *E. coli* W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. *BMC Genomics* (2011), **12**, 9. 35

[120] Poolman MG, Miguet L, Sweetlove LJ, *et al.* A genome-scale metabolic model of Arabidopsis and some of its properties. *Plant Physiol* (2009), **151**(3), 1570–1581. 35, 47, 79, 85, 98, 141

[121] de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, *et al.* AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant Physiol* (2010), **152**(2), 579–589. 35, 79, 141

[122] Mintz-Oron S, Meir S, Malitsky S, *et al.* Reconstruction of *Arabidopsis* metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc Natl Acad Sci U S A* (2012), **109**(1), 339–344. 35

[123] Duarte NC, Becker SA, Jamshidi N, *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* (2007), **104**(6), 1777–1782. 35

[124] Thiele I, Swainston N, Fleming RMT, *et al.* A community-driven global reconstruction of human metabolism. *Nat Biotechnol* (2013), **31**(5), 419–425. 35

[125] Thiele I and Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protocols* (2010), **5**(1), 93–121. 35

[126] Kanehisa M and Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* (2000), **28**(1), 27–30. 36

[127] Kanehisa M, Goto S, Sato Y, *et al.* KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* (2012), **40**(Database issue), D109–D114. 36

[128] Caspi R, Foerster H, Fulcher CA, *et al.* The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* (2008), **36**(Database issue), D623–D631. 36

[129] Scheer M, Grote A, Chang A, *et al.* BRENDA, the enzyme information system in 2011. *Nucleic Acids Res* (2011), **39**(Database issue), D670–D676. 36

[130] Schomburg I, Chang A, Placzek S, *et al.* BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res* (2013), **41**(Database issue), D764–D772. 36

[131] Karp PD, Paley SM, Krummenacker M, *et al.* Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* (2010), **11**(1), 40–79. 36, 52, 72

[132] Pinney JW, Shirley MW, McConkey GA, *et al.* metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. *Nucleic Acids Res* (2005), **33**(4), 1399–1409. 36

[133] Kharchenko P, Vitkup D, and Church GM. Filling gaps in a metabolic network using expression information. *Bioinformatics* (2004), **20**(Suppl 1), i178–i185. 37

[134] Kumar VS, Dasika M, and Maranas C. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* (2007), **8**(1), 212. 37

[135] Christian N, May P, Kempa S, *et al.* An integrative approach towards completing genome-scale metabolic networks. *Mol BioSyst* (2009), **5**(12), 1889–1903. 37, 59

[136] Kruse K and Ebenhöh O. Comparing flux balance analysis to network expansion: producibility, sustainability and the scope of compounds. *Genome Inform* (2008), **20**, 91–101. 38, 79

[137] Benyamini T, Folger O, Ruppin E, *et al.* Flux balance analysis accounting for metabolite dilution. *Genome Biol* (2010), **11**(4), R43. 38, 49, 79

[138] Beard DA, Liang SD, and Qian H. Energy Balance for Analysis of Complex Metabolic Networks. *Biophys J* (2002), **83**(1), 79–86. 38

[139] Price ND, Thiele I, and Palsson BØ. Candidate states of *Helicobacter pylori*'s genome-scale metabolic network upon application of "loop law" thermodynamic constraints. *Biophys J* (2006), **90**(11), 3919–3928. 39

[140] Heazlewood JL, Verboom RE, Tonti-Filippini J, *et al.* SUBA: the Arabidopsis Subcellular Database. *Nucleic Acids Res* (2007), **35**(suppl 1), D213–D218. 40, 59

[141] Sprenger J, Fink JL, Karunaratne S, *et al.* LOCATE: a mammalian protein subcellular localization database. *Nucleic Acids Res* (2008), **36**(Database issue), D230–D233. 40

[142] Briesemeister S, Blum T, Brady S, *et al.* SherLoc2: a high-accuracy hybrid method for predicting subcellular localization of proteins. *J Proteome Res* (2009), **8**(11), 5363–5366. 40

[143] Yu NY, Wagner JR, Laird MR, *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* (2010), **26**(13), 1608–1615. 40

[144] Swainston N, Smallbone K, Mendes P, *et al.* The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks. *J Integr Bioinform* (2011), **8**(2), 186. 40

[145] Mintz-Oron S, Aharoni A, Ruppin E, *et al.* Network-based prediction of metabolic enzymes' subcellular localization. *Bioinformatics* (2009), **25**(12), i247–i1252. 40

[146] Segel L and Slemrod M. The Quasi-Steady-State Assumption: A Case Study in Perturbation. *SIAM Rev.* (1989), **31**(3), 446–477. 41

[147] Lay DC (2006). *Linear Algebra and Its Applications*. Pearson. 42

[148] Poolman MG, Sebu C, Pidcock MK, *et al.* Modular decomposition of metabolic systems via null-space analysis. *J Theor Biol* (2007), **249**(4), 691–705. 43

[149] Palsson BØ (2006). *Systems Biology - Properties of Reconstructed Networks*. Cambridge University Press. 43

[150] Schuster S and Hilgetag C. On elementary flux modes in biochemical reaction systems at steady state. *J Biol Syst* (1994), **2**(2), 165–182. 43

[151] Heinrich R and Schuster S. (1996). *The Regulation of Cellular Systems*. Chapman & Hall, London, UK. 43

[152] Pfeiffer T, Sánchez-Valdenebro I, no JCN, *et al.* METATOOL: for studying metabolic networks. *Bioinformatics* (1999), **15**(3), 251–257. 43, 44

[153] Poolman MG, Fell DA, and Raines CA. Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. *Eur J Biochem* (2003), **270**(3), 430–439. 43

[154] Trinh CT, Wlaschin A, and Srienc F. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Microbiol Biotechnol* (2009), **81**(5), 813–826. 43

[155] Terzer M and Stelling J. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics* (2008), **24**(19), 2229–2235. 43

[156] Wiback SJ and Palsson BØ. Extreme pathway analysis of human red blood cell metabolism. *Biophys J* (2002), **83**(2), 808–818. 43

[157] Price ND, Papin JA, and Palsson BØ. Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. *Genome Res* (2002), **12**(5), 760–769. Case Study. 43

[158] Klamt S and Gilles ED. Minimal cut sets in biochemical reaction networks. *Bioinformatics* (2004), **20**(2), 226–234. 43

[159] Klamt S. Generalized concept of minimal cut sets in biochemical networks. *Biosystems* (2006), **83**(2-3), 233–247. 43

[160] Klamt S and Stelling J. Combinatorial complexity of pathway analysis in metabolic networks. *Mol Biol Rep* (2002), **29**(1-2), 233–236. 44

[161] Schilling CH, Letscher D, and Palsson BØ. Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective. *J Theor Biol* (2000), **203**(3), 229–248. 44

[162] Schuster R and Schuster S. Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed. *Comput Appl Biosci* (1993), **9**(1), 79–85. 44

[163] de Figueiredo LF, Podhorski A, Rubio A, *et al.* Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics* (2009), **25**(23), 3158–3165. 44

[164] Varma A and Palsson BØ. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Nat Biotech* (1994), **12**(10), 994–998. 45

[165] Dantzig GB, Orden A, and Wolfe PS. (1954). *Notes on Linear Programming: Part I: The Generalized Simplex Method for Minimizing a Linear Form Under Linear Inequality Restraints.* Santa Monica, CA: RAND Corporation. 45

[166] Schuetz R, Kuepfer L, and Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol* (2007), **3**, 119. 46

[167] Feist AM and Palsson BØ. The biomass objective function. *Curr Opin Microbiol* (2010), **13**(3), 344–349. 46

[168] Edwards JS, Ibarra RU, and Palsson BØ. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* (2001), **19**(2), 125–130. 46

[169] Schilling CH, Covert MW, Famili I, *et al.* Genome-scale metabolic model of *Helicobacter pylori* 26695. *J Bacteriol* (2002), **184**(16), 4582–4593. 46

[170] Gianchandani EP, Oberhardt MA, Burgard AP, *et al.* Predicting biological system objectives *de novo* from internal state measurements. *BMC Bioinformatics* (2008), **9**, 43. 46

[171] Poolman MG, Kundu S, Shaw R, *et al.* Responses to Light Intensity in a Genome-Scale Model of Rice Metabolism. *Plant Physiol* (2013). 47, 79, 141

[172] Blank LM, Kuepfer L, and Sauer U. Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol* (2005), **6**(6), R49. 47

[173] Lee S, Phalakornkule C, Domach MM, *et al.* Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput Chem Eng* (2000), **24**(2-7), 711–716. 47

[174] Mahadevan R and Schilling C. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* (2003), **5**(4), 264–276. 47, 48

[175] Murabito E, Simeonidis E, Smallbone K, *et al.* Capturing the essence of a metabolic network: A flux balance analysis approach. *J Theor Biol* (2009), **260**(3), 445–452. 47

[176] Gudmundsson S and Thiele I. Computationally efficient flux variability analysis. *BMC Bioinformatics* (2010), **11**, 489. 48

[177] Burgard AP, Pharkya P, and Maranas CD. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnol Bioeng* (2003), **84**(6), 647–657. 48

[178] Tepper N and Shlomi T. Predicting Metabolic Engineering Knockout Strategies for Chemical Production: Accounting for Competing Pathways. *Bioinformatics* (2010), **26**(4), 536–543. 48

[179] Segrè D, Vitkup D, and Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* (2002), **99**(23), 15112–15117. 48

[180] Shlomi T, Berkman O, and Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A* (2005), **102**(21), 7695–7700. 48

[181] Handorf T, Ebenhöh O, and Heinrich R. Expanding Metabolic Networks: Scopes of Compounds, Robustness, and Evolution. *J Mol Evol* (2005), **61**(4), 498–512. 49

[182] Gevorgyan A, Poolman MG, and Fell DA. Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics* (2008), **24**(19), 2245–2251. 55

[183] Farag MA, Huhman DV, Lei Z, *et al.* Metabolic profiling and systematic identification of flavonoids and isoflavonoids in roots and cell suspension cultures of *Medicago truncatula* using HPLC-UV-ESI-MS and GC-MS. *Phytochemistry* (2007), **68**(3), 342 – 354. 56, 78

[184] Pollier J, Morreel K, Geelen D, *et al.* Metabolite Profiling of Triterpene Saponins in *Medicago truncatula* Hairy Roots by Liquid Chromatography Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J Nat Prod* (2011), **74**(6), 1462–1476. 56

[185] Huhman DV and Sumner LW. Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. *Phytochemistry* (2002), **59**(3), 347 – 360. 56

[186] Kowalska I, Stochmal A, Kapusta I, *et al.* Flavonoids from Barrel Medic (*Medicago truncatula*) Aerial Parts. *J Agric Food Chem* (2007), **55**(7), 2645–2652. 56

[187] Daher Z, Recorbet G, Valot B, *et al.* Proteomic analysis of *Medicago truncatula* root plastids. *Proteomics* (2010), **10**(11), 2123–2137. 58, 59

[188] Dubinin J, Braun HP, Schmitz U, *et al.* The mitochondrial proteome of the model legume *Medicago truncatula*. *BBA-Proteins Proteom* (2011), **1814**(12), 1658–1668. 58, 59

[189] Lamesch P, Berardini TZ, Li D, *et al.* The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res* (2012), **40**(D1), D1202–D1210. 59

[190] Quackenbush J, Cho J, Lee D, *et al.* The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Res* (2001), **29**(1), 159–164. 59

[191] Benveniste P. Biosynthesis and accumulation of sterols. *Annu Rev Plant Biol* (2004), **55**, 429–457. 60

[192] Reinhart MP. Intracellular sterol trafficking. *Experientia* (1990), **46**(6), 599–611. 60

[193] Linka N and Weber AP. Intracellular Metabolite Transporters in Plants. *Molecular Plant* (2010), **3**(1), 21–53. 61, 164, 165, 166, 167, 170

[194] Helliwell CA, Sullivan JA, Mould RM, *et al.* A plastid envelope location of *Arabidopsis ent*-kaurene oxidase links the plastid and endoplasmic reticulum steps of the gibberellin biosynthesis pathway. *Plant J* (2001), **28**(2), 201–208. 61

[195] León J and Sánchez-Serrano JJ. Molecular biology of jasmonic acid biosynthesis in plants. *Plant Physiol Bioch* (1999), **37**(5), 373–380. 61

[196] Babujee L, Wurtz V, Ma C, *et al.* The proteome map of spinach leaf peroxisomes indicates partial compartmentalization of phylloquinone (vitamin K1) biosynthesis in plant peroxisomes. *J Exp Bot* (2010), **61**(5), 1441–1453. 61

[197] Flamholz A, Noor E, Bar-Even A, *et al.* eQuilibrator–the biochemical thermodynamics calculator. *Nucleic Acids Res* (2012), **40**(Database issue), D770–D775. 62, 130

[198] Shlomi T, Cabili MN, Herrgård MJ, *et al.* Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* (2008), **26**(9), 1003–1010. 66, 121, 122

[199] Lisec J, Schauer N, Kopka J, *et al.* Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc* (2006), **1**(1), 387–396. 67, 69

[200] Hara A and Radin NS. Lipid extraction of tissues with a low-toxicity solvent. *Anal Biochem* (1978), **90**(1), 420–426. 68

[201] Masakapalli SK (2011). *Network flux analysis of central metabolism in plants.* PhD thesis, University of Oxford. 68, 69, 70

[202] Sambrook J and Russel D. (2000). *Molecular Cloning - A Laboratory Manual.* Cold Spring Harbour Laboratory Press. 69

[203] Masakapalli SK, Kruger NJ, and Ratcliffe RG. The metabolic flux phenotype of heterotrophic Arabidopsis cells reveals a complex response to changes in nitrogen supply. *Plant J* (2013), **74**(4), 569–582. 69, 141, 142, 143

[204] Antoniewicz MR, Kelleher JK, and Stephanopoulos G. Accurate Assessment of Amino Acid Mass Isotopomer Distributions for Metabolic Flux Analysis. *Anal Chem* (2007), **79**(19), 7554–7559. 69

[205] González-Segovia R, Quintanar JL, Salinas E, *et al.* Effect of the flavonoid quercetin on inflammation and lipid peroxidation induced by *Helicobacter pylori* in gastric mucosa of guinea pig. *J Gastroenterol* (2008), **43**(6), 441–447. 78, 146

[206] Latunde-Dada AO, Cabello-Hurtado F, Czittrich N, *et al.* Flavonoid 6-hydroxylase from soybean (*Glycine max* L.), a novel plant P-450 monooxygenase. *J Biol Chem* (2001), **276**(3), 1688–1695. 78

[207] Winkel-Shirley B. Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. *Plant Physiol* (2001), **126**(2), 485–493. 78

[208] Winkel-Shirley B. Biosynthesis of flavonoids and effects of stress. *Curr Opin Plant Biol* (2002), **5**(3), 218–223. 78

[209] Rolland N, Curien G, Finazzi G, *et al.* The Biosynthetic Capacities of the Plastids and Integration Between Cytoplasmic and Chloroplast Processes. *Annu Rev Genet* (2012), **46**, 233–264. 79

[210] Lötscher M, Klumpp K, and Schnyder H. Growth and maintenance respiration for individual plants in hierarchically structured canopies of *Medicago sativa* and *Helianthus annuus*: the contribution of current and old assimilates. *New Phytol* (2004), **164**(2), 305–316. 81, 126

[211] Yousfi N, Slama I, Ghnaya T, *et al.* Effects of water deficit stress on growth, water relations and osmolyte accumulation in *Medicago truncatula* and *M. laciniata* populations. *C R Biol* (2010), **333**(3), 205–213. 85

[212] Bakoglu A, Bagci E, Kocak A, *et al.* Fatty Acid Composition of Some *Medicago* L. (Fabaceae) Species From Turkey. *Asian J Chem* (2010), **22**(1), 651–656. 85, 86

[213] Jung H and Engels F. Alfalfa Stem Tissues: Cell Wall Deposition, Composition, and Degradability. *Crop science.* (2002), **42 (2)**, 524–534. 85, 86

[214] Nakashima J, Chen F, Jackson L, *et al.* Multi-site genetic modification of monolignol biosynthesis in alfalfa (*Medicago sativa*): effects on lignin composition in specific cell types. *New Phytol* (2008), **179**(3), 738–750. 85, 86

[215] Johnson JMF, Barbour NW, and Weyers SL. Chemical Composition of Crop Biomass Impacts Its Decomposition. *Soil Sci Soc Am J* (2007), **71**(1), 155–162. 85, 86

[216] Blondon F, Marie D, Brown S, *et al.* Genome size and base composition in *Medicago sativa* and *M. truncatula* species. *Genome* (1994), **37**(2), 264–270. 85

[217] De Vries FWTP. The Cost of Maintenance Processes in Plant Cells. *Ann Bot* (1975), **39**(1), 77–92. 90, 97, 98

[218] Ros R, Cascales-Miñana B, Segura J, *et al.* Serine biosynthesis by photorespiratory and non-photorespiratory pathways: an interesting interplay with unknown regulatory networks. *Plant Biol (Stuttg)* (2012). 96

[219] Stephanopoulos G, Aristidou AA, and Nielsen J. (1998). *Maetabolic Engineering: Principles and Methodologies*, chapter 2, page 74. Academic Press. 98

[220] Graf A, Schlereth A, Stitt M, *et al.* Circadian control of carbohydrate availability for growth in *Arabidopsis* plants at night. *Proc Natl Acad Sci U S A* (2010), **107**(20), 9458–9463. 99

[221] Schubert S and Yan F. Nitrate and ammonium nutrition of plants: Effects on acid/base balance and adaptation of root cell plasmalemma H+ ATPase. *Z. Pflanzenernähr. Bodenk.* (1997), **160**(2), 275–281. 106

[222] Ohmori M, Ohmori K, and Strotmann H. Inhibition of nitrate uptake by ammonia in a blue-green alga, *Anabaena cylindrica* (1977), **114**(3), 225–229–. 111

[223] Alam SM, Naqvi SS, and Ansari R. (2007). *Handbook of Plant and Crop Stress - Impact of Soil pH on Nutrient Uptake by Crop Plants*, chapter 3, pages 51–60. Dekker, 2nd edition. 112

[224] de la Haba P, Agüera E, and Maldonado JM. Differential effects of ammonium and tungsten on nitrate and nitrite uptake and reduction by sunflower plants. *Plant Science* (1990), **70**(1), 21–26. 118

[225] Schortemeyer M, Feil B, and Stamp P. Root Morphology and Nitrogen Uptake of Maize Simultaneously Supplied with Ammonium and Nitrate in a Split-root System. *Ann Bot* (1993), **72**(2), 107–115. 118

[226] McRae DG, Miller RW, Berndt WB, *et al.* Transport of C(4)-Dicarboxylates and Amino Acids by *Rhizobium meliloti* Bacteroids. *Mol Plant-Microbe Interact* (1989), **2**(5), 273–278. 124, 125

[227] Bergersen F. Physiological and biochemical aspects of nitrogen fixation by bacteroids in soybean nodule cells. *Soil Biol Biochem* (1997), **29**(56), 875–880. 124, 125, 132

[228] Vance CP, Heichel GH, Barnes DK, *et al.* Nitrogen Fixation, Nodule Development, and Vegetative Regrowth of Alfalfa (*Medicago sativa* L.) following Harvest. *Plant Physiol* (1979), **64**(1), 1–8. 124

[229] Aydi S, Drevon JJ, and Abdelly C. Effect of salinity on root-nodule conductance to the oxygen diffusion in the *Medicago truncatula-Sinorhizobium meliloti* symbiosis. *Plant Physiol Biochem* (2004), **42**(10), 833–840. 124

[230] Sutton WD, Jepsen NM, and Shaw BD. Changes in the Number, Viability, and Amino-acid-incorporating Activity of *Rhizobium* Bacteroids during Lupin Nodule Development. *Plant Physiol* (1977), **59**(4), 741–744. 124

[231] Terpolilli JJ, O'Hara GW, Tiwari RP, *et al.* The model legume *Medicago truncatula* A17 is poorly matched for N2 fixation with the sequenced microsymbiont *Sinorhizobium meliloti* 1021. *New Phytol* (2008), **179**(1), 62–66. 126

[232] Russell JB and Cook GM. Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol Rev* (1995), **59**(1), 48–62. 128

[233] Dunn MF. Tricarboxylic acid cycle and anaplerotic enzymes in rhizobia. *FEMS Microbiol Rev* (1998), **22**(2), 105–123. 129

[234] Das D and Veziroğlu T. Hydrogen production by biological processes: a survey of literature. *Int J Hydrogen Energy* (2001), **26**(1), 13–28. 139

[235] Merberg D, O'Hara EB, and Maier RJ. Regulation of hydrogenase in *Rhizobium japonicum*: analysis of mutants altered in regulation by carbon substrates and oxygen. *J Bacteriol* (1983), **156**(3), 1236–1242. 139

[236] Albrecht SL, Maier RJ, Hanus FJ, *et al.* Hydrogenase in Rhizobium japonicum Increases Nitrogen Fixation by Nodulated Soybeans. *Science* (1979), **203**(4386), 1255–1257. 140

[237] Schubert KR, Jennings NT, and Evans HJ. Hydrogen Reactions of Nodulated Leguminous Plants: II. Effects on Dry Matter Accumulation and Nitrogen Fixation. *Plant Physiol* (1978), **61**(3), 398–401. 140

[238] Ciccolella CO, Raynard NA, Mei JHM, *et al.* Symbiotic Legume Nodules Employ Both Rhizobial *Exo-* and *Endo*-Hydrogenases to Recycle Hydrogen Produced by Nitrogen Fixation. *PLoS One* (2010), **5**(8), e12094. 140

[239] Horchani F, Hajri R, and Aschi-Smiti S. Is the sensitivity to ammonium nutrition related to nitrogen accumulation? *Current Botany* (2011), **2**(2). 142, 143

[240] Pearson J and Stewart GR. The deposition of atmospheric ammonia and its effects on plants. *New Phytol* (1993), **125**(2), 283–305. 142

[241] Britto DT, Siddiqi MY, Glass AD, *et al.* Futile transmembrane NH4(+) cycling: A cellular hypothesis to explain ammonium toxicity in plants. *Proc Natl Acad Sci U S A* (2001), **98**(7), 4255–4258. 142

[242] Dunn MF, Encarnación S, Araíza G, *et al.* Pyruvate carboxylase from *Rhizobium etli*: mutant characterization, nucleotide sequence, and physiological role. *J Bacteriol* (1996), **178**(20), 5960–5970. 144

[243] Cheung CYM, Williams TCR, Poolman MG, *et al.* A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. *Plant J* (2013). 144, 145

[244] Cooper GM (2000). *The Cell - A Molecular Approach 2nd Edition.* Sunderland (MA): Sinauer Associates. 164, 165, 166, 167, 170

[245] Schubert KR. Products of Biological Nitrogen Fixation in Higher Plants: Synthesis, Transport, and Metabolism. *Annu. Rev. Plant. Physiol.* (1986), **37**(1), 539–574. 165, 166, 167

[246] Hänsch R, Lang C, Riebeseel E, *et al.* Plant Sulfite Oxidase as Novel Producer of H_2O_2: *COMBINATION OF ENZYME CATALYSIS WITH A SUBSEQUENT NON-ENZYMATIC REACTION STEP. J Biol Chem* (2006), **281**(10), 6884–6888. 166

[247] Kaur N, Reumann S, and Hu J. Peroxisome Biogenesis and Function. *The Arabidopsis Book* (2009), pages e0123–. 166

[248] Baker A, Graham IA, Holdsworth M, *et al.* Chewing the fat: β-oxidation in signalling and development. *Trends Plant Sci* (2006), **11**(3), 124–132. 166, 167

[249] Jander G and Joshi V. Aspartate-Derived Amino Acid Biosynthesis in *Arabidopsis thaliana*. *The Arabidopsis Book* (2009), pages e0121–. 168

[250] Ingle RA. Histidine biosynthesis. *The Arabidopsis Book* (2011), **9**, e0141. 168

[251] Binder S. Branched-Chain Amino Acid Metabolism in *Arabidopsis thaliana*. *The Arabidopsis Book* (2010), pages e0137–. 168

[252] Hudson AO, Bless C, Macedo P, *et al.* Biosynthesis of lysine in plants: evidence for a variant of the known bacterial pathways. *BBA-Gen Subjects* (2005), **1721**(13), 27–36. 168

[253] Maeda H, Shasany AK, Schnepp J, *et al.* RNAi Suppression of Arogenate Dehydratase1 Reveals That Phenylalanine Is Synthesized Predominantly via the Arogenate Pathway in Petunia Petals. *The Plant Cell Online* (2010), **22**(3), 832–849. 168

[254] Ho CL and Saito K. Molecular biology of the plastidic phosphorylated serine biosynthetic pathway in *Arabidopsis thaliana*. *Amino Acids* (2001), **20**(3), 243–259. 168

[255] Ravanel S, Block MA, Rippert P, *et al.* Methionine Metabolism in Plants. *J Biol Chem* (2004), **279**(21), 22548–22557. 168, 170

[256] Hesse H, Nikiforova V, Gakière B, *et al.* Molecular analysis and control of cysteine biosynthesis: integration of nitrogen and sulphur metabolism. *J Exp Bot* (2004), **55**(401), 1283–1292. 168

[257] Ravanel S, Cherest H, Jabrin S, *et al.* Tetrahydrofolate biosynthesis in plants: Molecular and functional characterization of dihydrofolate synthetase and three isoforms of folylpolyglutamate synthetase in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* (2001), **98**(26), 15360–15365. 170

[258] Christensen KE and MacKenzie RE. Mitochondrial one-carbon metabolism is adapted to the specific needs of yeast, plants and mammals. *BioEssays* (2006), **28**(6), 595–605. 170

[259] Quinlivan EP, Roje S, Basset G, *et al.* The Folate Precursor *p*-Aminobenzoate Is Reversibly Converted to Its Glucose Ester in the Plant Cytosol. *J Biol Chem* (2003), **278**(23), 20731–20737. 170

[260] Li-Beisson Y, Shorrosh B, Beisson F, *et al.* Acyl-Lipid Metabolism. *The Arabidopsis Book* (2010), pages e0133–. 170

[261] Coxon KM, Chakauya E, Ottenhof HH, *et al.* Pantothenate biosynthesis in higher plants. *Biochem Soc Trans* (2005), **33**(Pt 4), 743–746. 170

[262] MacKenzie S and McIntosh L. Higher Plant Mitochondria. *Plant Cell* (1999), **11**(4), 571–585. 170