Modeling photosynthesis and related metabolic processes
– from detailed examination to consideration of the metabolic context –

Dissertation

zur Erlangung des akademischen Grades
"doctor rerum naturalium" (Dr. rer. nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

eingereicht von

ANNE ARNOLD

Arbeitsgruppe Systembiologie und mathematische Modellierung
Max-Planck-Institut für molekulare Pflanzenphysiologie

Potsdam im März 2014

Abstract

Mathematical modeling of biological systems is a powerful tool to systematically investigate the functions of biological processes and their relationship with the environment. To obtain accurate and biologically interpretable predictions, a modeling framework has to be devised whose assumptions best approximate the examined scenario and which copes with the trade-off of complexity of the underlying mathematical description: with attention to detail or high coverage. Correspondingly, the system can be examined in detail on a smaller scale or in a simplified manner on a larger scale. In this thesis, the role of photosynthesis and its related biochemical processes in the context of plant metabolism was dissected by employing modeling approaches ranging from kinetic to stoichiometric models.

The Calvin-Benson cycle, as primary pathway of carbon fixation in C_3 plants, is the initial step for producing starch and sucrose, necessary for plant growth. Based on an integrative analysis for model ranking applied on the largest compendium of (kinetic) models for the Calvin-Benson cycle, those suitable for development of metabolic engineering strategies were identified.

Driven by the question why starch rather than sucrose is the predominant transitory carbon storage in higher plants, the metabolic costs for their synthesis were examined. The incorporation of the maintenance costs for the involved enzymes provided a model-based support for the preference of starch as transitory carbon storage, by only exploiting the stoichiometry of synthesis pathways.

Many photosynthetic organisms have to cope with processes which compete with carbon fixation, such as photorespiration whose impact on plant metabolism is still controversial. A systematic model-oriented review provided a detailed assessment for the role of this pathway in inhibiting the rate of carbon fixation, bridging carbon and nitrogen metabolism, shaping the C_1 metabolism, and influencing redox signal transduction.

The demand of understanding photosynthesis in its metabolic context calls for the examination of the related processes of the primary carbon metabolism. To this end, the *Arabidopsis core model* was assembled via a bottom-up approach. This large-scale model can be used to simulate photoautotrophic biomass production, as an indicator for plant growth, under so-called optimal, carbon-limiting and nitrogen-limiting growth conditions.

Finally, the introduced model was employed to investigate the effects of the environment, in particular, nitrogen, carbon and energy sources, on the metabolic behavior. This resulted in a purely stoichiometry-based explanation for the experimental evidence for preferred simultaneous acquisition of nitrogen in both forms, as nitrate and ammonium, for optimal growth in various plant species.

The findings presented in this thesis provide new insights into plant system's behavior, further support existing opinions for which mounting experimental evidences arise, and posit novel hypotheses for further directed large-scale experiments.

iv

Zusammenfassung

Mathematische Modellierung biologischer Systeme eröffnet die Möglichkeit systematisch die Funktionsweise biologischer Prozesse und ihrer Wechselwirkungen mit der Umgebung zu untersuchen. Um präzise und biologisch relevante Vorhersagen treffen zu können, muss eine Modellierungsstrategie konzipiert werden, deren Annahmen das untersuchte Szenario bestmöglichst widerspiegelt und die dem *Trade-off* der Komplexität der zugrunde liegenden mathematischen Beschreibung gerecht wird: Detailtreue gegenüber Größe. Dementsprechend kann das System detailliert, in kleinerem Umfang oder in vereinfachter Darstellung im größeren Maßstab untersucht werden. In dieser Arbeit wird mittels verschiedener Modellierungsansätze, wie kinetischen und stöchiometrischen Modellen, die Rolle der Photosynthese und damit zusammenhängender biochemischer Prozesse im Rahmen des Pflanzenstoffwechsels analysiert.

Der Calvin-Benson-Zyklus, als primärer Stoffwechselweg der Kohlenstofffixierung in C_3-Pflanzen, ist der erste Schritt der Stärke- und Saccharoseproduktion, welche maßgeblich für das Wachstum von Pflanzen sind. Basierend auf einer integrativen Analyse zur Modellklassifizierung wurden aus der größten bekannten Sammlung von (kinetischen) Modellen des Calvin-Benson-Zyklus diejenigen ermittelt, die für die Entwicklung von Metabolic-Engineering-Strategien geeignet sind.

Angeregt von der Fragestellung warum Kohlenstoff transitorisch vorwiegend in Form von Stärke anstatt Saccharose gespeichert wird, wurden die metabolischen Kosten beider Syntheseprozesse genauer betrachtet. Die Einbeziehung der Bereitstellungskosten der beteiligten Enzyme stützt die Tatsache, dass bevorzugt Stärke als temporärer Kohlenstoffspeicher dient. Die entprechende Untersuchung erfolgte einzig auf Grundlage der Stöchiometrie der Synthesewege.

In vielen photosynthetisch-aktiven Organismen findet zudem Photorespiration statt, die der Kohlenstofffixierung entgegenwirkt. Die genaue Bedeutung der Photorespiration für den Pflanzenmetabolismus ist noch umstritten. Eine detaillierte Einschätzung der Rolle dieses Stoffwechselweges bezüglich der Inhibierung der Kohlenstofffixierungsrate, der Verknüpfung von Kohlenstoff- und Stickstoffmetabolismus, der Ausprägung des C_1-Stoffwechsels sowie die Einflussnahme auf die Signaltransduktion wurde in einer modell-basierten, kritischen Analyse vorgenommen.

Um die Photosynthese in ihrem metabolischen Kontext verstehen zu können, ist die Betrachtung der angrenzenden Prozesse des primären Kohlenstoffmetabolismus unverzichtbar. Hierzu wurde in einem Bottom-up Ansatz das *Arabidopsis core* Modell entworfen, mittels dessen die Biomasseproduktion, als Indikator für Pflanzenwachtum, unter photoautotrophen Bedingungen simuliert werden kann. Neben sogenannten optimalen Wachstumsbedingungen kann dieses großangelegte Modell auch kohlenstoff- und stickstofflimitierende Umweltbedingungen simulieren.

Abschließend wurde das vorgestellte Modell zur Untersuchung von Umwelteinflüssen auf

das Stoffwechselverhalten herangezogen, im speziellen verschiedene Stickstoff-, Kohlenstoff- und Energiequellen. Diese auschließlich auf der Stöchiometrie basierende Analyse bietet eine Erklärung für die bevorzugte, gleichzeitige Aufnahme von Nitrat und Ammonium, wie sie in verschiedenen Spezies für optimales Wachstum experimentell beobachtet wurde.

Die Resultate dieser Arbeit liefern neue Einsichten in das Verhalten von pflanzlichen Systemen, stützen existierende Ansichten, für die zunehmend experimentelle Hinweise vorhanden sind, und postulieren neue Hypothesen für weiterführende großangelegte Experimente.

Acknowledgment

I would like to deeply thank my supervisor, Zoran Nikoloski, for putting his confidence in me and encouraging me in all possible ways during my PhD. His unconditional support, the valuable advice, and all the inspiring discussions are the basis for this thesis.

Furthermore, I would like to thank Alisdair Fernie and Mark Stitt for many fruitful discussions, expert advice and insights in plant metabolism.

Of course, I like to thank Tabea Mettler for answering my questions regarding diverse biological phenomena.

I especially thank Marco Ende, Axel Nagel and Jost Neigenfind for keeping the computers running and technical support, whenever needed.

Many thanks to all present and former members of the Systems Biology and Mathematical Modeling Group for the nice working atmosphere: Heike Aßmus, Georg Basler, David Breuer, Nils Christian, Robert Heise, Florian Hollandt, Önder Katal, Sabrina Kleeßen, Jeanne Marie Mbebi, Nooshin Omranian, Semidán Robaina Estévez, Moritz Schütte, Alexander Skupin, Katrin Tirok, and Nadine Töpfer.

Special thanks go to Max Sajitz-Hermstein for his support and patience in a sometimes pretty uneasy relationship.

Moreover, I sincerely thank my family for the mental support, and their never-ending encouragement, patience and *love.*

Finally, Sergio Grimbs deserves my deepest gratitude for his scientific advice and constant discussions, for having the faith in me and this work, for backing me during hard times, and always being there for me.

Contents

Contents

xii

Abbreviations

2PG	2-phosphoglycolate
A-CoA	Acetyl-CoA
ACP	aceyl carrier protein
AdenylP	adenylphosphate
ADP	adenosine diphosphate
ADPG	adenosine diphosphate-glucose
AGPase	ADP-glucose pyrophosphorylase (EC 2.7.7.27)
Ala	alanine
amDHP	aminomethyldihydrolipoylprotein
AMP	adenosine monophosphate
Arabidopsis	*Arabidopsis thaliana*
Arg	arginine
Asn	asparagine
Asp	aspartate
ATP	adenosine triphosphate
ATPase	ATP synthase = ATPS
ATPS	ATP synthase
BRENDA	Braunschweig Enzyme Database
CBC	Calvin-Benson cycle
cDNA	complementary DNA
Chl	chlorophyll
Chlamydomonas	*Chlamydomonas reinhardtii*
Cit	citrate
CO_2	carbon dioxide
CoA	coenzyme A
CTP	cytidine triphosphate
Cys	cysteine
cyt	cytosolic
Cyt*c*	cytochrome *c*
D. carota	*Daucus carota*
dATP	deoxyadenosine triphosphate
DC-AMP	dicarboxyethyl-AMP = adenylosuccinate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DHAP	dihydroxyacetone phosphate
DHP	dihydrolipoylamide
DNA	deoxyribonucleic acid
DPGA	1,3-bisphosphoglycerate
dTTP	deoxythymidine triphosphate
DW	dry weight
DWC	dry weight content
E. coli	*Escherichia coli*
E4P	erythrose-4-phosphate
EC	Enzyme Commission
F26BP	fructose-2,6-bisphosphate
F6P	fructose-6-phosphate
FBA	flux balance analysis
FBP	fructose-1,6-bisphosphate
FBPA	aldolase interconverting FBP (EC 4.1.2.13)
FBPase	fructose-1,6-bisphosphatase (EC 3.1.3.11)
FCA	flux coupling analysis
Fd	ferredoxin
FNR	Fd-NADP reductase (EC 1.18.1.2)
Fru	fructose
Fum	fumarate

Contents

NH$_3$	ammonia
NH$_4$	ammonium
NiR	nitrite reductase (EC 1.7.7.1)
NO$_2$	nitrite
NO$_3$	nitrate
NR	nitrate reductase (EC 1.7.1.1)
NTP	nucleotide triphosphate
O$_2$	oxygen
OAA	oxalacetate
ODE	ordinary differential equation
Orn	ornithine
ox	oxidized
OxalAc	oxalacetate = OAA
P. aeruginosa	*Pseudomonas aeruginosa*
P. falciparum	*Plasmodium falciparum*
P. sativum	*Pisum sativum*
P. vulgaris	*Proteus vulgaris*
P5C	1-pyrroline-5-carboxylate
PAL	phenylalanine ammonia-lyase (EC 4.3.1.249
PEP	phosphoenolpyruvate
PF2K	6-phosphofructo-2-kinase (EC 2.7.1.105)
PFD	photon flux density
PFK	phosphofrucokinase (EC 2.7.1.11)
PFP	pyrophosphate-dependent phosphofructokinase
PGA	3-phosphoglycerate
PGAK	3-phosphoglycerate kinase (EC 2.7.2.3)
PGI	glucose-6-phosphate isomerase (EC 5.3.1.9)
PGM	phosphoglucomutase (EC 5.4.2.2)
PGPase	phosphoglycolate phosphatase (EC 3.1.3.18)
Phe	phenylalanine
Pi	inorganic phosphate
PPDB	Plant Proteome Database
PPi	pyrophosphate
PRI	ribose-5-phosphate isomerase (EC 5.3.1.6)
PRK	phosphoribulokinase (EC 2.7.1.19)
PRPP	phosphoribosyl pyrophosphate synthesis
Pro	proline
Pyr	pyruvate
Q	ubiquinone
QH$_2$	ubiquinol
R5P	ribose-5-phosphate
red	reduced
RED	reduction
REG	regeneration
rev	reversible
RNA	ribonucleic acid
RPE	ribulose-5-phosphate epimerase (EC 5.1.3.1)
RSS	residual sum of squares
Ru5P	ribulose-5-phosphate
RuBisCO	ribulose-1,5-bisphosphate carboxylase/oxygenase (EC 4.1.1.39)
RuBP	ribulose-1,5-bisphosphate
rxn	reaction
S. cerevisiae	*Saccharomyces cerevisiae*
S. chungbukensis	*Spingomonas chungbukensis*
S. oleracea	*Spinacia oleracea*
S. tuberosum	*Solanum tuberosum*
S7P	sedoheptulose-7-phosphate
SA	shikimate
SBML	Systems Biology Markup Language
SBP	sedoheptulose-1,7-bisphosphate
SBPA	aldolase interconverting SBP (EC 4.1.2.13)
SBPase	sedoheptulose-1,7-bisphosphatase (EC 3.1.3.37)
SCA	succinate
SeA	semialdehyde

Contents

Chapter 1.

Introduction

1.1. The merits and intricacies of investigating plants

Plants have a key role for life on earth. They are one of the major primary producers of organic material and oxygen (O_2), required by most other organisms for viability. While herbivores use diverse parts of plant for sustenance, *e.g.*, leaves (koalas), wood (termites), seeds (birds), and pollen (bees), human beings make use of them beyond nutrition. We cultivate plants to harvest additional specific products as fuels and for other economic purposes, such as: clothes (cotton), medicine (codeine, morphine) and drugs (nicotine, caffeine). Therefore, crop plants are the key element in agriculture and industry such that they are increasingly in the focus of research, with two principle questions: "How to improve crop yield?" [1] and "How does crop plants adapt to changing climatic conditions?" [2].

Crop plants are usually bred to increase yield or improve their fitness. These undertakings, in combination with the demand for enhancing the production of particular compounds (*e.g.*, lipids), often result in altered chemical composition. To address the aforementioned questions, it is important to understand plant responses on the metabolic level.

Metabolic processes subsume the entirety of biochemical processes occurring within a biological system (*e.g.*, cell or organism) that support its viability. The biochemical reactions converting simple substances to more complex ones form the anabolism, while the catabolism captures the processes of breaking down metabolites to yield energy. The most remarkable metabolic feature of plants is their ability to grow under photoautotrophic conditions, whereby they use energy from photons and convert it into cellular energy, enabling *photosynthesis*.

Outside the kingdom of plants, there are only six other groups where photosynthetic organisms are found: algae, amongst the eukaryotes, as well as cyanobacteria, purple bacteria, green sulfur bacteria, green nonsulfur bacteria and heliobacteria, amongst the prokaryotes. The photosynthetic pathways differ amongst these groups in various aspects, *e.g.*, underlying biochemistry, anatomy and/or overall morphology [3]. The main categorization into oxygenic and anoxygenic photosynthesis is based on the utilized electron donor, namely, water (H_2O) [4], sulfide [5] or nitrite [6]. Eukaryotes and cyanobacteria perform the oxygenic form, whereby O_2 is produced in the photosynthetic reactions by oxidizing H_2O (see section 1.4.1).

Based on the ability to utilize light energy, plants are able to synthesize all organic compounds starting from low-energetic precursors, while heterotrophic organisms require high-energetic precursors to drive the metabolism. In other words, plants can utilize distinct sources for energy and the necessary chemical elements, such as carbon dioxide (CO_2), H_2O, nitrate and/or ammonium, sulfate or hydrogen sulfide, and inorganic phosphate,

rather than to take up a single molecule combining energy and nutrients. This facilitates the identification and investigation of the individual steps of the (plant) anabolic processes.

The major stages of anabolism are threefold: (1) synthesis of the metabolites serving as building blocks, *i.e.*, monosaccharides, amino acids, and nucleotides, (2) the (re)activation using cellular energy from compounds, such as adenosine triphosphate, and (3) the assembly of these building blocks to macromolecules composing the biomass of each cell. The key metabolites of the plant anabolism include carbohydrates, such as sugars, starch and cellulose, free amino acids and proteins, nucleic acids, including DNA and RNA, and the diverse biochemical compound group of lipids. In addition, plants synthesize more complex metabolites, such as vitamins and the so-called secondary metabolites, for feeding deterrence, pathogen defense, stress-protection, attraction of pollinators and seed-dispersing animals, and as agents of plant-plant competition [7]. Nevertheless, already the synthesis pathways of these building blocks are numerous, intricate and interrelated. For instance, amino acid synthesis bridges carbon and nitrogen metabolism, and, in the case of cysteine and methionine, the synthesis pathways additionally link to sulfur metabolism. Nucleotide synthesis, moreover, interrelates to phosphorus metabolism such that considering solely the building block synthesis pathways results in a highly complex network. To elucidate the complexity of the metabolism, it is important to understand the underlying molecular mechanisms and their interrelationship [8]. This reveals insights into the emerging principles of metabolic behavior.

The metabolic state of a cell is determined by the interaction of metabolism with the remaining system's levels, including genome, transcriptome, and proteome. For instance, the metabolic behavior of the (active) biochemical processes is in part affected by the activity of the underlying enzymes. These, in turn, can be influenced via preceding transcriptional control and RNA processing, as well as post-translational modification and/or allosteric modulation [9].

Aside from the modulation by the different cellular levels, metabolism is also shaped by the environment, namely, abiotic and biotic factors. Unfavorable conditions, termed stresses in the extreme cases, can affect metabolism either directly, via limitations on the availability of key nutrients, or indirectly, due to its internal modulation by the other cellular processes [10, 11]. Following this line of argumentation, it is evident that plant phenotype (including the state of metabolism) emerges from system-wide effects of genetic and/or environmental perturbations [12, 13].

1.2. Approaches for system's investigation

The classical approach to address the questions raised bases on experimental *in vivo* or *in vitro* studies only, whereas the emerging approaches often involve the *in silico* guided examination of systemic effects of perturbations. Pursuing the classical approach for testing hypotheses in well-investigated model organisms is the standard in modern biology. For instance, this approach was employed to determine the combination of genetic modifications which permits the production of potato starch with very high amylose content [14]. In this particular case, the structure and synthesis pathways as well as the impact in industrial applications are well-documented [15]. These provide the necessary knowledge which can be used as a starting point for a feasible number of targeted experiments to identify the

most promising candidate modifications.

This classical approach is not suitable for dissecting processes which have not been systematically investigated, such as metabolic adaptation to environmental changes or specialized (secondary) metabolism. For instance, identifying the genes pertaining to synthesis pathway of vitamin E in tomato proved to be difficult due to the lack of detailed background knowledge [16]. Although the biosynthesis pathways of the vitamin E compounds, tocopherol and tocotrienol, were elucidated in 1979 [17], the involved key enzyme encoding genes have been identified only in the mid 1990s, but only for *Arabidopsis thaliana* (Arabidopsis) and *Synechocystis* sp. PCC6803 [18]. Thus, pursuing the solely experimental reverse genetics approach and testing systematically the impact of all genes in tomato would be time consuming and potentially uncertain since all gene knockouts would have to be created, established, and analyzed. In contrast, to narrow down candidate genes putatively affecting tocopherol content Almeida and coworkers [19] integrated high pressure liquid chromatography measurements and *in silico* methods, more precisely, BLAST-based analyses against Arabidopsis sequences and quantitative trait loci analysis for characterizing and mapping the genes involved in the biosynthesis pathway. These tocopherol-related pathway candidate genes provide, then, the starting point for further reverse genetic analyses to verify the genes of the vitamin E metabolism.

The main difference of these two approaches is their understanding of the metabolic system as a 'reductionist' and 'integrationist', respectively [20]. Methodological reduction is described by Ayala [21] as "the claim that the best strategy of research is to study living phenomena at increasingly lower levels of complexity and, ultimately, at the level of atoms and molecules" and, therefore, accounts for the local effects of the system (*e.g.*, metabolism). On the other hand, according to Mishler and Brandon [22], integration refers to the "active interaction among parts of an entity. In other words, does the presence or activity of one part of an entity matter to another part?" Hence, by pursuing the concept of integration the global, system-wide effects can be examined. By only considering the local effects, the inferred mechanisms might be confounding due, *e.g.*, misinterpreting the experimental data. In this manner, false-negative and false-positive implications might arise from measuring no metabolic signal (although unexamined metabolites may show effects) and metabolic changes which may be actually caused by other biochemical reactions (due to complexity and intricate interrelationship of metabolic processes). In contrast, considering the system-wide impact facilitates the consideration of the interdependence of subsystem effects of interpretation, however, at the cost of not revealing all molecular mechanisms.

Benefiting by both research strategies leads to the conceptual approach of *systems biology* combining "the identification and detailed characterization of the parts, with the investigation of their interaction with each other and with their wider environment, to elucidate the maintenance of the entity" [20]. The claim of understanding the complexity of the biological system propelled the various omics technologies. Therein, the respective system level is systematically examined through high-throughput experiments resulting in a surge of data. Analyzing and interpreting these data and the underlying mechanisms insistently calls for the integration by *in silico* investigations.

1.3. *In silico* investigations

In *in silico* approaches, based on continuous and discrete data, both the deterministic and stochastic processes are examined by means of *mathematical models* of the underlying system, *e.g.*, by mechanistic kinetic and stoichiometric models. Often it is assumed that "the observed dynamics are driven exclusively by internal deterministic mechanisms" [23]. However, for instance pharmacological processes are exposed to influences that are not completely understood or not feasible to model explicitly. In order to model stochastic processes, a system of stochastic differential equations, as a natural extension to the deterministic, is used [24].

Focusing on dynamic, continuous modeling of metabolism, the corresponding processes are typically described by ordinary differential equations. While the "scope of kinetic models is to quantify the rates of the [bio]chemical reaction as a function of system [...] properties", *e.g.*, metabolite concentration and kinetic parameters [25], stoichiometric models aim at determining a feasible distribution of reaction rates (fluxes) for each reaction, termed flux distribution, without using any knowledge of the underlying kinetic form.

Kinetic modeling, the conventional approach to simulate metabolic processes, describes the changes of metabolite concentrations over time and, thereby, enables to answer questions about detailed scenarios, their transient behavior, and/or effects of potential regulatory events. The various models of reaction kinetics can be categorized regarding their levels of reaction modeling, for instance, law of mass action and equilibrium approximation are based on substrate-product level, whereas Michaelis-Menten and Hill kinetics rely on substrate-enzyme level. By considering substrate binding to the enzyme, the mathematical formulation of reaction rate becomes more complex. However, this allows the incorporation of regulatory processes. In general, kinetic modeling is based on detailed background knowledge about the dynamic behavior of the system, *e.g.*, enzyme action, substrate binding affinities in terms of kinetic parameters (k in mass action and K_m in Michaelis-Menten), and regulatory mechanisms. Additionally, it acts on the assumption that the system behaves similarly *in vivo* and *in vitro* such that the predominantly *in vitro* measured kinetic parameters are a proper approximation. Unfortunately, the availability of this information mostly restricts the applicability of kinetic modeling to small-scale and at most medium-scale models.

The examination of large-scale metabolic networks is rendered possible via *stoichiometric modeling*. Here, the complexity of the reaction rate description is reduced to the topology of the underlying system, *i.e.*, only the stoichiometry and the directionality of the biochemical reactions as well as physiologically plausible minimum and maximum reaction rates (flux boundaries) are taken into account. Based on the quasi-steady state approximation for the internal metabolites, the systemic behavior and respective properties pertaining to homeostatic conditions can be examined (see section 1.4.2). By means of network-based pathway analyses, such as elementary modes [26] and extreme pathways [27], insights into pathway structure and reaction essentiality can be gained. The usage of additional (invariant and/or environment-specific) restrictions of the solution space render constraint-based analyses possible. On the one hand, measured (mainly extracellular) fluxes and/or (de)activating regulatory constraints enable the elucidation of reaction contributions to overall metabolic processes (metabolic flux analysis). On the other hand, assuming an optimal realization of metabolic processes in a cell owing to the evolutionary pressure

provides the means for predictive analyses by using the so-called flux balance analysis (FBA; see section 1.4.3) [28]. By means of linear programming the model is rendered "capable of predicting the [molecular] phenotype (flux distribution) that will be expressed under certain conditions (given by the input constraints)" [29]. The drawback of those investigations within the metabolic context is disregarding the dynamic behavior of the internal processes, *i.e.*, understanding the metabolite concentration adaptations and detailed regulatory processes. However, the simplification of the metabolic system does not constrain the solution space, since any steady-state solution of a kinetic model is also a solution of a stoichiometric model with proper flux boundaries. Moreover, it is noteworthy that the quasi-steady state approximation does not imply that "the dynamic nature of the entire process is completely disregarded, because dynamic extracellular processes, such as substrates uptake and product formation, can still be considered" [29].

Compliant with the systems biology approach, the research efforts for modeling metabolic processes should be directed towards rendering kinetic modeling applicable on a large scale. This would allow both, a detailed characterization of molecular mechanisms and their dynamic behavior as well as the examination of the interaction with all system's level and the environmental factors. Since such a model is not yet available, the crucial step in modeling metabolic processes is the selection of the modeling approach according to the biological question(s) to address. Moreover, the modeling outcome highly depends on the available data, and the assumptions which are closest to the examined scenario. Knowledge about the shortcomings and advantages of the respective approaches, in combination with these three criteria, can reduce the risk of misinterpreting the data and the emerging principles underlined by the molecular mechanisms.

1.4. Underlying concepts and methods of the thesis

This section briefly covers the key biological and modeling concepts and methods used in several chapters. The methods exclusively applied in one chapter are provided in those respective method sections.

1.4.1. Plant photosynthesis

Photosynthesis in plants describes the process of carbon fixation, in terms of CO_2 assimilation, by means of splitting H_2O and releasing O_2. This process is realized in distinct ways due to plants' adaptation to different climatic conditions. Increased light intensities, elevated temperatures and dryness [30] result in morphological modifications, *e.g.*, increased thickened and fleshy leaves, as well as anatomical adaptations, *e.g.*, Kranz anatomy with almost uniform mesophyll layer and absence of palisade layer, respectively. Biochemically, C_4 and CAM plants differ from C_3 plants in the ways in which carbon is fixed. In C_4 and CAM plants, a prior CO_2 fixation step is introduced that separates the initial and final CO_2 assimilation spatially and temporally, respectively. As a result, while C_4 plants can increase the photosynthetic efficiency regarding ribulose-1,5-bisphosphate carboxylase/oxygenase, CAM plants can enhance the efficiency of H_2O utilization. In the following, the focus is on the basic photosynthesis processes of C_3 plants, dealt with in this thesis.

The photosynthetic pathway comprises the light reactions involving the related oxygen-evolving complex, and the Calvin-Benson cycle (CBC). In the CBC pathway the inorganic carbon is assimilated, for which the light reactions provide the required cellular energy [31]. In order to integrate the light reaction processes within the scope of metabolic modeling, the photon absorption, the electron transfer through the protein complexes, and the proton translocation across thylakoid membrane have to be included (see Chapter 5). According to the available Enzyme Commission number in AraCyc [32], the light reactions can be encompassed by five biochemical reactions. For simplification, the underlying O_2 production by splitting H_2O is merged with the processes conducted at the photosystem II. In contrast, the underlying biochemical reactions of the CBC are modeled on different levels of details, depending on the considered aspects of photosynthesis (see Chapter 2). The CBC comprises 13 biochemical reactions catalyzed by 11 different enzymes, and can be divided into three stages: (1) the carboxylation of ribulose 1,5-bisphosphate, forming two molecules of 3-phosphoglycerate, (2) the reduction of 3-phosphoglycerate to glyceraldehyde-3-phosphate, the branch intermediate for sucrose synthesis, and (3) the regeneration of the CO_2 acceptor. The regeneration stage of the CBC contains, the branch-point towards starch synthesis as well as several shared reactions with gluconeogenesis and pentose phosphate pathway. This position as the branching point to all further anabolic processes in plants is the reason for the great importance of the CBC.

1.4.2. Quasi-steady state approximation

The quasi-steady state approximation describes the simplification of the metabolic system with respect to the time-hierarchy of the underlying biochemical reactions. Based on this time-hierarchy, the reactions can be categorized into *slow* and *fast* [33]. After a short transient period, the metabolite concentrations of the substrates pertaining to a fast reaction can be considered constant (although, they may exhibit minor changes) [34]. In addition, these changes are very small compared to those in slow reactions, supporting the approximation of constant metabolite concentrations over time, termed quasi-steady state approximation. This assumption is well-established for enzyme-substrate complex formation compared to product formation in kinetic modeling [35] as well as for intracellular processes compared to extracellular ones in stoichiometric modeling [29].

Based on the quasi-steady state approximation, often homeostatic conditions are simulated. Cannon [36] devised the term homeostasis as: "The coordinated physiological processes which maintain most of the steady states in the organism are so complex and so peculiar to living beings [. . .] that I have suggested a special designation for these states, homeostasis. The word does not imply something set and immobile, a stagnation. It means a condition–a condition which may vary, but which is relatively constant" [36]. In line with this, homeostasis is a property maintaining the state of the system rather than its functions [37], and, accordingly, is at best described by a stable steady state referring to "unstressed" environmental conditions.

1.4.3. Constraint-based methods for stoichiometric modeling

Flux balance analysis[1]   allows predicting the synthesis and/or consumption rate of specific metabolites or combinations thereof, *e.g.*, representatives of cell biomass as a growth indicator. The underlying assumption is that the metabolic processes of cells have evolved towards an optimal behavior, forced by the evolutionary pressure, *i.e.*, the metabolic fluxes are regulated such that an optimal flux distribution is achieved [28, 29]. Based on an ascertained metabolic objective, *e.g.*, starch synthesis or growth enhancement in terms of biomass production, a linear program is formulated composed of: (1) the optimization of the objective function, $\mathbf{c}^T\mathbf{v}$, which can be any linear combination of fluxes $\mathbf{v}$, where $\mathbf{c}$ denote the objective coefficients, (2) the steady state postulation, where $\mathbf{S}$ is the stoichiometric matrix, (3) the specifications of the reaction directionalities via lower and upper flux boundaries, $\mathbf{v}^{min}$ and $\mathbf{v}^{max}$, (4) and possible further constraints (Equation 1.1):

$$
\begin{aligned}
\max/\min \quad & \mathbf{c}^T\mathbf{v} \\
\text{s.t.} \quad & \mathbf{S}\cdot\mathbf{v} = 0 \\
& \mathbf{v}^{min} \leq \mathbf{v} \leq \mathbf{v}^{max}
\end{aligned}
\qquad . \tag{1.1}
$$

It is noteworthy that the resulting flux distribution may not be unique, *i.e.*, it is only a representative achieving the maximum/minimum of the objective function and obeying the constraints. Modifications of the conventional FBA formulation enable the examination of quadratic problems, transience of metabolism (dynamic FBA) [39], and provide unique solutions by performing an additional optimization step [29, 40], or selecting a specific distribution (geometric FBA) [41].

Certainly, FBA has limitations, *e.g.*, by neglecting the reaction kinetics, it cannot be used to predict metabolite concentrations and can only account for specific regulatory effects, such as reaction (de)activation [28].

Flux variability analysis[1]   enables the determination of alternative pathways, in terms of different combinations of reaction sequences, achieving the same optimal behavior of the cell with respect to a predefined metabolic objective. Such alternative pathways are shunts, by-passes or overlapping biochemical pathways, such as the CBC and pentose phosphate pathway, and reflect the redundancy which occurs in most metabolic networks [42]. According to this, flux variability analysis quantifies the flexibility of the underlying metabolic processes while maintaining the prespecified yield. By a series of optimization steps, the feasible range of steady-state fluxes for each reaction is determined. Here, a reaction is denoted as variable if the potential flux range is more than 1 % of the total flux range.

Flux coupling analysis[1]   facilitates the analysis of metabolic relationships and the detection of functionally related reactions, by examining dependencies between two reactions that are "active" together, in terms of carrying a nonzero flux in a steady state [43]. By a series of optimization steps, the effects of changing the flux of one coupling partner on the second reaction can be investigated [44]. Thereby, one distinguishes three types of couplings,

---

[1]Based partially on the supplemental material of Arnold and Nikoloski (2014) [38] (see Chapter 5).

namely fully-, partially- and directionally-coupled reactions. Two reactions are fully-coupled, if changing the reaction flux of one coupling partner results in a fixed proportional change of the flux of the other reaction. On the other hand, for partially-coupled reactions such a change only causes an arbitrarily change of the corresponding coupling partner. Importantly, those effects are mutual, whereby the change of each of both reactions induce the respective change of the other reaction flux. In directionally-coupled reactions, this relationship is asymmetric. All reactions which do not fall into these categories are termed uncoupled [45].

As the computation of flux couplings only relies on the stoichiometry of the metabolic network, constraints such as flux boundaries are ignored. The compared models dramatically differ in the number of import and export reactions which highly influence the coupling of reactions. Consequently, all nonessential importers and exporters are removed to establish a comparable framework. Moreover, all "nonactive", so-called blocked, reactions are eliminated, including also those arising from the removal of the importers and exporters (in accordance to the algorithm's procedure) [44]. Accordingly, a functional networks is obtained for which the couplings are identified.

1.5. Organization of the thesis

The focus of this thesis is to illustrate different aspects of modeling photosynthesis and related metabolic processes, and to reveal their applicability for addressing biological questions. Depending on the different question(s) to address, the suitable modeling approach is chosen for each analysis, from detailed small-scale reaction kinetic towards large-scale stoichiometric modeling. The analyses are applied to Arabidopsis, as the best-investigated C_3 plant model organism.

In order to investigate the photosynthetic processes in detail, first of all, either a model has to be assembled or a proper model can be selected from the multitude of existing ones. In Chapter 2, a quantitative comparative analysis of existing CBC models is presented to determine the best-performing ones with respect to a combination of criteria ensuring biological functionality. For that purpose, several models had to be amended to get them properly working and allow comparability (published as Arnold and Nikoloski, Arnold and Nikoloski, 2011, 2014 [46, 47][2]).

The most-promising candidate of CBC models, capable for metabolic engineering applications, is utilized in Chapter 3 to demonstrate the impact of enzyme turnover and synthesis by integrating enzyme action and stoichiometric modeling. In doing so, the costs of enzyme *de novo* synthesis can be incorporated[3] which provide the means to expand the concept of metabolic costs by the maintenance costs of enzyme provision. As a result, the metabolic costs for synthesizing the photosynthesis related end-products are determined, thus, elucidating the transitory carbon storage mechanism.

In Chapter 4, the modeling attempts of the photorespiratory pathway in various level of detail and considered context are characterized. Thereby, the interconnection of photore-

---

[2]The review article in Trends in Plant Science was intended as a succeeding publication to the original article in Mathematics and Computers in Simulation (special issue for outcomes arising from the Workshop DIEBM2010). Repeatedly delay in the publication process required the inclusion of the underlying methods in the supplemental material of Arnold and Nikoloski, 2011 [46].

[3]Amongst others, based on the model presented in Chapter 5, Arnold and Nikoloski, 2014 [38].

spiration and photosynthesis is highlighted, and its characteristic of bridging carbon and nitrogen metabolism and the corresponding still controversial role are assessed (published as Arnold and Nikoloski, 2013 [48]).

For the purpose of investigating the photosynthetic processes within the immediate metabolic context, the demand for a large-scale model of Arabidopsis' primary metabolism of high quality is raised in Chapter 5. In contrast to the existing models, a bottom-up assembled model is devised, capable of simulating photoautotrophic conditions by obeying the law of mass conservation and without the usage of gap-filling algorithms. Moreover, the additional annotation of enzyme complex structures enables the calculation of enzyme synthesis costs which, in turn, is used to explore the trade-off between protein synthesis and growth (under revision as Arnold and Nikoloski, 2014 [38]).

Metabolic costs and their relevance highly depend on the present environmental conditions. Chapter 6 reveals the effects of different nitrogen sources as well as light and carbon-source conditions on the metabolic costs of amino acid synthesis. Moreover, the comparative analysis of different models across the considered conditions reinforces the importance of the model selection and a careful curation (under revision as Arnold *et al.*, 2014 [49]).

The results and the drawn conclusions of each analysis are summarized and discussed in Chapter 7. Moreover, a potential extension is considered offering the perspective of modeling on a large scale by at least partially accounting for dynamic information of the metabolism. Such an approach may set the course for diminishing the shortcomings of the currently existing two distinct approaches of modeling metabolism.

Chapter 2.

A quantitative comparison of Calvin-Benson cycle models[4]

The Calvin-Benson cycle provides the precursors for biomass synthesis necessary for plant growth. The dynamic behavior and yield of the CBC depend on the environmental conditions and regulation of the cellular state. Accurate quantitative models hold the promise of identifying the key determinants of the tightly regulated CBC function and their effects on the responses in future climates. We provide an integrative analysis of the largest compendium of existing models for photosynthetic processes. Based on the proposed ranking, our framework facilitates the discovery of best-performing models with regard to metabolomics data and of candidates for metabolic engineering.

2.1. Modeling as a step towards understanding

Mathematical modeling of integrated processes lends itself as a useful tool for *in silico* probing of biological systems. The existing modeling paradigms hold the promise to tackle one of the greatest challenges in plant physiology, improving the understanding of the CBC, its limiting steps, and the relation to plant growth [50]. Mathematical modeling allows for placing this important metabolic pathway in the context of its cellular milieu (*e.g.*, surrounding pathways) and the entire carbon cycle. Moreover, it provides the means for predicting systemic behavior on various levels of the system, rendering it valuable in planning laborious experiments aimed at confirming posited hypotheses.

Here we present a comprehensive critical review of the existing models of the CBC. By assembling and implementing a compendium of 15 models, our aim is to identify those model candidates that provide quantitatively accurate predictions for the levels of CBC intermediates and show biologically plausible dynamics. These candidates, in turn, can be used in metabolic engineering [51] and in the design of synthetic metabolic pathways for improved carbon fixation, growth and yield [52].

2.2. Overview of model components

The challenge of modeling involves selection of the relevant biochemical reactions. In the case of the CBC, the most general description includes the three stages: carboxylation, reduction and regeneration, which can further be divided into more specific sub-processes

---

down to the level of single reactions. One of the key reactions is the initial step, whereby carbon dioxide (CO_2) enters the CBC. It is termed the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO; EC 4.1.1.39) reaction, although the corresponding enzyme name is misleading [53]. In fact, the enzyme initially reacts with ribulose-1,5-bisphosphate (RuBP), resulting in the enediol-enzyme complex which can then capture CO_2 (carboxylation) or oxygen (O_2, oxygenation) [54, 55]. Due to its biochemical importance, this step has been included in almost all modeling attempts either at reaction level or its representation as substrate-enzyme and product-enzyme steps [56]. The other key reactions of the CBC, such as those catalyzed by fructose-1,6-bisphosphatase (FBPase; EC 3.1.3.11), sedoheptulose-1,7-bisphosphatase (SBPase; EC 3.1.3.37) and ribulose-5-phosphate (Ru5P) kinase (EC 2.7.1.19), also appear in almost every model at the reaction level [51, 56–61].

In the case of the end-product processes, the key reactions involve adenosine diphosphate (ADP) glucose pyrophosphorylase (AGPase, EC 2.7.7.27) for starch synthesis, uridine triphosphate glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) for sucrose synthesis and RuBisCO (oxygenase function) for photorespiration. Moreover, the steps branching from the CBC and competing for the branch-point intermediates have to be carefully considered. For instance, the triose phosphates (TP) are the branch intermediates for sucrose synthesis, and the export of TP out of the chloroplast is the corresponding branching point. There is a strict counter-exchange of TP with inorganic phosphate (Pi) via the triose-phosphate translocator (TPT) which does not follow classical enzyme kinetics, thus requiring careful selection of modeling strategies [62].

The regulatory processes, activation and inhibition, enable the adaptation to different conditions and represent yet another aspect of the modeled pathways. Light activation, as the best characterized effect of an external regulator, affects several enzymes of CBC and of the end-product processes, which undergo a 2 to 40-fold increase in activity after the onset of light [63–65].

A common example of an external metabolic regulator of the CBC is the CO_2–O_2 competition for RuBisCO. In both (*i.e.*, carboxylation and oxygenation reactions), each gaseous substrate acts as an inhibitor of the reaction involving the other, *i.e.*, CO_2 and O_2 influence the CBC/photorespiration ratio [66] (Appendix A.1.3).

Internal metabolic regulators can be distinguished by the direction of their effect in a reaction chain. If a reaction does not take place, its substrate(s) will accumulate and its product(s) will be depleted. The concentration of substrates or products could then inhibit the downstream reactions, termed forward regulation. However, in backward regulation, metabolites may activate or inhibit previous reactions by affecting the activity of the corresponding enzymes or by producing the necessary cofactor.

Forward regulation for photosynthetic processes is experimentally proven for AGPase. The CBC intermediate 3-phosphoglycerate (PGA) increases the activity of AGPase by 9 to 80-fold and enforces starch production [67, 68]. By contrast, Pi inhibits this reaction and it was shown that the stromal PGA/Pi ratio, in fact, exerts control over CO_2 incorporation into starch [69, 70].

A typical backward regulation within the CBC is the product inhibition (also termed product competition) of FBPase, whereby increased concentration of fructose-6-phosphate (F6P) inhibits the enzyme activity of the previous reaction [71]. More complex cases of regulatory processes exist, such as substrate competition for a CBC enzyme. Transketolase

(EC 2.2.1.1) catalyzes two reversible reactions, the transformation of F6P or sedoheptulose-7-phosphate with glyceraldehyde-3-phosphate (GAP) to erythrose-4-phosphate or ribose-5-phosphate (R5P) with xylulose-5-phosphate (X5P), respectively (Appendix Equation A.32). It may be expected that the different substrates inhibit the competing reaction, although only the inhibition of F6P and R5P has been experimentally established. However, the unproven inhibition by the remaining substrates has often been used in modeling of the CBC [51, 56, 61], although it may lead to doubtful conclusions. Moreover, this inhibitory effect has been proven only for *Leishmania mexicana* [72], an obligate intracellular protozoan parasite. The integration of such regulatory processes, not experimentally proven for higher plants, or even for the model plant *Arabidopsis thaliana* (Arabidopsis), could result in more implausible predictions. This exemplifies the fact that regulatory processes are the most challenging part of modeling the CBC and, therefore, have to be examined with special care.

2.3. Classification of CBC models

By thoroughly reviewing the literature spanning the past three decades, we assembled the largest existing compendium of models for the CBC. The compendium consists of 15 models, including the initial modeling attempts covering various contexts of photosynthesis-related processes and some of their widely cited extensions [51, 56–61, 73–80] (Appendix Table A.1). We provide a detailed classification of the models included in the compendium based on (1) model boundaries, *i.e.*, the coverage of CBC together with the end-product pathways; (2) levels of cellular organization, *i.e.*, leaf, cell or compartment; (3) complexity of kinetics [81], translating the model structure into mathematical equations for analyzing spatiotemporal properties; and (4) included regulatory processes, specifying the regulators, their types and the resulting formalization [82].

The considered models differ in their boundaries due to the aspects of photosynthesis they cover. Five of the models focus on the RuBisCO reaction and merge the remaining steps of the CBC (Table 2.1, column 3). There are three models describing the CBC in detail but omit related processes. A group of three models investigate the processes taking place in the chloroplast, namely the CBC and starch synthesis, and the remaining four models additionally integrate the different end-product pathways. The model of Laisk *et al.* (2006) [56] originally includes the photosystems and electron transport chain; however, in the model comparison, we use a reduced version capturing the CBC and the end-product pathways (Appendix A.1.2).

The model boundaries partly affect the levels of cellular organization. Models focusing exclusively on the CBC and the processes in the chloroplast were modeled at compartment level (Table 12.1, column 2), whereas models including sucrose synthesis (and photorespiration) span the cell level. As a result, the cell-level models include details concerning compartmentalization and the related transport steps. The remaining models describe photosynthesis at the most complex biological level, *i.e.*, the leaf. This level necessitates consideration of diffusion and the difference of atmospheric and intercellular partial pressure of gases. However, these details were merely included in the RuBisCO-focusing group, consisting of the smallest models.

A critical part of kinetic modeling is the translation of the model structure into ma-

Table 2.1.: Classification of the 15 CBC models.

Model	Classification				Ref
	Level	Boundary	Kinetics	Regulation	
Farquhar *et al.*			[red] [blue] [black]	c	[73]
Medlyn *et al.*			[red] [purple] [black]	c	[74]
Schultz	Leaf	RuBisCO	[red] [purple] [black]	c	[75]
Sharkey *et al.*			[red] [purple] [black]	c	[76]
Damour and Urban			[red] [purple] [black]	c	[77]
Fridlyand and Scheibe	Plastid		[blue]		[57]
Zhu *et al.* (2009)			[blue]		[78]
Giersch *et al.*			[blue] [purple] [black]	p	[79]
Hahn		++	[red]		[80]
Poolman *et al.*		+	[red] [blue] [black]	a c m n p	[60]
Pettersson and Ryde-Pettersson	Cell	+ CBC	[green] [blue] [black]	a c m n p	[58]
Woodrow and Mott		+	[green] [blue] [purple]	c p	[59]
Laisk *et al.* (1989)		++	[green] [blue] [black]	a c	[61]
Laisk *et al.* (2006)		++	[green] [blue] [purple] [black]	a x	[56]
Zhu *et al.* (2007)		+++	[green] [blue] [purple] [black]	a c m n p	[51]

+ – CBC & starch; ++ – CBC & starch & sucrose; +++ – CBC & starch & sucrose & photorespiration; [red] – mass action; [green] – equilibrium approximation; [blue] – Michaelis–Menten; [purple] – Michaelis–Menten like; [black] – special function; a – activation; c – competitive; m – mixed; n – non-competitive; p – competitive product; x – unidentified inhibition.

thematical equations, requiring specification of the reaction kinetics. The most common approaches for modeling reaction kinetics include mass-action and Michaelis-Menten kinetics (Table 2.1, column 4; Appendix A.1.3). In the compendium, other types of kinetics in the form of Michaelis-Menten-like and special functions are also considered. The special functions comprise the kinetics proposed by Giersch, specifically tailored to TPT [62], and all remaining kinetics within the compendium which do not fall into the common theoretical frameworks. Regardless of the type, the kinetic functions describe the velocities of the modeled reactions, and determine the temporal changes of the metabolite concentrations which can be mathematically described by differential equations.

Nevertheless, there exist simplifications of the kinetics, such as the equilibrium approximation, which are not suitable for time-resolved description. For reactions modeled by this approximation such as TP isomerase (EC 5.3.1.1), one assumes rapid settling into equilibrium. This assumption may be advantageous as it allows reducing the system size. The metabolites involved in reactions, which are very fast compared to the adjacent reactions, are merged into a metabolic pool and their transient behavior is determined by the dependence to this pool [51, 56, 61] (Appendix Equation A.20). Such simplifications are common and reliable for small metabolic pools including two or three different intermediates. However, equilibrium approximation of many reactions forming one pool may restrict or even disable any temporal analysis of the remaining system, as the differential equations can be evaluated solely at steady state [58, 59], without any additional assumption (Appendix A.1.3). As a result, equilibrium approximation has not been used to model

the key reactions discussed above.

The employed kinetics can be regarded as the determinants of the regulatory processes. For instance, the mass-action kinetics and the equilibrium approximation do not allow the inclusion of any regulation term. For the CBC models, Michaelis-Menten, Michaelis-Menten-like as well as special-functions kinetics can integrate regulatory processes and reinforce their usage for modeling the key reactions. In the compendium, there are several combinations of regulators and regulated reactions: altogether, 19 reactions are regulated by 19 regulators (Appendix A.1.3). We further distinguished six different (sub)types of regulation: one activation and five inhibition (*i.e.*, competitive, mixed, non-competitive, competitive product and unidentified) (Table 2.1, column 5; Appendix Tables A.3 and A.4). This resulted in 29 different combinations of regulators and regulation types across all reactions, and up to six different terms of regulation for a single reaction.

2.4. Model analysis and comparison

To capture biologically realistic scenarios, a model of the CBC should be stable and robust to small parameter perturbations. Moreover, it should reflect the experimental data. These three criteria form the basis for the ranking of the models in the compendium. Furthermore, models whose dynamics are more similar to those determined as well-performing will be considered more reliable. To this end, we carried out the following analyses with respect to (1) sensitivity, (2) stability, (3) robustness and (4) residual sum of squares (RSS) at the resulting steady states.

The implementation of the models in the compendium has proven a challenging step, especially the case of reproducing the published results. To facilitate the usage of the compendium, we implemented the included models in Systems Biology Markup Language (SBML) suitable for analysis tools, such as: SBtoolbox2 and COPASI (Appendix A.1.4, Table A.5). Nevertheless, further challenges arose: the models of Pettersson and Ryde-Pettersson [58], Poolman *et al.* [60] and Zhu *et al.* (2007) [51] use the same incorrect velocity for the AGPase reaction due to a unit inconsistency first appearing in Pettersson and Ryde-Pettersson [58] (Appendix A.1.2). We resolved this issue in accordance with the authors' suggestion. A similar inconsistency was corrected for the reaction describing the transformation of PGA to sink in the model of Zhu *et al.* (2009) [78]. In addition, for some models not all kinetic parameters are given. Here, we resolved this issue by using values obtained from prior model versions [61, 83, 84], alternative models [51, 56, 58, 61, 73, 74] and literature [85–87] (Appendix Table A.2). Nevertheless, even after this step, the model of Fridlyand and Scheibe [57] has missing parameter values. As a result, this model was excluded from further analyses.

As the temporal analysis requires model output in the form of time series, the models based on the steady-state assumption cannot be evaluated in the time domain; therefore, the models of Pettersson and Ryde-Pettersson [58] and Woodrow and Mott [59] were excluded from temporal analyses and were not implemented in SBML. Furthermore, for the model of Laisk *et al.* (1989) [61], imaginary concentrations for ADP arise (Appendix Equation A.11) at physiologically plausible values for ADP-glucose and adenylphosphate (Appendix Equations A.13 and A.14). Therefore, this model was also been excluded from any further analysis.

To enable a fair comparison, the initial conditions for all models must be the same. To this end, we chose the inner and outer metabolite concentrations from Zhu *et al.* (2007) [51] as a reference data set of initial values, due to its extensive coverage and literature support. This includes the boundary conditions for the CBC – namely, CO_2, O_2, the energy equivalents provided by the electron transport chain and the Pi pool. Whether the boundary conditions are inner metabolites, outer metabolites, or even not integrated, varies from model to model. However, if integrated, the initial values of the boundary conditions are almost the same within the compendium (Table 2.3). Only the model of Hahn [80] uses very different initial values due to its special metabolites, such as thiamine pyrophosphate glycoaldehyde. Based on these values, we determined the steady state by using the resulting time series or calculated it by using the steady-state assumption. Startlingly, very different steady-state solutions were obtained from the thirteen suitable models, *e.g.*, for PGA and TP within 0.147–215.042 and 0.113–8.503 mM (Figure 2.1; Appendix Table A.6), respectively. The set of obtained steady-state solutions for the metabolite concentrations is the starting point for the analyses.

2.4.1. Stability and robustness analyses

Metabolic systems have evolved to provide stable and robust operation in a well-defined physiological range, reflecting the effects of external and internal perturbations, respectively. To check which of the models are endowed with these properties, we carried out a stability analysis (Methods 2.6.2) to investigate the effects of external perturbations on the system. All of the 11 models suitable for this analysis were stable at the given steady state (Table 2.2, column 2; Appendix Tables A.7–A.12).

To investigate the robustness, namely, the effects of changing values for the kinetic parameters, as internal perturbations, we tested whether the system exhibits small deviations from the given steady state (Methods 2.6.3). This evaluation was repeated 10^5 times and the relative frequency of robust instances was recorded (Table 2.2, column 3; Algorithm 2.1). Among the six most robust models are the five of the RuBisCO-focusing group [73–77], and the model of Zhu *et al.* (2009) [78]. These models are also the six smallest within the compendium, leading to the claim that robustness decreases with increasing model complexity. The order of the remaining models supports this claim: Giersch *et al.* [79] (ranked seventh) covers the CBC only (but is larger than that of Zhu *et al.*, 2009 [78]); whereas Poolman *et al.* [60] (ranked eighth), Laisk *et al.* (2006) [56] (ninth) and Zhu *et al.* (2007) [51] (tenth) comprise each one additional end-product process. The model of Hahn [80] is the only which does not follow the rule. The underlying mass-action kinetics does not permit inclusion of regulatory processes which are needed to stabilize thermodynamically highly irreversible fluxes, in particular, the carbon flux through the CBC down to the end-product processes.

2.4.2. Compliance with data

The biological relevance of the models is reinforced by the compliance of the computational predictions to experimental data. We calculated the RSS between the given steady-state concentrations and metabolomics data of Arabidopsis [88] to investigate the physiological plausibility of the predictions of each model. The data include one of the most recent
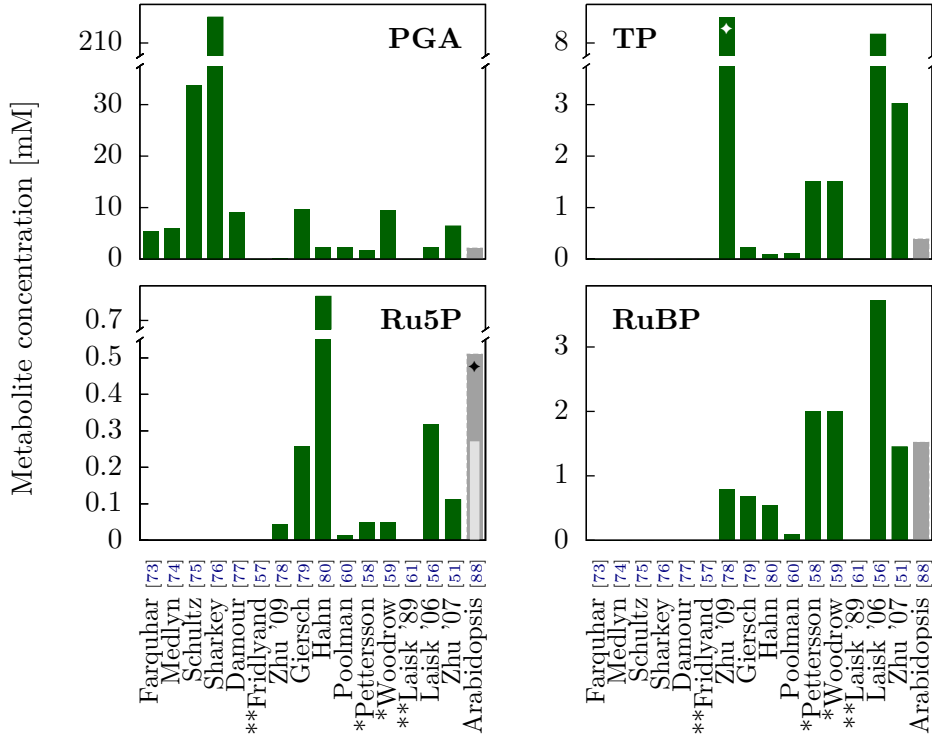
Figure 2.1.: Quantitative steady-state concentrations of PGA, TP, Ru5P and RuBP for each model within the compendium (green) and the corresponding experimental data (gray). For Zhu *et al.* (2009) [78], only GAP is available instead of TP ($\diamond$). Experimentally, Ru5P could be measured only together with X5P ($\blacklozenge$). The brighter gray indicates the extrapolated concentration of Ru5P using the assignment of Zhu *et al.* (2007) [51]. The values of the models marked with * are analytically calculated, those marked with ** could not be evaluated (and are excluded for this reason) and the remaining are simulated.

measurements capturing almost all CBC metabolites. Moreover, the experimental data were obtained under non-limiting CO_2, O_2 and light conditions in accordance with the assumptions of most of the models. As mentioned above, the model of Hahn [80] is the only outlier (Appendix Table A.5). We note that CBC metabolites are compartmentalized, but only partial experimental support is available for them [89, 90]. To use these data, we converted them by using the subcellular volumes [91] together with assumptions, such as GAP is 5 % of dihydroxyacetone phosphate [92] (Methods 2.6.4; Appendix Table A.13). We would like to stress that no time series is required for these calculations and, consequently, the RSS can be calculated for 13 models in this compendium. Because of the different model boundaries, the RSS involves different numbers of metabolites for each model. For model comparison, we employed the averaged RSS, $\overline{\text{RSS}}$ (Table 2.2, column 4; Methods 2.6.4). The models of Pettersson and Ryde-Pettersson [58], Poolman *et al.* [60] and Woodrow and Mott [59] best describe the Arabidopsis data. The high ranking of this group of models is comprehensible taking into account that the two last models are extensions of

the first and, therefore, are similar in structure, size and model boundaries. However, the model of Farquhar *et al.* [73] has the fourth high compliance to the data, whereas the other RuBisCO-focusing models are ranked between ninth and thirteenth. This shows that similar model structure only does not imply similar compliance to data. The relationship between low rank and high model complexity observed in the case of robustness analysis is not found for the case of RSS.

2.4.3. Sensitivity analysis

Models describing the same pathway should show similar dynamic behavior in the vicinity of their steady states. At present, a comparison of dynamic behavior across the models is hampered by the lack of time-resolved experimental data for the concentration of metabolites appearing in more than one compartment. Therefore, to determine the similarity for two models, we carried out a sensitivity analysis comprising metabolic control analysis (MCA) for each model [93] and the Kendall rank correlation of the outcomes for each pair of models [94]. The concept of MCA allows investigation of the effects of small parameter perturbations on a steady state. We focused on the elasticity and flux control coefficients which describe the contribution of perturbations in metabolite concentrations on the velocity of a single reaction and changes in enzyme activity on all velocities, respectively. To enable the comparison of the different models, six new matrices were created from each original MCA matrix by reduction and merging of entries, corresponding to the model boundaries (Appendix Tables A.14 and A.15). The similarities between the models could then be investigated for the most general form of the CBC, more detailed versions of the CBC with five steps, as well as for the entire CBC with starch and/or sucrose synthesis.

In addition, for any pair of models, the Kendall rank correlation τ was used to quantify the similarity of the MCA results (Methods 2.6.5, Equation 2.8). Larger τ implies more similar behavior between the compared models (Figure 2.2; Appendix Tables A.16–A.26, Figures A.1 and A.2). We note that τ cannot be calculated in the case of only one substrate and two reactions, since it requires at least six arguments. Therefore, five of the remaining 11 models, focusing on the RuBisCO reaction, are excluded from this analysis.

For the models of Zhu *et al.* (2009) [78], Giersch *et al.* [79], Hahn [80], Poolman *et al.* [60], Laisk *et al.* (2006) [56] and Zhu *et al.* (2007) [51], the Kendall correlation of the elasticities, τ_E, yields a value greater than zero in each of the six cases for each possible pair of models. This provides a similarity between all models with respect to local effects on reaction velocities. However, regarding the global effects captured by the Kendall correlation of the flux control coefficients, τ_E, the models differ considerably, which is demonstrated by $\tau_E \in [-1, 1]$. Large variations do not only appear for different pairwise combinations, but also for the same combination among the different cases. For a given model, the average of the six cases is used to identify the most similar candidate from the compendium. This is done for the elasticity ($\bar{\tau}_E$) and the flux control coefficients ($\bar{\tau}_E$), as well as for their combination, $\bar{\tau}_{av}$ (Appendix Table A.27). The mutually most similar pairs of models are: Zhu *et al.* (2009) [78] and Hahn [80], Giersch *et al.* [79] and Laisk *et al.* (2006) [56], and Poolman *et al.* [60] and Zhu *et al.* (2007) [51]. These similarities, $\bar{\tau}_{av}$, were integrated into the final ranking (Appendix Table A.28, column 11), so that models showing similar dynamic behavior to well-performing ones became higher ranked. Obviously, worse-performing models and those similar to them are then ranked lower.
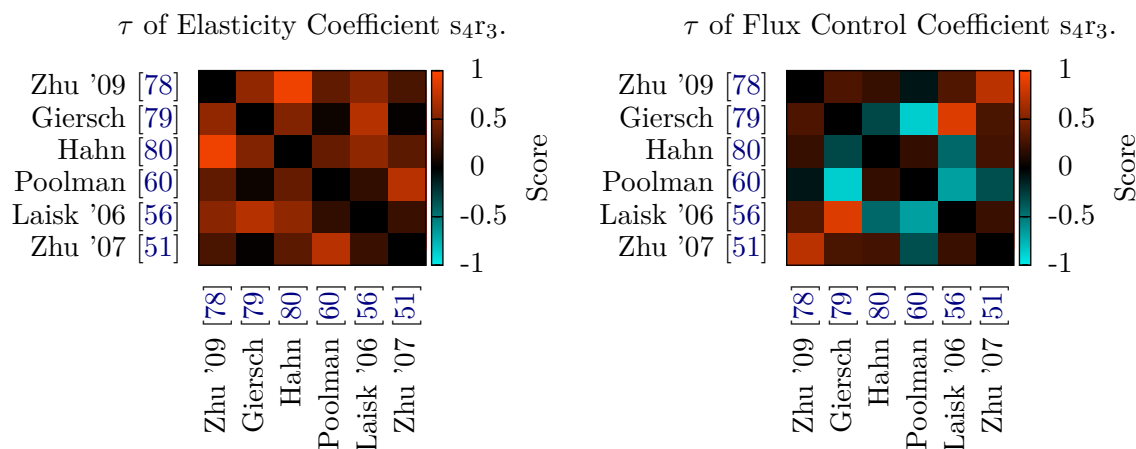
Figure 2.2.: Kendall τ's for pairwise comparison of the category considering 4 substrates and 3 reactions (s_4r_3; Appendix Tables A.16 and A.17) where: red indicates very similar, black neutral and turquoise very different dynamic behavior (legend right hand side).

2.4.4. Ranking

The applied analyses favor different models as best-performing (Appendix Table A.28). To determine the best-performing models, we worked out a score combining the ranking of each criterion with respect to their relevance. In our opinion, for a biologically reliable CBC model, (1) the stability is of paramount importance, since small perturbations of metabolite concentrations should not lead to the system's break-down; (2) the compliance of the model predictions to experimental data is the second crucial property, because regardless of how stable a model is, if the steady state is physiologically unreliable the model is useless for biological predictions; and (3) the model robustness is consequently the third basic principle. To demonstrate their relevance, the criteria are weighted by decreasing factors, here chosen as: four, two and one, respectively (Appendix Equation A.33).

The resulting preliminary score for each model (Table 2.2, column 5; Appendix Table A.28) is further combined with the one of the most similar model via linear combination. For the models applicable for sensitivity analysis, the similarity value ($\bar{\tau}_{av}$) is used as the weighting factor (Appendix Equation A.34). So, the greater the similarity of the mutually most similar models, the more the preliminary ranking of the most similar model influence the final score of a model (Table 2.2, column 7; Appendix Table A.28). For the models unsuitable for sensitivity analysis, the weighting factor is set to zero. Therefore, their final rank is affected only by the preliminary score.

The overall best-performing model of this ranking is that of Farquhar *et al.* [73]. Although the model and its extensions are criticized [53], when restricted to the ranking criteria it performs best to investigate carbon fixation processes. Due to the focus on the RuBisCO reaction and the merging of the remaining steps of the CBC, *i.e.*, the small model size, the RuBisCO-focusing models [73–77] (ranked first, fourth, sixth, tenth and eleventh) provide insufficient information for metabolic engineering. Moreover, the last four models

Table 2.2.: Results and model rankings of the investigated 15 models. The results of stability and robustness analysis as well as the compliance with data ($\overline{RSS}$) are included in columns 2–4. Column 6 comprise the most similar model as far as the model is suitable for sensitivity analysis. Columns 5 and 7 contain the preliminary and the final model rankings, respectively. The shadings indicate the candidates for metabolic engineering.

Model	Results			Ranking			Ref
	Stability	Robustness	$\overline{RSS}[mM^2]$	Preliminary	Similarity	**Final**	
Farquhar *et al.*	✓	0.4156	11.0248	1		**1**	[73]
Medlyn *et al.*	✓	0.5960	14.9133	4		**4**	[74]
Schultz	✓	0.6628	994.0520	9		**10**	[75]
Sharkey *et al.*	✓	0.0304	45343.9510	11		**11**	[76]
Damour and Urban	✓	1	55.5920	5		**6**	[77]
Fridlyand and Scheibe	‡	–	–	14		**14**	[57]
Zhu *et al.* (2009)	✓	0.0642	25.4759	8	[80]	**9**	[78]
Giersch *et al.*	✓	0.0293	11.8853	3	[56]	**3**	[79]
Hahn	✓	0	14.3666	5	[78]	**7**	[80]
Poolman *et al.*	✓	0.0187	0.7460	1	[51]	**2**	[60]
Pettersson and Ryde-Pettersson	†	–	0.2811	12		**12**	[58]
Woodrow and Mott	†	–	4.2994	13		**13**	[59]
Laisk *et al.* (1989)	‡	–	–	14		**14**	[61]
Laisk *et al.* (2006)	✓	0.0150	14.4639	5	[79]	**5**	[56]
Zhu *et al.* (2007)	✓	0.0004	14.5347	9	[60]	**8**	[51]

✓ – stable steady state; † – unstable steady state; ‡ – no steady state at all

[57–59, 61] (with respect to the ranking) are inappropriate for such applications because they cannot be evaluated in the time domain.

By dividing the models into these two groups, best-performing regarding carbon fixation or metabolic engineering, the RuBisCO-focusing [73–77] as well as the last four models [57–59, 61] should be excluded from the metabolic engineering application category. Thus, the model of Poolman *et al.* [60], ranked second, is the best-performing model suitable for metabolic engineering applications. The compliance to the Arabidopsis data is very high, promising reliable predictions. Furthermore, the models of Giersch *et al.* [79] and Laisk *et al.* (2006) [56] are promising candidates for metabolic engineering applications. The model of Giersch *et al.* [79] is more robust than that of Poolman *et al.* [60], whereas the model of Laisk *et al.* (2006) [56] provides the connection in the original model boundaries to include photosystems and the electron transport chain. Interestingly, the next model of the ranking, that of Hahn [80] (ranked seventh), is one of the simplest, as it is based on mass-action kinetics and, consequently, does not consider regulatory processes. However, these modeling assumptions undermine the importance of metabolic regulation, and may lead to inaccurate predictions, especially under changing environmental scenarios [95].

2.5. Modeling challenges ahead

This newly assembled compendium of CBC models allows for some important questions directly related to biotechnology applications to be readdressed. To this end, our model classification can serve as an easy reference of CBC models, components and relations between them, based on different modeling aspects. Moreover, the proposed analysis framework can be employed to address the optimization of the net carbon fixation, identification of missing regulators, or the maximization of biomass production. Furthermore, it can facilitate analysis of the CBC spatio-temporal dynamics of which rigorous treatments are scarce [81]. As challenges of metabolic engineering, these issues are largely unexplored in plant research. Our findings further reinforce the potential of the framework in developing models of better performance for other metabolic processes (*e.g.*, cancer development).

As modeling is an iterative process, the compendium can only point out the best-performing models if the ranking criteria together with the compendium itself are regularly updated. To this end, the latest models, as well as the new insights into the underlying biology of the photosynthetic processes, can be incorporated. For instance, at present there are no time-resolved experimental data that clearly include the metabolite concentrations in the different compartments to enable the comparison of predictions regarding the dynamic behavior of the investigated process. In this context, methods can also be incorporated to push forth various aspects of the biologically relevant theoretical analyses. For instance, the integration of the Variational Bayes measure [96] can then take into account the model complexity within the analysis of compliance to data which will be part of our future studies.

2.6. Methods

2.6.1. Steady state analysis

Each method applied in this study integrates a steady state solution of the given models. To provide the same initial values and, therefore, enable a fair comparison, we have chosen a reference data set of metabolite concentrations [51]. Hence, we can also provide similar boundary conditions, namely, CO_2, O_2, adenosine triphosphate (ATP), nicotinamide adenine dinucleotide phosphate (NADPH) and the Pi pool for almost all models (Table 2.3). The model of Hahn [80], is the only model deviating from the common conditions due to their unusual metabolite and reaction definitions, such as: thiamine pyrophosphate glycoaldehyde and the corresponding splitting of the transketolase reaction. The very different steady state values (Appendix Table A.6) give an idea of the undifferentiated modeling situation and of the importance of this study.

2.6.2. Stability analysis

Stability Analysis is a well-established method comprising the calculation of the Jacobian at the given steady state and its evaluation by considering the eigenvalues (Appendix Tables A.7–A.12). Thereby, one has to use a full-rank system which can require an initial system reduction. Therefore, the mass-conservation equations for the dependent metabolites

Table 2.3.: Initial boundary conditions for the 13 involved models within the compendium. The shaded entries denote values obtained by assignment rules.

Model	Boundary conditions [mM]					Ref
	CO_2[a]	O_2[b]	ATP	NADPH	Pi	
Farquhar *et al.*	0.230	0.210		0.210		[73]
Medlyn *et al.*	0.245	0.210		0.210		[74]
Schultz	0.350	0.210		0.210		[75]
Sharkey *et al.*	0.245	0.210		0.210		[76]
Damour and Urban	0.245	0.210		0.210		[77]
Zhu *et al.* (2009)			0.680		6.348	[78]
Giersch *et al.*			0.213		5.000	[79]
Hahn	923.790		3.875		2.500	[80]
Poolman *et al.*			0.680	0.210	0.978	[60]
Pettersson and Ryde-Pettersson			0.680	0.210	0.975	[58]
Woodrow and Mott			0.680		0.975	[59]
Laisk *et al.* (2006)	0.360	0.212	0.680	0.210	0.011	[56]
Zhu *et al.* (2007)	0.268	0.206	0.680	0.210	0.975	[51]

[a] – [µbar]; [b] – [mbar]

are used to rewrite the velocity functions in terms of the independent metabolites. Additionally, the stoichiometric matrix has to be reduced by eliminating the rows corresponding to the dependent metabolites. Such reduction is required for the models of Giersch *et al.* [79] and Hahn [80]. To determine the stability of the steady state, one has to examine the sign of the (greatest) eigenvalue(s), leading to the following conditions:

(1) If the real part of all eigenvalues is *negative*, the steady state is *asymptotically stable*.

(2) If the real part of the greatest eigenvalue is *positive*, the given steady state is *unstable*.

(3) If the real part of the greatest eigenvalue is *zero*, for a linear system the steady state is *marginal stable*. For non-linear systems the steady state has to be further investigated.

2.6.3. Robustness analysis

To test the models for robustness, we have introduced small parameter perturbations, $p = 0.05$, at the steady state, x^*, and have checked whether the system reaches the "same" steady state (Algorithm 2.1). The parameter perturbations are of the form (Equation 2.1):

$$k_{per} = \left(1 + (-1)^{r_{\{0;1\}}} p \cdot r_{[0;1]}\right) k_{orig}. \tag{2.1}$$

The "same" steady state is defined as the original with a deviation up to $d = 0.05$. We have repeated this procedure $n = 10^5$ times and have recorded the positive instances.

Algorithm 2.1: Steps of robustness analysis

Data:

perturbation, $p = 0.05$,

numbers of repetitions, $n = 10^5$,

deviation, $d = 0.05$

Result: Relative frequency of robust instances, f

begin

> Determine the original steady state, $x^*(k_{orig})$ via the original kinetic parameter, k_{orig}
>
> **for** $j = 1$ **to** n **do**
>
>> Set random variables: $r_{\{0;1\}} = \text{rand}(\{0;1\})$, $r_{[0;1]} = \text{rand}([0;1])$
>>
>> Set perturbed kinetic parameter: $k_{per} = \left(1 + (-1)^{r_{\{0;1\}}} p \cdot r_{[0;1]}\right) k_{orig}$
>>
>> Determine the perturbed steady state, $x^*(k_{per})$
>>
>> **for** *each element k of* $x^*(k_{per})$ **do**
>>
>>> **if** $|x^*(k_{per}) - x^*(k_{orig})| < d \cdot x^*(k_{orig})$ **then**
>>>
>>>> increase counter, c
>>>
>>> **end**
>>
>> **end**
>
> **end**
>
> $f = \frac{c}{n}$

end

2.6.4. Compliance with data

To support the biological relevance of the predicted steady states, we have performed a comparison with metabolomics of Arabidopsis [88] by using RSS (Equation 2.2):

$$\text{RSS} = \sum_{i=1}^{m} (y_i - x_i^*)^2 \,, \tag{2.2}$$

$$\overline{\text{RSS}} = \frac{1}{m} \sum_{i=1}^{m} (y_i - x_i^*)^2 \,. \tag{2.3}$$

The experimental data were obtained by ion pair reverse-phase chromatography coupled by a triple quadrupole mass spectrometer (IPC-MS/MS) measurements of metabolites of the CBC, starch and sucrose synthesis as well as the involved cofactors. Unfortunately, these data do not consider the compartmentalization and, therefore a recalculation of the data is required. Moreover, the data have to be transformed from $\left[\frac{\text{nmol}}{\text{g FW}}\right]$ to $[\text{mM}]$ to enable the comparison. The underlying assumptions for recalculation and transformation comprise: (1) the concentration of GAP is 5 % of the one of dihydroxyacetone phosphate (DHAP) [92], (2) 1 g fresh weight (FW) corresponds approximately to 1 mg chlorophyll (Chl) [97], (3) PGA, DHAP, fructose-1,6-bisphosphate, F6P, glucose-6-phosphate (G6P) and uridine diphosphate-glucose are compartmentalized as shown for spinach and wheat

[89, 90], (4) GAP is compartmentalized as $DHAP^5$, (5) glucose-1-phosphate as $G6P^1$, and (6) Arabidopsis leaves have the same subcellular volume per mg Chl as shown for spinach [91]. Then, the resulting concentrations (y; Appendix Table A.13) and the predicted steady state concentrations (x^*) are combined to the RSS. Because of the different model boundaries the RSS involves different numbers of metabolites for each model and, therefore we have employed the averaged RSS, $\overline{\text{RSS}}$ (Equation 2.3).

2.6.5. Sensitivity analysis

The similarity of two models is investigated by carrying out a sensitivity analysis comprising Metabolic Control Analysis (MCA) and Kendall rank correlation.

Metabolic Control Analysis

The concept of MCA provides the means to investigate the effects of parameter perturbations on the steady state of a system. There exist three different properties: (1) local effects on single reaction rates due to perturbations of metabolite concentration (elasticity), and (2) global effects on all velocities due to perturbations of metabolite concentration (concentration control) or (3) due to perturbations of enzyme activities (flux control). We investigate the CBC regarding the elasticity (E, Equation 2.4) and flux control coefficients (F, Equation 2.5):

$$\varepsilon_{S_i}^{v_k} = \frac{S_i}{v_k} \frac{\partial v_k}{\partial S_i}, \tag{2.4}$$

$$C_{v_k}^{J_j} = \frac{p_k}{J_j} \frac{\partial J_j}{\partial p_k} \left(\frac{p_k}{v_k} \frac{\partial v_k}{\partial p_k} \right)^{-1}. \tag{2.5}$$

The resulting elasticity and flux control matrices have different sizes due to the different model boundaries. To enable the comparison of pairs of models, we reduce and merge these matrices into seven different sizes (Appendix Table A.14).

Kendall rank correlation

As a result of the MCA, we arrive at six different categories (Appendix Table A.15) to test the similarities using Kendall rank correlation.

The Kendall rank correlation is a measure of the association between two sets of measured quantities. More precisely, the similarity of the orderings of the data sets when ranked by each of the quantities is determined [94]. In our framework, we compare the results of the MCA between each pairs of models, namely, the order of the most influential perturbations of metabolite concentration and enzyme activities, respectively. The seventh category, s_1r_2, includes only one substrate and two reactions which are too few arguments for a Kendall rank correlation. Therefore, at least 6 arguments are required to obtain potential significant statements (note that $6! = 720$, resulting in a p-value of 0.0014). For the other

---

[5]personal communication with Mark Stitt

categories, each matrix is converted into a row vector, w, by appending the rows of the matrix (Equation 2.7):

$$M = (a_{ij})_{i=1,\ldots,m;\; j=1,\ldots,n} \qquad\qquad M \in \mathcal{K}^{\,m \times n}, \qquad (2.6)$$

$$w = (a_{11} \ldots a_{1n} \;\ldots\; a_{m1} \ldots a_{mn}) \qquad\qquad w \in \mathcal{K}^{\,m \cdot n}. \qquad (2.7)$$

The similarity of two vectors by the Kendall's τ is defined as the difference of concordant and discordant pairs divided by the total number of pairs (Equation 2.8):

$$\tau = \frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\text{total number of pairs}}, \qquad (2.8)$$

$$= P\,(\text{same order}) - P\,(\text{different order})\,. \qquad (2.9)$$

Therefore, larger (positive) τ's represent greater similarity whereas negative τ's stand for oppositional behavior. For the six cases and two MCA coefficients, the results are arranged in twelve matrices (Appendix Tables A.16–A.26) which are visualized via heatmaps (Appendix Figures A.1 and A.2). Moreover, for the six categories the Kendall τ coefficients for the same pairs of models are summed up to determine the overall $\bar{\tau}_E$, $\bar{\tau}_F$, and $\bar{\tau}_{\text{av}}$, respectively (Equation 2.10):

$$\bar{\tau}_{\text{av}} = \frac{\bar{\tau}_E + \bar{\tau}_F}{2}, \qquad \bar{\tau}_{E/F} = \frac{1}{n} \sum_{j=1}^{n} \tau_{E/F\,j}. \qquad (2.10)$$

Because not each pair of models appears in the same number of categories, the average is considered (Appendix Table A.27). The largest $\bar{\tau}$ for a given model, then, determines the most similar model. For the final ranking, the most similar model with respect to $\bar{\tau}_{\text{av}}$ is chosen as local and global effects are both important.

Chapter 3.

Integrating enzyme action and stoichiometric modeling: The case of the Calvin-Benson cycle and related end-product processes

3.1. Background

Ever since the seminal work of Sumner (1926), enzymatic proteins are known as the primary catalysts of chemical reactions in biological systems. Enzymes accelerate the rate of biochemical reactions without being consumed by providing an alternative route of lower activation energy. They are usually selective, *i.e.*, a small number of highly specific substrate molecules fit in the so-called active site of an enzyme. Protein complexes, acting as enzymes, usually include more than one active site, so that they are able to catalyze more than one reaction simultaneously. To achieve full biochemical activity, many enzymes require one or several chemical, non-protein compounds, called cofactors. The probably most thoroughly studied cofactors are adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide (phosphate; NAD(P)H), so-called coenzymes which bind loosely to the enzyme to facilitate the progression of energetically unfavorable reactions. Upon binding, ATP and NAD(P)H release energy stored in energy-rich covalent bonds and high-energy electrons, respectively. The vast majority of biochemical reactions are enzymatically driven which is why, often, the enzyme name is used as a synonym for the catalyzed reaction. This is also the reason why enzymes are commonly understood to enable metabolic functionality.

Modeling the kinetics of biochemical processes allows the incorporation of the enzymatic properties for the respective reactions via kinetic parameters. In mass-action kinetics, this is taken into account indirectly by the reaction rate constants, k. On the other hand, the enzymatic properties, such as concentration, E, and maximum reaction rate, V_{max}, respectively, are directly incorporated in Michaelis-Menten kinetics via $V_{max} = k_{cat}E$, where k_{cat} denotes the turnover rate. However, the incorporation of kinetic information for larger systems is often problematic due to insufficient knowledge and/or the lack of experimental data. An attempt to overcome this issue is the application of parameter estimation via optimization which is, unfortunately, very computationally intensive for medium- and large-scale models [98].

Alternatively, stoichiometric modeling can be utilized for *in silico* investigations of the modeled system. Assuming steady-state conditions, stoichiometric modeling enables the examination of the biochemical reaction solely based on stoichiometry and reaction directionality. Nevertheless, this simplification is in line with the predictions from kinetic modeling, since any solution of a kinetic model is also a solution of the corresponding

stoichiometric model with proper flux boundaries.

In order to incorporate the involved enzyme of a biochemical reaction in stoichiometric modeling, we examined the system on the substrate-enzyme-product level of reaction modeling. Traditionally, stoichiometric models cannot capture the involved enzymes of a biochemical reactions as the net stoichiometry of each reactant is taken into account ($S + \cancel{E} \rightarrow P + \cancel{E}$). Here, we present a novel approach to circumvent this obstacle of integrating enzymes and, accordingly, to enable the consideration of this key element of biochemical systems. This allows us to incorporate enzyme-specific turnover rates such that we account at least partially for the dynamic features of enzymes rather than to consider them as static pools of infinite capacity.

This approach also enables the investigation of the impact of the enzyme synthesis costs. In accordance with Craig and Weber [99], the metabolic costs of the enzymes are estimated as the energy demand, in terms of ATP, to synthesize the required amino acids and to assemble the complete polypeptide sequence including post-translational processing. Thus, we were able to extend the current concept of metabolic costs.

The metabolic costs, in terms of energy, usually account only for the energy demand of the synthesis process itself, *i.e.*, interconverting the provided biochemical precursor(s) into the desired metabolite (composition). However, metabolic costs also comprise the maintenance costs for the employed underlying machinery, enzymes synthesis and degradation [100]. Thereby, assuming steady-state or even homeostatic conditions implicates the maintenance of the levels of the involved enzyme, which corresponds to an ascertained flux distribution in a conventional stoichiometric model. Pursuing the extended stoichiometric modeling approach additionally constrains the respective included enzyme turnover and synthesis rates to be equal.

Here, we show that integrating enzymes and stoichiometric modeling allows to amend the metabolic costs reinforcing biological findings. Illustrated on the case of a small-scale stoichiometric model of the Calvin-Benson cycle (CBC) which also includes the end-product processes of starch and sucrose synthesis, we compare the metabolic synthesis costs of the two end-products, in terms of ATP requirements, according to the traditional and the extended concept.

3.2. Small-scale Calvin-Benson cycle model

To arrive at an accurate stoichiometric model of the CBC, we adapted a kinetic model examined in the comparative analysis and ranking of CBC models from Arnold and Nikoloski (2011) [46] (see Chapter 2). We utilized the model of Poolman *et al.* [60], the best-performing model which comprises a detailed description of the CBC and is, additionally, suitable for metabolic engineering applications [46]. By converting the original kinetic model, we introduced a few metabolites, namely, water (H_2O), adenosine diphosphate glucose (ADPG), and two artificial starch metabolites comprising one and two glucose units, respectively. In this fashion, we not only ensured that the model is mass-balanced, but also enabled export of a starch dimer via an extension of the starch synthesis pathway. Moreover, we included the biochemical reactions and metabolites according to the well-investigated sucrose synthesis pathway (Figure 3.1). It is noteworthy that the ADPG pyrophosphorylase (AGPase) and the uridine triphosphate glucose-1-phosphate uridylyltransferase (UGPase) reactions are

both provided as merged reactions comprising the inorganic diphosphatase (EC 3.6.1.1) converting pyrophosphate and H_2O to inorganic phosphate (Table 3.1, highlighted by *). As a result, we obtain a stoichiometric model comprising 43 metabolites and 43 biochemical reactions of which seven are import and export reactions.
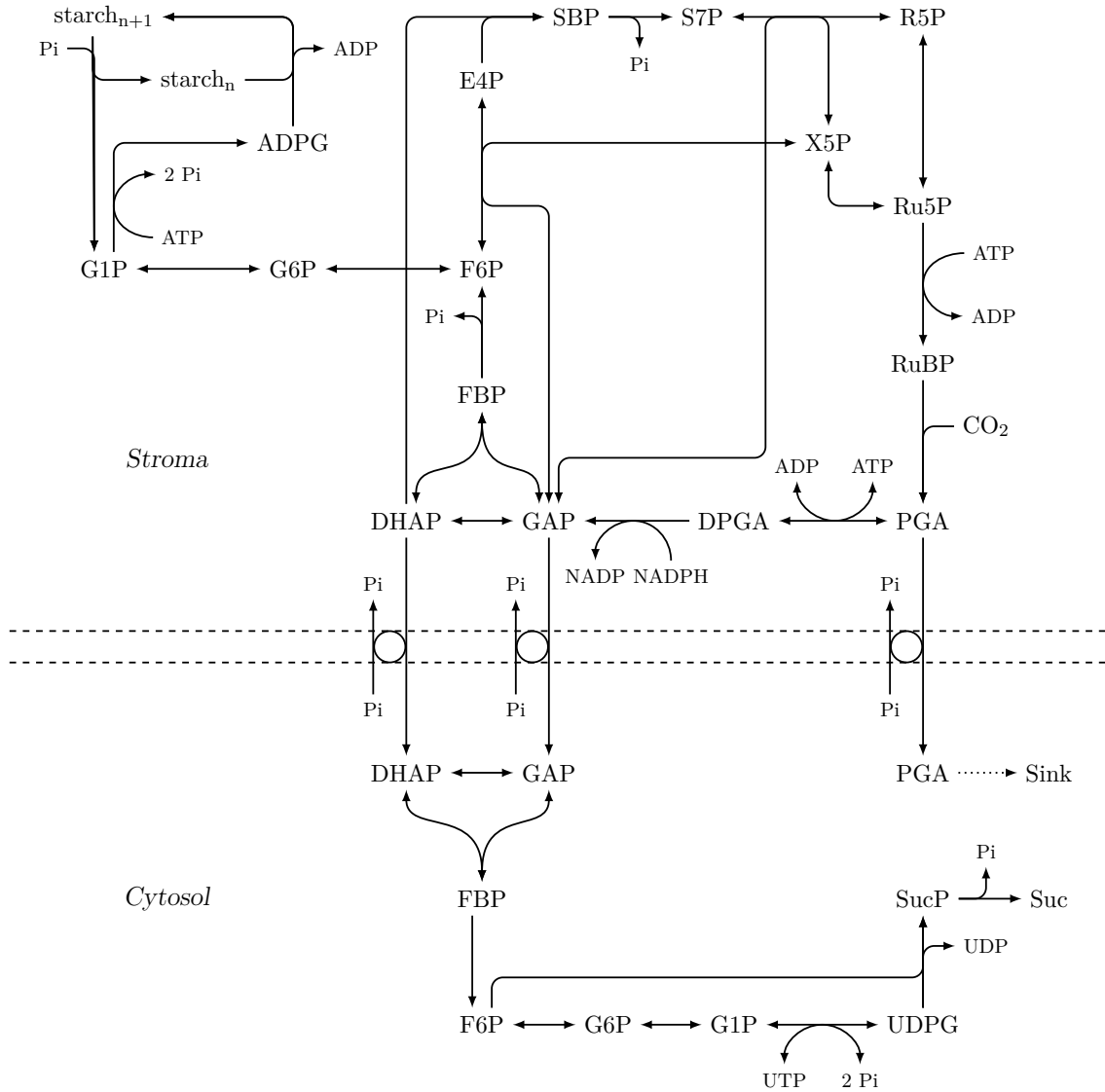
Figure 3.1.: Schematic representation of the small-scale stoichiometric model of the CBC and associated end-product processes of starch and sucrose synthesis.

Table 3.1.: Reaction list according to AraCyc [32] of the small-scale stoichiometric model of the CBC and associated end-product processes of starch and sucrose synthesis. The reactions denoted by * are merged reactions comprising the inorganic diphosphatase (EC 3.6.1.1) converting pyrophosphate and H_2O to Pi.

Calvin-Benson cycle (Chloroplast)	
$RuBP + CO_2 + H_2O \rightarrow 2\,PGA$	RuBisCO (EC 4.1.1.39)
$PGA + ATP \rightleftharpoons DPGA + ADP$	PGA kinase (EC 2.7.2.3)
$DPGA + NADPH \rightarrow GAP + NADP + Pi$	GAP dehydrogenase (EC 1.2.1.13)
$GAP \rightleftharpoons DHAP$	TP isomerase (EC 5.3.1.1)
$GAP + DHAP \rightleftharpoons FBP$	FBP aldolase (EC 4.1.2.13)
$FBP + H_2O \rightarrow F6P + Pi$	FBPase (EC 3.1.3.11)
$GAP + F6P \rightleftharpoons E4P + X5P$	F6P transketolase (EC 2.2.1.1)
$DHAP + E4P \rightarrow SBP$	SBP aldolase (EC 4.1.2.13)
$SBP + H_2O \rightarrow S7P + Pi$	SBPase (EC 3.1.3.37)
$GAP + S7P \rightleftharpoons R5P + X5P$	S7P transketolase (EC 2.2.1.1)
$X5P \rightleftharpoons Ru5P$	Ru5P epimerase (EC 5.1.3.1)
$R5P \rightleftharpoons Ru5P$	R5P isomerase (EC 5.3.1.6)
$Ru5P + ATP \rightarrow RuBP + ADP$	Phosphoribulokinase (EC 2.7.1.19)
Starch synthesis and degradation (Chloroplast)	
$F6P \rightleftharpoons G6P$	G6P isomerase (EC 5.3.1.9)
$G6P \rightleftharpoons G1P$	Phosphoglucomutase (EC 5.4.2.2)
$G1P + ATP + H_2O \rightarrow ADPG + 2\,Pi$	ADPG pyrophosphorylase (EC 2.7.7.27)*
$ADPG \rightarrow starch_1 + ADP$	starch synthase (EC 2.4.1.21)
$ADPG + starch_1 \rightarrow starch_2 + ADP$	starch synthase (EC 2.4.1.21)
$starch_2 + Pi \rightarrow starch_1 + G1P$	(starch) phosphorylase (EC 2.4.1.1)
$starch_1 + Pi \rightarrow G1P$	(starch) phosphorylase (EC 2.4.1.1)
Sucrose synthesis (Cytosol)	
$GAP \rightleftharpoons DHAP$	TP isomerase (EC 5.3.1.1)
$GAP + DHAP \rightleftharpoons FBP$	FBP aldolase (EC 4.1.2.13)
$FBP + H_2O \rightarrow F6P + Pi$	FBPase (EC 3.1.3.11)
$F6P \rightleftharpoons G6P$	G6P isomerase (EC 5.3.1.9)
$G6P \rightleftharpoons G1P$	Phosphoglucomutase (EC 5.4.2.2)
$G1P + UTP + H_2O \rightleftharpoons UDPG + 2\,Pi$	UDPG pyrophosphorylase (EC 2.7.7.9)*
$F6P + UDPG \rightarrow SucP + UDP$	SucP synthase (EC 2.4.1.14)
$SucP \rightarrow Suc + Pi$	SucP phosphorylase (EC 3.1.3.24)
energy transformers (Chloroplast,Cytosol)	
$ADP + Pi \rightarrow ATP + H_2O$	ATP synthase (EC 3.6.3.14)
$NADP \rightarrow NADPH$	Fd-NADP reductase (EC 1.18.1.2)
$UTP + ADP \rightleftharpoons UDP + ATP$	NDP kinase (EC 2.7.4.6)
Transport reactions	
$PGA + Pi_c \rightarrow PGA_c + Pi$	TP translocator (Fischer, 2011)
$GAP + Pi_c \rightarrow GAP_c + Pi$	TP translocator (Fischer, 2011)
$DHAP + Pi_c \rightarrow DHAP_c + Pi$	TP translocator (Fischer, 2011)
$ATP_c + ADP + Pi \rightleftharpoons ATP + ADP_c + Pi_c$	NTP transporter (Fischer, 2011)
$H_2O \rightarrow H_2O_c$	Diffusion

3.3. The extended stoichiometric model

The consideration of enzyme synthesis and degradation allows the quantification of metabolic costs of building blocks with respect to the *de novo* synthesis costs of the catalyzing enzymes. For this purpose, we first incorporated the enzymes as internal metabolites of the stoichiometric metabolic network. As the stoichiometric matrix comprises only the net stoichiometric coefficients of the underlying reaction equations, we have to distinguish between the portion of the enzyme that catalyzes a reaction, E (termed "active enzyme"), and the released portion of the enzyme after catalyzing the reaction, E' (termed "inactive enzyme"; Figure 3.2). Accordingly, reversible reactions are split into two reactions, each with one active enzyme (Figure 3.2B). Reactivation of an inactive enzyme is achieved by including a corresponding regeneration reaction specific to the considered enzyme. Moreover, based on the distinction of active and inactive enzyme, we defined that enzyme synthesis, E_{syn}, results only in active enzyme, and the degradation, E_{deg}, acts only on the inactive enzyme.

Accordingly, for each biochemical reaction, three to four enzyme-related reactions are incorporated, depending on the reaction reversibility. Implementing these modifications and additions results in a extended stoichiometric model including 97 metabolites and 140 reactions.

A

$$E + S \xrightarrow{v} E' + P$$

$$E_{\mathrm{syn}} \longrightarrow \qquad \longrightarrow E_{\mathrm{deg}}$$
$$v_{\mathrm{deg}} = t \cdot v$$

B

$$E + S \underset{}{\overset{v}{\rightleftharpoons}} E + P$$
$$E \; + S \xrightarrow{v_f} \; E' + P$$
$$\; + P \xrightarrow{v_r} \; + S$$

$$E_{\mathrm{syn}} \longrightarrow \qquad \longrightarrow E_{\mathrm{deg}}$$
$$v_{\mathrm{deg}} = t \left(v_f + v_r \right)$$
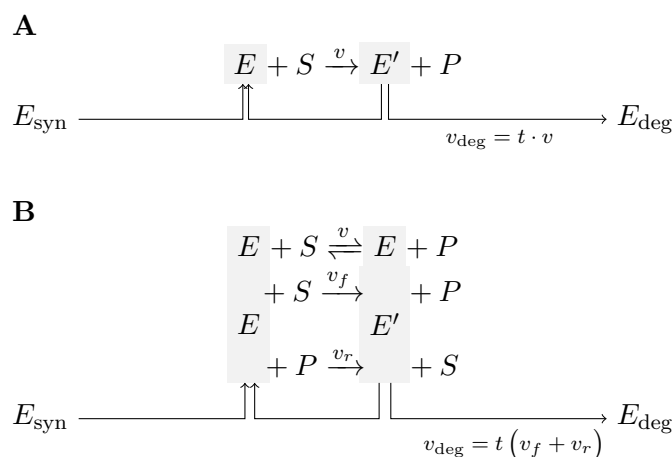
Figure 3.2.: Reaction schemes of an irreversible biochemical reaction (A) and a reversible reaction (B) including enzyme synthesis, E_{syn}, and degradation, E_{deg}, as well as regeneration reaction.

3.3.1. Enzyme synthesis

Enzyme synthesis and the corresponding synthesis costs, in terms of ATP, were modeled as an ATP hydrolyzing import of the active enzyme (Equation 3.1):

$$x \text{ ATP} + x \text{ H}_2\text{O} \rightarrow E + x \text{ ADP} + x \text{ Pi}, \tag{3.1}$$
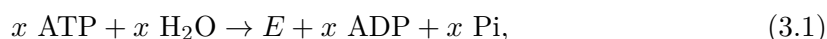
where x is the energy demand of the *de novo* synthesis of the catalyzing enzyme. The metabolic costs of each protein can be decomposed into: (1) the biosynthesis costs of the protein building blocks, namely, amino acids, and (2) the assembly and maturation costs of a protein. For the cost determination, we first had to ascertain the subunit structure as well as the encoding genes of each enzyme (Tables 3.2).

Table 3.2.: Subunit structure, gene-encoding and EC number of the involved, catalyzing enzymes. The slanted genes encode a non-catalytic subunit.

Enzyme	EC	Structure	Genes	
RuBisCO	4.1.1.39	$[RbcL]_8[RbcSx]_8$	RbcL	ATCG00490
			RbcS1A	AT1G67090
			RbcS1B	AT5G38430
			RbcS2B	AT5G38420
			RbcS3B	AT5G38410
PGA kinase	2.7.2.3	$[PGKx]$	PGK1	AT3G12780
			PGK2	AT1G56190
GAP dehydrogenase	1.2.1.13	$[GapA\text{-}x]_4$ OR $[GapA\text{-}x]_2[GapB]_2$	GapA-1	AT3G26650
			GapA-2	AT1G12900
			GapB	AT1G42970
TP isomerase	5.3.1.1	$[TIM]$	TIM	AT2G21170
Aldolase	4.1.2.13	$[SFBA\text{-}x]_4$	SFBA-1	AT2G21330
			SFBA-2	AT4G38970
			SFBA-3	AT2G01140
FBPase	3.1.3.11	$[FBP]_4$	FBP	AT3G54050
Transketolase	2.2.1.1	$[TKL\text{-}x]$	TKL-1	AT3G60750
			TKL-2	AT2G45290
SBPase	3.1.3.37	$[SBP]_2$	SBP	AT3G55800
Ru5P epimerase	5.1.3.1	$[RPE]$	RPE	AT5G61410
R5P isomerase	5.3.1.6	$[PRI]$	PRI	AT3G04790
Phosphoribulokinase	2.7.1.19	$[PRK]_2$	PRK	AT1G32060
G6P isomerase	5.3.1.9	$[PGI]_2$	PGI	AT4G24620
Phosphoglucomutase	5.4.2.2	$[PGM1]$	PGM1	AT5G51820
ADPG pyrophosphorylase	2.7.7.27	$[ApLx]_2[ApSx]_2$	ApL1	AT5G19220
			ApL2	AT1G27680
			ApL3	AT4G39210
			ApL4	AT2G21590
			ApS1	AT5G48300
			ApS2	AT1G05610
Starch synthase	2.4.1.21	$[SSx]$	SS1	AT5G24300
			SS2	AT3G01180
			SS3	AT1G11720
			SS4	AT4G18240
(Starch) Phosphorylase	2.4.1.1	$[PHSx]$	PHS1	AT3G29320
			PHS2	AT3G46970

Table 3.2: Subunit structure, gene-encoding and EC number of the involved, catalyzing enzymes – continued

Enzyme	EC	Structure	Genes	
ATP synthase	3.6.3.14	$[ATPA]_3[ATPB]_3$	ATPA	ATCG00120
		$[ATPC1][ATPD]$	ATPB	ATCG00480
		$[ATPE][ATPF]$	ATPC1	AT4G04640
		$[ATPG][ATPH]_{14}$	ATPD	AT4G09650
		$[ATPI]$	ATPE	ATCG00470
			ATPF	ATCG00130
			ATPG	AT4G32260
			ATPH	ATCG00140
			ATPI	ATCG00150
Fd-NADP reductase	1.18.1.2	[FNR-1][FNR-2]	FNR-1	AT5G66190
			FNR-2	AT1G20020
cyt TP isomerase	5.3.1.1	[CTIM]	CTIM	AT3G55440
cyt aldolase	4.1.2.13	$[SFBA-7]_4$	SFBA-7	AT4G26520
cyt FBPase	3.1.3.11	$[CFBP]_4$	CFBP	AT1G43670
cyt G6P isomerase	5.3.1.9	$[PGIC]_2$	PGIC	AT5G42740
cyt phosphoglucomutase	5.4.2.2	[PGMx]	PGM2	AT1G70730
			PGM3	AT1G23190
UDPG pyrophosphorylase	2.7.7.9	[UGPx]	UGP1	AT3G03250
			UGP2	AT5G17310
SucP synthase	2.4.1.14	$[SPSxF]_2$ OR $[SPSxF]_4$	SPS1F	AT5G20280
			SPS2F	AT5G11110
			SPS3F	AT1G04920
			SPS4F	AT4G10120
SucP phosphatase	3.1.3.24	$[SPPx]_2$	SPP1	AT1G51420
			SPP2	AT3G52340
			SPPa	AT2G35840
			SPPb	AT3G54270
NDP kinase	2.7.4.6	$[NDPKx]_6$	NDPK1	AT4G09320
			NDPK2	AT5G63310
			NDPK3	AT4G11010
			NDPK4	AT4G23900

Amino acid biosynthesis costs

The synthesis of the amino acids, as protein building blocks, represents the first part of the enzyme costs. To this end, we first had to determine the underlying amino acid composition of each enzyme such that we could calculate, subsequently, their individual costs. In analogy to previous studies [101–103], we approximated the costs based on large-scale metabolic models comprising the various and intricate synthesis pathways of the different amino acids. We focused on growing leaf cells assuming that energy efficiency is an adequate objective of these cells, *i.e.*, the synthesis pathways requiring the least amount of energy, in terms of ATP, are preferably utilized. The consequent optimizations were solved by means of

flux balance analysis obtaining a minimum cost estimation for the amino acids. Here, we pursued two approaches based on the studies of (1) Sajitz-Hermstein and Nikoloski [101], and (2) Arnold and Nikoloski (2014) [38] (see Chapter 5).

Most of the enzymes are not uniquely defined by a single amino acid sequence, there often exist isozymes with different subunit structures and/or encoding genes, and isoforms of genes, *i.e.*, different splicing forms. In order to reinforce the biological plausibility of the costs, we decided to select a single representative polypeptide sequence for each case rather than to average over all possible combinations resulting in an artificial sequence. Therefore, we accounted for all possible amino acid compositions, A, and, subsequently, determined the representative using of the medoid (for Euclidean distance). A medoid determines the representative(s) of a data set, r^{dist}, with the minimum average dissimilarity to all other candidates (Equation 3.2). Here, we described the dissimilarity by the euclidean distance of the relative amino acid frequency of two amino acid compositions, x and y (Equation 3.3):

$$r^{\text{euc}} = \arg\min_{x \in A} \frac{1}{|A|} \sum_{y=1}^{|A|} d(x, y)\,, \tag{3.2}$$

$$d(x, y) = \sqrt{\sum_{i=1}^{20} \left(\frac{n_i^x}{\sum_j n_j^x} - \frac{n_i^y}{\sum_j n_j^y} \right)^2}\,, \tag{3.3}$$

where $|A|$ is the cardinality of set A, n_i^x is number of amino acid i in amino acid composition x. In doing so, we obtained one representative polypeptide sequence for each subunit for the catalyzing enzymes and, accordingly, a unique amino acid composition for each enzyme (Table 3.3).

In order to assign the amino acid biosynthesis costs of each involved enzyme, we integrated the respective unique amino acid composition and a metabolic costs determination approach. In the study of Sajitz-Hermstein and Nikoloski [101], the costs for synthesizing the amino acids individually were obtained based on the Arabidopsis metabolic network of Poolman *et al.* [104] (Table 3.4). These single amino acid costs were multiplied with the respective amino acid composition resulting in the enzyme-specific building block synthesis costs. We emphasize that the model of Poolman *et al.* [104] simulates heterotrophic growth conditions and, moreover, the consequent costs refer to day scenario. In contrast, the Arabidopsis core model of Arnold and Nikoloski (2014) [38] (see Chapter 5) represents photoautotrophic growth conditions matching the here investigated small-scale model of the CBC and related end-product synthesis processes. Based on the Arabidopsis core model, we utilized the respective amino acid composition as objective function to calculate the energy demand of each enzyme, in terms of ATP. More precisely, we incorporated the number of each amino acid as stoichiometric coefficients of the biomass reaction `Bio_AA` and determined the minimum ATP consumption while synthesizing one unit of the amino acid composition using flux balance analysis (Algorithm B.1). The Arabidopsis core model and the corresponding minimum ATP consumption algorithm are detailed in Chapter 5.

Table 3.3.: Amino acid composition for the representative polypeptide sequences (AGI code) of the involved enzymes identified by the medoid from the Euclidean distance.

EC		AGI code	Medoid	Ala	Arg	Asn	Asp	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
4.1.1.39	L	ATCG00490.1	–	43	27	15	26	9	35	12	47	15	21	41	24	8	22	24	19	33	8	17	33
	S	AT5G38430.1	0.1664	13	7	6	8	5	11	4	11	2	9	12	13	5	10	12	15	13	4	8	13
2.7.2.3		AT1G56190.1	0.0565	52	16	13	25	2	27	4	46	7	25	50	38	6	19	21	40	29	2	4	52
1.2.1.13	A	AT1G12900.3	0.1344	30	17	17	30	5	11	6	31	6	22	29	25	5	9	12	24	20	5	7	39
		AT1G12900.4	0.1344	30	17	17	30	5	11	6	31	6	22	29	25	5	9	12	24	20	5	7	39
	B	AT1G42970.1	–	47	19	18	31	8	20	9	36	9	25	34	29	9	13	20	33	26	5	7	49
5.3.1.1		AT2G21170.1	0.01	33	12	10	15	4	19	8	31	6	18	19	23	2	13	10	37	12	6	4	33
		AT2G21170.2	0.01	33	12	9	15	3	19	8	29	6	18	19	22	2	11	10	37	12	5	4	32
4.1.2.13		AT2G21330.1	0.146	51	16	15	14	3	26	17	32	4	17	38	22	11	8	19	32	23	7	15	29
3.1.3.11		AT3G54050.1	1	32	18	17	23	8	26	16	37	6	27	35	21	8	11	16	46	20	2	19	29
		AT3G54050.2	1	32	18	17	23	8	26	16	37	6	27	35	21	8	11	16	46	20	2	19	29
2.2.1.1		AT3G60750.1	0.0151	80	32	26	36	8	50	17	68	21	41	58	41	14	28	39	63	43	11	22	43
		AT3G60750.2	0.0151	80	32	26	36	8	49	17	68	21	41	58	41	14	28	39	63	43	11	22	43
3.1.3.37		AT3G55800.1	–	28	14	11	19	7	24	11	37	4	20	37	27	8	18	17	37	32	2	13	27
5.1.3.1		AT5G61410.1	1	28	13	9	15	5	12	9	21	6	20	21	16	6	10	15	26	10	2	4	33
		AT5G61410.2	1	28	13	9	15	5	12	9	21	6	20	21	16	6	10	15	26	10	2	4	33
5.3.1.6		AT3G04790.1	–	24	10	6	19	2	14	7	28	3	15	31	20	6	12	10	23	15	1	4	26
2.7.1.19		AT1G32060.1	–	25	20	15	29	5	23	16	23	8	27	37	25	8	19	20	29	25	2	16	23
5.3.1.9		AT4G24620.1	0.0106	59	28	26	34	7	38	22	46	11	35	66	31	14	21	34	48	30	8	19	36
5.3.1.9		AT4G24620.2	0.0106	55	26	23	31	5	35	22	43	10	29	63	30	12	21	33	45	30	8	16	33
5.4.2.2		AT5G51820.1	–	48	25	24	47	2	30	11	57	9	43	45	41	11	37	31	58	32	6	23	43
2.4.1.21		AT1G11720.2	0.2627	80	60	56	72	10	81	39	73	30	52	95	82	16	55	50	61	45	24	31	82
2.4.1.1		AT3G29320.1	0.0339	74	49	46	53	10	87	27	58	11	54	76	70	22	45	43	67	47	18	32	73
		AT3G46970.1	0.0339	65	43	40	51	7	62	33	54	14	44	82	58	17	36	38	50	45	18	30	54
3.6.3.14	A	ATCG00120.1	–	48	29	16	18	1	39	38	41	2	46	51	19	11	12	16	36	31	0	18	35
	B	ATCG00480.1	–	40	27	22	23	1	36	19	47	4	30	49	24	15	18	26	26	36	0	12	43
	C	AT4G04640.1	–	33	20	14	19	5	26	13	15	1	18	42	27	6	13	18	37	23	1	8	34
	D	AT4G09650.1	–	23	10	12	11	0	13	16	8	1	16	25	17	5	9	7	21	16	0	3	21
	E	ATCG00470.1	–	16	9	12	5	1	10	6	7	1	15	15	5	3	1	3	4	10	2	0	7
	F	ATCG00130.1	–	9	13	20	6	1	15	10	12	2	14	22	9	3	10	2	9	13	1	4	9
	G	AT4G32260.1	–	32	6	6	9	0	21	10	2	0	12	26	23	8	8	10	24	8	0	2	12
	H	ATCG00140.1	–	16	2	2	0	0	4	3	11	0	5	12	1	2	3	4	4	3	0	1	8
	I	ATCG00150.1	–	17	3	9	5	1	12	11	20	4	25	34	7	3	18	14	16	14	5	8	23
1.18.1.2	1	AT5G66190.1	0.0405	19	12	13	18	6	30	8	27	3	20	28	39	11	15	17	27	22	6	12	27
		AT5G66190.2	0.0405	11	9	10	15	4	23	5	23	2	14	25	27	11	12	13	16	12	5	10	15
	2	AT1G20020.3	0.0153	29	12	18	21	6	22	10	26	4	17	25	37	13	16	16	24	26	6	15	26

Table 3.4.: Published amino acid synthesis costs, in terms of ATP, of Sajitz-Hermstein and Nikoloski [101] for heterotrophic growth conditions during the day.

Amino acid	*Ala*	*Arg*	*Asn*	*Asp*	*Cys*	*Glu*	*Gln*	*Gly*	*His*	*Ile*
Costs [ATP]	19.33	41.32	26.66	25.33	19.33	31.33	32.66	13.33	39.99	37.33

Amino acid	*Leu*	*Lys*	*Met*	*Phe*	*Pro*	*Ser*	*Thr*	*Trp*	*Tyr*	*Val*
Costs [ATP]	37.33	38.66	31.33	55.33	31.33	19.33	25.33	68.66	55.33	31.33

Protein assembly costs[6]

While the costs for the amino acid biosynthesis differ for the underlying biochemical pathways and, consequently, vary over proteins with different amino acid composition, the costs for the protein assembly can be approximated per amino acid or polypeptide. Besides the actual protein assembly, we took into account costs arising from protein maturation.

The protein assembly or biosynthesis comprises four energy demanding steps: (1) the amino acid activation, (2) the formation of the initiation complex, (3) the amino-acid-sequence elongation, and (4) the finalizing of the polypeptide within the termination phase. Before an amino acid is capable of being included into protein biosynthesis, it has to be activated by forming the corresponding aminoacyl-transfer RNA (tRNA). This activation process requires one molecule of ATP and releases one molecule of adenosine monophosphate (AMP) and pyrophosphate, respectively. The recycling of AMP into ADP requires a second molecule of ATP, so that, for convenience, the cost can be estimated as two molecules ATP (Formulas 3.4–3.6, Table 3.5):

$$\text{aa} + \text{ATP} + \text{tRNA} \rightleftharpoons \text{aa-tRNA} + \text{AMP} + \text{PPi}, \tag{3.4}$$

$$\text{AMP} + \text{ATP} \rightleftharpoons 2\,\text{ADP}, \tag{3.5}$$

$$\text{PPi} + \text{H}_2\text{O} \rightarrow 2\,\text{Pi}. \tag{3.6}$$
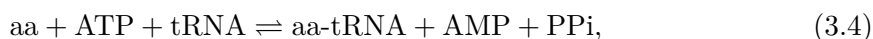
The actual biosynthesis starts with the formation of the initiation complex. Thereby, first the small ribosome subunit, the translation initiation factors and the ternary complex including one molecule guanosine triphosphate (GTP) form the 43S preinitiation complex [105]. By means of further translation initiation factors the messenger RNA (mRNA) is ATP-dependent activated and, subsequently, attached to build the 48S initiation complex. Finally, the large ribosome subunit and a further initiation factor comprising a further GTP are added to form the 80S initiation complex whereby the associated initiation factors are displaced [105]. In the end, both GTP are hydrolyzed and released and the nucleoside-diphosphate kinase (EC 2.7.4.6) catalyzes the recycling of GTP via ATP such that the costs for the initiation can be approximated by three molecules ATP per polypeptide (Formulas 3.7 and 3.8, Table 3.5):

---

[6]Based on the supplemental material of Arnold and Nikoloski (2014) [38] (Chapter 5).

$$40S + 60S + \text{mRNA} + \text{eIFs} + 2\,\text{GTP} + \text{ATP} \rightarrow 80S + \text{eIFs} + 2\,\text{GDP} + \text{ADP} + 3\,\text{Pi}, \quad (3.7)$$
$$2\,\text{GDP} + 2\,\text{ATP} \rightleftharpoons 2\,\text{GTP} + 2\,\text{ADP}. \quad (3.8)$$
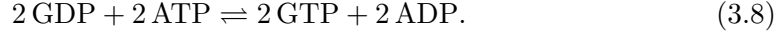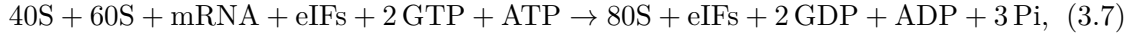
During an elongation cycle, first an elongation factor binds the aminoacyl-tRNA forming the ternary complex and, subsequently, releases the aminoacyl-tRNA into the so-called aminoacyl (A) site of the ribosome and catalyzes thereby the peptide bond formation. Secondly, another elongation factor catalyzes the translocation of the peptidyl-tRNA into the peptidyl (P) site while the deacylated tRNA is pushed to the exit (E) site. The application of both elongation factors requires one molecule of GTP released as guanosine diphosphate (GDP), respectively. By means of the nucleoside-diphosphate kinase the cost of the elongation cycle can be estimated with two molecules ATP per peptide bound (Formulas 3.9 and 3.10, Table 3.5):

$$80S\text{–aa}_\text{n} + \text{aa-tRNA} + \text{eEFs} + 2\,\text{GTP} \rightarrow 80S\text{–aa}_{\text{n+1}} + \text{eEFs} + 2\,\text{GDP} + 2\,\text{Pi}, \quad (3.9)$$
$$2\,\text{ATP} + 2\,\text{GDP} \rightleftharpoons 2\,\text{ADP} + 2\,\text{GTP}. \quad (3.10)$$

The finalizing release of the polypeptide sequence in the termination phase requires the binding of a termination factor. Again, this factor is associated with a molecule of GTP which is hydrolyzed before the factor is released. By recycling the GDP to GTP, the costs for the termination phase is given by one molecule ATP per polypeptide (Formulas 3.11 and 3.12, Table 3.5):

$$80S\text{–aa}_\text{N} + \text{eRF} + \text{GTP} \rightarrow \text{post-termination complex} + \text{peptide} + \text{eRF} + \text{GDP} + \text{Pi},$$
$$(3.11)$$
$$\text{ATP} + \text{GDP} \rightleftharpoons \text{ADP} + \text{GTP}. \quad (3.12)$$
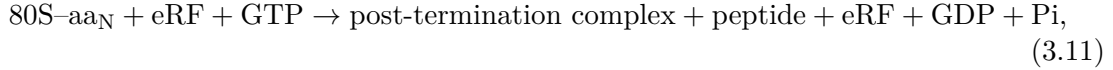
Furthermore, we accounted for the costs of protein maturation including the energy demand for maintenance processes, the secretion of signaling sequences and post-translational modifications. Altogether, these processes consume approximately one molecule ATP per amino acid as specified in Table 3.5. Consequently, the overall costs for the protein assembly can be estimated as five molecules of ATP hydrolyzed per amino acid of synthesized protein.

3.3.2. Enzyme degradation

Biologically determined enzyme turnover rates describe the portion of degraded protein per time. However, as stoichiometric models take into account the stoichiometry of the internal metabolites rather than the respective concentrations, we defined the enzyme degradation as degraded protein per catalyzed reaction. In order to reinforce the biological relevance, we estimated the turnover based on the turnover number, k_cat, and the half-life, $T_{1/2}$, of each enzyme. In doing so, we determined the maximum number of catalyzed reactions, k_max, as the product of half-life and turnover number, and the turnover rate per reaction

Table 3.5.: Itemization of the protein assembly costs comprising the protein biosynthesis and the maturation.

Process	Costs [ATP]		Ref
Protein biosynthesis			
Amino acid activation	2	per amino acid	[106]
Initiation	3	per polypeptide	[105]
Elongation – peptide bond formation and translocation	2	per peptide bond	[106]
Termination	1	per polypeptide	[106]
Tool maintenance	0.16	per amino acid	[106]
Signal sequences – $0.7 \cdot 0.18 \cdot$ (costs from biosynthesis and biodegradation of polypeptide)	0.65	per amino acid	[106]
Amino acid turnover	0	per peptide bond	[107]
Post-translational processing			
Methylation, acetylation, etc.	0.1	per peptide bond	[107]
Phosphorylation, *e.g.* enzyme (de)activation	0.1	per peptide bond	[107]
Protein biodegradation	1	per peptide bond	[106]
Total	≈ 5	per amino acid	[108]

by considering the respective reciprocal value (Equation 3.13):

$$k_{\text{deg}} = \frac{1}{k_{\text{max}}}, \qquad k_{\text{max}} = k_{\text{cat}} \cdot T_{1/2}, \tag{3.13}$$

The turnover numbers of the considered enzymes were extracted from the Braunschweig Enzyme Database (BRENDA), with preference to values specific to Arabidopsis or other plants (Table 3.6). For the half-lives, we chose a data set of *Chlamydomonas reinhardtii* (Chlamydomonas) from Mastrobuoni *et al.* [109], due to its extensive coverage of the catalyzing enzymes and the deficiency of comparable data for Arabidopsis. Mastrobuoni and colleagues identified more than 2500 proteins and measured the ratio of protein labeled with heavy amino acids and light amino acids at different time points. Assuming exponential cell growth, an exponential decay of the light protein, they estimated the half-life of the proteins [109].

As a result, the enzyme turnover rates are coupled with the number of catalyzed reactions. However, as several enzymes, in particular complexes, comprise more than one active sites, we had to account for the number of active sites in the catalyzing reaction. For this purpose, we included the reciprocal value of the number of active sites as the stoichiometric coefficient of the enzyme (Equation 3.14):

$$\frac{1}{c} E + S \rightarrow \frac{1}{c} E' + P. \tag{3.14}$$

Table 3.6.: Enzyme degradation rates per catalyzed reaction and the corresponding half-lives from Chlamydomonas [109] and the turnover numbers from given organisms.

Enzyme	Half-life time [h]		Turnover number $\left[\frac{1}{s}\right]$		Reference organism		Degradation rate
RuBisCO	57.586		3.4		*N. tabacum*	[110]	$704,852.64^{-1}$
			3.7		*S. oleraceae*	[110]	$767,045.52^{-1}$
			5.8		Chlamydomonas	[110]	$1,202,395.68^{-1}$
PGAK	69.024		31.5		*V. radiata*	[111]	$7,827,321.6^{-1}$
GAPDH	56.976		61		*S. oleraceae*	[112]	$12,511,929.6^{-1}$
			223	(PGA)	Chlamydomonas	[113]	$45,740,332.8^{-1}$
			251	(NADPH)	Chlamydomonas	[113]	$51,483,513.6^{-1}$
			430	(NADPH)	Chlamydomonas	[114]	$88,198,848.0^{-1}$
TPI	188.48		4700	(GAP)	*S. cerevisiae*	[115]	$3,189,081,600.0^{-1}$
			500	(DHAP)	*S. cerevisiae*	[115]	$339,264,000.0^{-1}$
			4300	(GAP)	*T. molitor*	[116]	$2,917,670,400.0^{-1}$
			4467	(GAP)	*P. falciparum*	[117, 118]	$3,030,984,576.0^{-1}$
			9000	(GAP)	*E. coli*	[115]	$6,106,752,000.0^{-1}$
SFBA	45.847		40	(FBP)	*V. radiata*	[119]	$6,601,968.0^{-1}$
			33.3	(SBP)	*V. radiata*	[119]	$5,496,138.36^{-1}$
			64.5	(FBP)	*D. carota*	[120]	$10,645,673.4^{-1}$
			25.2	(SBP)	*D. carota*	[120]	$4,159,239.84^{-1}$
FBPase	73.986	(*SBPase*)	635	(reduced)	*S. oleracea*	[121]	$169,131,996.0^{-1}$
			22.9		*T. aestivum*	[122]	
TK	48.776		0.0187	(X5P)	*S. oleracea*	[123]	$3,283.60^{-1}$
			0.0148	(R5P)	*S. oleracea*	[123]	$2,598.79^{-1}$
SBPase	73.986		81	(reduced)	*S. oleracea*	[124]	$21,574,317.6^{-1}$
RPE	43.875		7100		*S. oleracea*	[123]	$1,121,445,000.0^{-1}$
			0.138		*S. oleracea*	[125]	$21,797.1^{-1}$
PRI	43.875	(*RPE*)	3440		*S. oleracea*	[126]	$543,348,000.0^{-1}$
PRK	30.359		463		Chlamydomonas	[127]	$50,602,381.2^{-1}$
			926		Chlamydomonas	[127]	$101,204,762.4^{-1}$
PGI	36.277		1000	(G6P)	*H. sapiens*	[128]	$130,597,200.0^{-1}$
			650	(F6P)	*H. sapiens*	[128]	$84,888,180.0^{-1}$
			18	(F6P)	*C. auratus*	[129]	$2,350,749.6^{-1}$
PGM	56.638		50	(G1P)	*P. aeruginosa*	[130]	$10,194,840.0^{-1}$
AGPase	33.107		158	(G1P)	*S. tuberosum*	[131]	$18,831,261.6^{-1}$
			166	(ATP)	*S. tuberosum*	[131]	$19,784,748.2^{-1}$
			38.17	(G1P)	*Z. mays*	[132]	$4,549,299.08^{-1}$
			43.32	(ATP)	*Z. mays*	[132]	$5,163,102.86^{-1}$
StS	36.101		235		*P. vulgaris*	[133]	$30,541,446.0^{-1}$
StP	32.506	(PHOA)	0.0037	(starch)	*S. tuberosum*	[134]	432.98^{-1}
			0.001	(G1P)	*S. tuberosum*	[134]	117.02^{-1}
UGPase	33.107	(*AGPase*)	3400		*T. latifolia*	[135]	$405,229,680.0^{-1}$
			27.04		*S. chungbukensis*	[136]	$3,222,767.81^{-1}$
SPS	36.101	(*StS*)	235	(*StS*)	*P. vulgaris*	[133]	$30,541,446.0^{-1}$
SPP	32.506	(*StP*)	0.0037	(*StP*)	*S. tuberosum*	[134]	432.98^{-1}
ATPS	44.083	(ATPC)	285	(ATP)	*E. coli*	[137]	$45,229,158.0^{-1}$
FNR	96.673		200	(Fd)	*S. oleracea*	[138]	$69,604,560.0^{-1}$
			139	(Fd)	*P. sativum*	[139]	$48,375,169.2^{-1}$
			90	(Fd)	Chlamydomonas	[138]	$31,322,052.0^{-1}$
NDPK	33.1	(*GDPK*)	147	(dTDP)	*P. sativum*	[140]	$17,516,520.0^{-1}$
			106	(dCDP)	*P. sativum*	[140]	$12,630,960.0^{-1}$
			2085	(ATP)	*P. sativum*	[141]	$248,448,600.0^{-1}$

3.4. Results and discussion

Based on the studies of Sajitz-Hermstein and Nikoloski [101], and Arnold and Nikoloski (2014) [38], we were able to determine the *de novo* synthesis costs of all enzymes considered in the modeled system. As mentioned above, the two approaches differ in the underlying large-scale model which results in, amongst others, different underlying growth conditions, namely, heterotrophic ones pursuing the approach of Sajitz-Hermstein and Nikoloski [101], and photoautotrophic ones based on the study of Arnold and Nikoloski (2014) [38]. The total enzyme synthesis costs in the photoautotrophic scenario are consistently larger than those under heterotrophic growth conditions (Table 3.7). This is caused by the higher amino acid synthesis costs under photoautotrophic conditions as there are no high-energetic precursors available nor amino acids can be taken up from soil. However, qualitatively the two measures are highly concordant as corroborated by Kendall rank correlation coefficient $\tau = 0.9943$.

Furthermore, the costs of an enzyme are, as expected, roughly in accordance with its molecular weight (Table 3.7). Large enzyme complexes, *i.e.*, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and ATP synthase, have large costs, while small monomeric or dimeric enzymes such as 3-phosphoglycerate (PGA) kinase, UGPase, and triose phosphate (TP) isomerase are least expensive, in terms of ATP demand. Interestingly, the "fast" enzymes with high turnover numbers, which are often modeled by mass-action kinetics or equilibrium approximations (see Chapter 2), have also low costs, *e.g.*, TP isomerase, Ribose-5-phosphate (R5P) isomerase, and UGPase. The exception is the octameric ribulose-5-phosphate (Ru5P) epimerase which would be the fourth least expensive enzyme if the costs per active site would be taken into account. However, since only the complete enzyme can be synthesized *de novo*, we account for the total energy demand.

Incorporating the determined enzymes costs and the corresponding turnover rates we were able to amend the metabolic costs for synthesizing the photosynthesis related end-products, *i.e.*, starch and sucrose. Therefore, we considered unphosphorylated sucrose and the artificial starch molecule comprising two $\alpha(1{\rightarrow}4)$ bound glucose units. Accordingly, both metabolites have the same molecular formula such that a comparison is unbiased with respect to material costs. The conventional examination of metabolic costs yield an equal energy demand of 38 ATP per molecule starch and sucrose. In contrast, the integration of enzymes results in shift towards distinct larger costs for sucrose synthesis, independently of the pursued approach (Table 3.8). More precisely, for the synthesis of one starch molecule of two glucose units about 170 and 219 molecules of ATP are needed under heterotrophic and photoautotrophic growth conditions, respectively, while producing one molecule of sucrose requires over 40 % more energy in both cases. These elevated costs suggest a purely stoichiometry-based explanation of storing carbon during the day as transitory starch rather than as sucrose. Certainly, this *in silico* analysis do not provide the complete, biological explanation but corroborates the observations from experiments in contrast to the conventional approach.

The analysis of the photosynthetic and the corresponding end-product processes shows the impact of supplementing the metabolic synthesis costs with the maintenance costs of the involved enzymes. Unfortunately, it is not applicable yet on large-scale metabolic models as the required data for determining the enzyme turnover rates, *i.e.*, enzyme half-lives, are

Table 3.7.: Total *de novo* synthesis costs of the catalyzing enzymes and their number of active sites (Act). SN and AN denote the pursued approach for the comprised amino acid biosynthesis costs of the respective enzyme, namely, Sajitz-Hermstein and Nikoloski [101], and Arnold and Nikoloski (2014) [38] (see Chapter 5).

Enzyme	EC	Act	Costs [ATP]		Molecular weight		
			SN	AN	[kDa]	Reference	
RuBisCO	4.1.1.39	8 (L)	193608	266916	560	Arabidopsis, Chlamydomonas	[142]
PGAK	2.7.2.3	1	16606	22603	39	*S. oleracea*	[143]
GAPDH	1.2.1.13	2 (A)	55964	75653	189	*S. oleracea*	[113]
TPI	5.3.1.1	2	21254	28847	59	*S. oleracea*	[144]
SFBA	4.1.2.13	4	56469	77236	148	*S. oleracea*	[145]
FBPase	3.1.3.11	4	58751	80076	160	*S. oleracea*	[146]
TK	2.2.1.1	2	52564	71867	160	*S. oleracea*	[123]
SBPase	3.1.3.37	2	27876	38121	66	*S. oleracea*	[147]
RPE	5.1.3.1	8	79187	108248	200	*S. oleracea*	[123]
PRI	5.3.1.6	2	19364	26425	49	*S. oleracea*	[126]
PRK	2.7.1.19	2	29213	40210	110	Arabidopsis	[148]
PGI	5.3.1.9	2	41069	56515	140	*S. oleracea*	[149]
PGM	5.4.2.2	1	22560	30784	60	*S. tuberosum*	[150]
AGPase	2.7.7.27	2 (S)	72952	99937	210	Arabidopsis	[151]
StS	2.4.1.21	1	41009	56608	70	*S. tuberosum*	[152]
StP	2.4.1.1	1	31386	43359	151	*S. oleracea*	[153]
ATPS	3.6.3.14	3 (B F_1)	195799	269065	≈500	*S. oleracea*	[154]
FNR	1.18.1.2	1	23306	31924	35	Arabidopsis	[155]
cTPI	5.3.1.1	2	17987	24380	59	*S. oleracea*	[144]
cSFBA	4.1.2.13	4	51511	70595	143	*S. oleracea*	[145]
cFBPase	3.1.3.11	4	48939	66839	130	*S. oleracea*	[146]
cPGI	5.3.1.9	2	40804	56329	125	*S. oleracea*	[149]
cPGM	5.4.2.2	1	20986	28601	60	*S. tuberosum*	[150]
UGPase	2.7.7.9	1	17181	23408	53	Arabidopsis, *S. tuberosum*	[156, 157]
SPS	2.4.1.14	2	77208	106277	270	*S. oleracea*	[158]
SPP	3.1.3.24	2	31432	43494	115	*P. sativum*	[159]
NDPK	2.7.4.6	6	51105	70130	111	*S. oleracea*	[160]

not available, particularly for Arabidopsis. Apart from this challenge, the bottleneck of the method is the determination of the unique amino acid composition as often the exact enzyme subunit structure is not recorded. However, the potential of this approach is, on the one hand, by approximating the actual metabolic costs to gain insight into resource allocation and application for, *e.g.*, storage, with respect to cost minimization. On the other hand, by varying the turnover rates regarding different environmental scenarios, we can begin to examine the effects of enzyme degradation on metabolic costs.

Table 3.8.: Metabolic costs of starch and sucrose synthesis with and without incorporation of enzyme degradation, respectively.

Biomass component	Conventional model	Extended model	
		based om SN	based om AN
Sucrose	38	242.2776	319.3640
Starch (two glucose units)	38	170.3808	219.3265

Chapter 4.

Comprehensive classification and perspective for modeling photorespiratory metabolism[7]

Biological processes involved in photorespiratory and photosynthetic metabolism operate concurrently and affect the interplay between carbon and nitrogen assimilation reflected in plant growth. Experimental evidence has indicated that photorespiratory metabolism has a wide-ranging influence not only on other principal metabolic pathways but also on a multitude of signaling cascades. Therefore, accurate quantitative models of photorespiration can provide a means for predicting and *in silico* probing of plant behavior at various levels of the system. We first present a comprehensive classification of current models of photorespiratory metabolism developed within the existing carbon-centric modeling paradigm. We then offer a perspective for modeling photorespiratory metabolism by considering the coupling of carbon and nitrogen metabolism in the context of compartmentalized, genome-scale metabolic models of C_3 plants. In addition, we outline the challenges stemming from the need to consider plant metabolic and signaling pathways in assessing the still controversial role of photorespiration and to confront the devised models with the ever increasing amounts of high-throughput data.

4.1. Background

Photorespiratory metabolism includes reactions involving multiple pools compartmentalized among the chloroplast, mitochondrion and peroxisome, and operates concurrently with photosynthetic metabolism. The role of photorespiratory metabolism goes beyond the obvious recycling of carbon in glycolate (GLC) back to glycerate (GA), and mounting evidence points to its important role in shaping the energetics of photosynthesis, compartmental reductant exchange, nitrate assimilation, one-carbon (C_1) metabolism and redox signal transduction (for recent reviews, see Foyer *et al.* [161], Bauwe *et al.* [162]). Nevertheless, understanding of the implications of photorespiration and modulation of its function on plant metabolism as a whole remains fragmented.

Quantitative modeling of photorespiration offers an alternative for *in silico* investigation of entire plant systems, and necessitates the inclusion of not only the metabolic level but also cell signaling and regulation of translation and transcription. This type of modeling alternative goes hand in hand with the high-throughput data obtained simultaneously from various system levels, *i.e.*, the transcriptome, proteome and metabolome. Furthermore,

---

[7]This chapter is based on the publication of Arnold and Nikoloski (2014) [47]. For consistency the abbreviations were adjusted and the language was changed to American English.

existing data indicate that the impact of photorespiratory metabolism increases with increasing light intensity, temperature, oxygen (O_2) concentration and water (H_2O) deficit [162–164]. Therefore, the study of photorespiratory metabolism requires examining the effect of concomitant variations in environmental conditions. Finally, the promise of quantitative modeling is that findings regarding the system's functioning under changing environmental conditions can be coupled with results of internal perturbations, *e.g.*, gene knockout, up- or down-regulation of gene expression and inclusion of novel reactions, to engineer a plant system having maximized yield.

Here, we first provide an overview of the model components of the photorespiratory pathway. We then provide a comprehensive classification of the existing modeling approaches of photorespiratory metabolism, and indicate assumptions, advantages and shortcomings of the existing models developed within the carbon-centric modeling paradigm. This succinct review includes models of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), its extensions and models coupling photosynthetic and photorespiratory metabolism. Subsequently, we present a detailed overview of existing genome-scale models incorporating the photorespiratory pathway. Finally, we present a perspective on how genome-scale modeling can be used to investigate the role of photorespiration in an integrated overview of metabolism.

4.2. Overview of model components

The plant photorespiration pathway involves 12 reactions that are partitioned among three compartments, namely the chloroplast, peroxisome and mitochondrion, with bypasses through the cytosol [165]. The reactions catalyzed by RuBisCO (EC 4.1.1.39) provide the interplay between photosynthesis and photorespiration. This enzyme initially reacts with ribulose-1,5-bisphosphate (RuBP), resulting in the enediol enzyme complex that can then react with O_2, which is termed oxygenation [54, 166], or react with carbon dioxide (CO_2), which is termed carboxylation. Upon oxygenation, one molecule of RuBP is transformed into one molecule of 3-phosphoglycerate (PGA) and one molecule of phosphoglycolate (2PG). The 2PG is hydrolyzed by phosphoglycolate phosphatase (EC 3.1.3.18) into GLC and inorganic phosphate (Pi); both of these reactions take place in the chloroplast. The resulting GLC is transported into the peroxisome and is there oxidized by glycolate oxidase (EC 1.1.3.15) to glyoxylate (GOX) and hydrogen peroxide (H_2O_2). Using catalase (EC 1.11.1.16), H_2O_2 is in turn degraded into H_2O and O_2. This step invokes the relationship between photorespiration and signaling pathways, since H_2O_2 serves as a trigger of many cascades [167]. Upon reacting with glutamate (*Glu*), a reaction catalyzed by glycine transaminase (EC 2.6.1.4) in the peroxisome, GOX is transformed into glycine (*Gly*) and α-ketoglutarate (KG). In the mitochondrion, through a series of four reactions, two molecules of *Gly* are transformed into one molecule each of serine (*Ser*), ammonia (NH_3) and CO_2. Three of these reactions, namely glycine dehydrogenase (decarboxylating) (EC 1.4.4.2), aminomethyltransferase (EC 2.1.2.10) and dihydrolipoyl dehydrogenase (EC 1.8.1.4), form the glycine cleavage system (GCS). The transformation of the second *Gly* into *Ser* is catalyzed by glycine hydroxymethyltransferase (EC 2.1.2.1). The resulting *Ser* is transported in the peroxisome where, together with GOX, it enters the reaction catalyzed by serine-glyoxylate transaminase (EC 2.6.1.45), resulting in 3-hydroxypyruvate (HPR) and *Gly*.

The glycerate dehydrogenase or hydroxypyruvate reductase, respectively, (EC 1.1.1.29 or 1.1.1.81; redundant) in turn transforms HPR into GA, also in the peroxisome. After which GA and adenosine triphosphate (ATP) enter the reaction catalyzed by glycerate kinase (EC 2.7.1.31) in the chloroplast, and are transformed into PGA and adenosine diphosphate (ADP), thus closing the photorespiratory cycle. For completeness, Table 4.1 includes the enzyme names, Enzyme Commission (EC) numbers, corresponding biochemical reactions together with their compartmentalization and reversibility, obtained from AraCyc 9.0 [32] and PlantCyc 6.0 [168].

4.3. Carbon-centric modeling paradigm

The existing models of photorespiration are carbon-centric as they only consider the effect of photorespiration on photosynthesis and disregard the role of nitrogen. They include a progressively larger set of rate-limiting steps to CO_2 assimilation, starting from mechanistic models of RuBisCO, to electron transport, production and consumption of reduced nicotinamide adenine dinucleotide phosphate (NADPH) and ATP, as well as inclusion of other metabolic pools. Moreover, some of the models have been extended from the level of a single enzyme, namely RuBisCO, to the chloroplast and the whole leaf. In the following, we present the assumptions on which the reviewed models are based, and provide a brief overview of their advantages and disadvantages with respect to the type of elicited conclusions.

4.3.1. Models of RuBisCO

The first attempts in modeling photorespiration and its relation to photosynthesis were focused on developing a mathematical description of RuBisCO. As carboxylation and oxygenation of RuBP directly connects photosynthetic carbon assimilation (*i.e.*, the Calvin-Benson cycle (CBC)) and photorespiration these two reactions are included in all considered models.

Let $[O_2]$ and $[CO_2]$ denote the atmospheric concentrations of O_2 and CO_2, respectively. Moreover, let Γ denote the concentration of CO_2 at which no net assimilation occurs (*i.e.*, rates of photosynthesis and photorespiration are equal), the so-called CO_2 compensation point. We will use v_o and v_c to denote the rates of oxygenation and carboxylation, and V_o and V_c for their maxima. The rates of dark respiration, photorespiration and (net) CO_2 assimilation will be denoted as R_d, v_p and A, respectively. We note that dark respiration includes the release of CO_2 from the mitochondria in the light, in addition to that resulting from photorespiration. Finally, K_{O_2} and K_{CO_2} will denote the Michaelis-Menten inhibition constants for O_2 and CO_2, respectively. Note that the maximum efficiency for a reaction of maximum rate $V_{\max}$ and the K_m value for its substrate is denoted by ψ, where $\psi = \frac{V_{\max}}{K_m}$.

A thorough literature search indicated that the earliest attempts to model the interplay between photorespiration and photosynthesis were those of Forrester *et al.* [170] and Tregunna *et al.* [171]. These authors set out to experimentally determine the effect of O_2 on the rate of photorespiration in detached soybean and tobacco leaves by measuring the release of CO_2 in both light and dark conditions. The authors indicated that $([CO_2] - \Gamma)$ represents the concentration of CO_2 available for photosynthesis. This quantity captures

Table 4.1.: Enzymes and reactions of photorespiration including different compartmentalization for Arabidopsis [161] and Chlamydomonas [169].

Enzyme name	EC	Reaction formula	Compartmentalization in Arabidopsis	Chlamydomonas
RuBisCO	4.1.1.39	$RuBP + O_2 \rightarrow PGA + 2PG$	[h]	[h]
2PG phosphatase	3.1.3.18	$2PG + H_2O \rightarrow GLC + Pi$	[h]	[h]
GLC oxidase	1.1.3.15	$GLC + O_2 \rightarrow GOX + H_2O_2$	[p]	–
GLC dehydrogenase	1.1.99.14	$GLC + Q \rightarrow GOX + QH_2$	–	[m]
Catalase	1.11.1.6	$2\ H_2O_2 \rightarrow 2\ H_2O + O_2$	[p]	–
Gly transaminase	2.6.1.4	$Glu + GOX \rightarrow Gly + KG$	[p]	–
Ala-GOX transaminase	2.6.1.44	$Ala + GLX \rightarrow Gly + Pyr$	–	[m]
Gly dehydrogenase (decarboxylating)	1.4.4.2	$Gly + LPL \rightarrow amDHP + CO_2$	[m]	[m]
Aminomethyltransferase	2.1.2.10	$amDHP + THF \rightarrow DHP + M\text{-}THF + NH_3$	[m]	[m]
DHP dehydrogenase	1.8.1.4	$DHP + NAD \rightleftharpoons LPL + NADH$	[m]	[m]
Gly hydroxymethyltransferase	2.1.2.1	$Gly + M\text{-}THF + H_2O \rightleftharpoons Ser + THF$	[m]	[m]
Ser-GOX transaminase	2.6.1.45	$Ser + GLX \rightarrow HPR + Gly$	[p]	[m]
GA dehydrogenase	1.1.1.29	$HPR + NADH \rightarrow GA + NAD$	[p]	–
HPR reductase	1.1.1.81	$HPR + NAD(P)H \rightarrow GA + NAD(P)$	–	[m]
GA kinase	2.7.1.31	$GA + ATP \rightarrow PGA + ADP$	[h]	[h]

Ala – alanine; amDHP – aminomethyldihydrolipoylprotein; DHP – dihydrolipoylamide; LPL – lipoylprotein; M-THF – 5,10-methylenetetrahydrofolate; Pyr – pyruvate; Q – ubiquinone; QH$_2$ – ubiquinol; THF – tetrahydrofolate

the inhibitory effect of O_2 as Γ increases linearly with $[O_2]$, given as $\Gamma \approx k\,[O_2]$. The rate of photorespiration can then be expressed as function of Γ, $[CO_2]$ and A as follows:

$$v_p = \frac{A}{CO_2 - \Gamma}\gamma. \tag{4.1}$$

Ogren and Bowes [166] defined A as the difference between the rate of true (gross) photosynthesis and the rate of photorespiration, *i.e.*,

$$A = \psi \cdot \exp\left(-\frac{[O_2]}{K_{O_2}}\right)[CO_2] - \psi \cdot \exp\left(-\frac{[O_2]}{K_{O_2}}\right)k\,[O_2], \tag{4.2}$$

where ψ is termed the *carboxylation efficiency of photosynthesis* in the absence of O_2. This parameter can be determined by measuring the rate of CO_2 evolution into CO_2-free gases at varying $[O_2]$, which is expressed as $-\psi \cdot \exp\left(-\frac{[O_2]}{K_{O_2}}\right)k\,[O_2]$. Rearranging the terms yields $v_p = \psi \cdot \exp\left(-\frac{[O_2]}{K_{O_2}}\right)\Gamma$ for the rate of photorespiration. These empirically developed expressions were used to demonstrate that O_2 has a large effect on CO_2 evolution in the light, *i.e.*, photorespiration, but has little to no effect on dark respiration, and thus historically contributes to delineation of the two processes.

Laing *et al.* [172] developed mathematical expressions for the velocities of carboxylation and oxygenation reactions catalyzed by RuBisCO, taking into account the competitive inhibition of CO_2 and O_2, resulting in the following:

$$v_c = \frac{V_c K_{O_2}\,[CO_2]}{K_{CO_2}K_{O_2} + K_{O_2}\,[CO_2] + K_{CO_2}\,[O_2]}, \tag{4.3}$$

$$v_o = \frac{V_o K_{CO_2}\,[O_2]}{K_{CO_2}K_{O_2} + K_{O_2}\,[CO_2] + K_{CO_2}\,[O_2]}. \tag{4.4}$$

Assuming that at low $[CO_2]$ the quantity $K_{O_2}\,[CO_2]$ becomes negligible, the rates of carboxylation and oxygenation can be written as:

$$w_c = \frac{V_c}{K_{CO_2}}\frac{K_{O_2}}{K_{O_2} + [O_2]}\,[CO_2], \tag{4.5}$$

$$w_o = \frac{V_o}{K_{CO_2}}\frac{K_{CO_2}}{K_{O_2} + [O_2]}\,[O_2]. \tag{4.6}$$

To take into account the proportion of GLC carbon that is released in photorespiration, one may introduce a scaling factor t into the oxygenation rate, which results in the following expression:

$$w_c - tw_o = \frac{V_c}{K_{CO_2}}\frac{K_{O_2}}{K_{O_2} + [O_2]}\left([CO_2] - t\frac{V_o K_{CO_2}}{V_c K_{O_2}}\,[O_2]\right). \tag{4.7}$$

Comparison of this expression with that empirically developed, above, indicates that $k = t\frac{V_o K_{CO_2}}{V_c K_{O_2}}$ and $\psi = \frac{V_c}{K_{CO_2}}$, since $\exp\left(-\frac{[O_2]}{K_{O_2}}\right) \approx \frac{K_{O_2}}{K_{CO_2} + [O_2]}$ from the Maclaurin series

expansion of e^{-x}. Since at Γ net photosynthesis is zero, it implies that

$$\Gamma = \frac{V_o K_{CO_2}}{V_c K_{O_2}} [O_2]. \tag{4.8}$$

Moreover, applying the same assumptions, the ratio between w_o and w_c, denoted as ϕ, can be expressed as

$$\phi = \frac{w_o}{w_c} = \frac{V_o K_{CO_2} [O_2]}{V_c K_{O_2}} [CO_2], \tag{4.9}$$

which captures the most essential relationship between photorespiration and photosynthesis as determined in RuBisCO-focused models. Farquhar [173] developed similar expressions for the rates of carboxylation and oxygenation, and indicated that they are appropriate for describing enzyme kinetics in conditions of excess RuBP, typically occurring when RuBisCO is assayed. Discarding the fact the carboxylation-oxygenation is an ordered reaction, and assuming that the dependence of the velocities of both reactions on free RuBP is the same (described by one apparent Michaelis-Menten constant, K_r), Farquhar [173] and Farquhar *et al.* [73] obtained

$$\frac{v_o}{w_o} = \frac{v_c}{w_c} = \frac{[RuBP]}{[RuBP] + K_r}. \tag{4.10}$$

This expression describes two cases with respect to whether the concentration of RuBP is smaller or larger than the total concentration of active sites. The authors pointed out that in vivo these are two cases that can become rate-limiting for carboxylation, later termed RuBisCO-limited and RuBP regeneration-limited photosynthesis. To further examine these cases, Farquhar *et al.* [73] also incorporated the adjacent photosynthetic reactions.

4.3.2. Extensions of RuBisCO-focused models

With the argument that in the photorespiratory pathway, oxygenation of RuBP with one molecule of O_2 releases 0.5 molecules CO_2, Farquhar *et al.* [73] described the net rate of CO_2 assimilation as

$$A = v_c - 0.5 v_o - R_d. \tag{4.11}$$

Therefore, the rate of photorespiration 'is defined' as half of the rate of oxygenation. At the compensation point, Γ, in absence of dark respiration, the oxygenation-carboxylation ratio is then $\phi = 2$. As the carboxylation reaction may not only be limited by the amount of RuBisCO, the authors incorporated RuBP availability as a rate-limiting step, the so-called RuBP regeneration-limited photosynthesis. They argued that regeneration itself is limited by the supply of NADPH and ATP from the light-dependent reactions and, therefore, consider their consumption.

Using stoichiometry-based arguments, Farquhar *et al.* [73] determined the rates of PGA production as well as NADPH and ATP consumption as functions of ϕ and v_c. Given the photon flux density (PFD), I, the fraction f lost as absorption by other than chloroplast lamellae, the maximum rate of carboxylation, J', allowed through electron transport is

then given as

$$J' = \frac{0.25 \left(1 - f\right) I}{2 + 2\phi},\tag{4.12}$$

where the denominator denotes the ratio between the rate of NADPH consumption and v_c. The actual rate of carboxylation, v_c, is then no greater than $\min(w_c, J')$. Using similar arguments, Farazdaghi and Edwards [174] determined the net rate of CO_2 assimilation in the absence of dark respiration.

Sharkey [175] introduced a third rate-limiting step for the RuBisCO reaction, namely, triose phosphate (TP) utilization. Starch and sucrose synthesis are required at one-third the rate of CO_2 fixation; otherwise, the level of Pi will decline and becomes rate limiting. This effect is included in many extensions of the model of Farquhar *et al.* [73] [75–77, 176], which, nevertheless, leads to only minor structural changes to the original model. The influence of different photosynthesis-related parameters, such as temperature of V_c, J' [74], R_d and Γ [75], light intensity of J' [75], leaf age of R_d and Γ [75], stomatal conductance to H_2O vapour [75] and starch accumulation (with and without the presence of sink activity) of J' [77], is taken into account. Although the inclusion of these parameters results in small changes to the model equations, the difference between the model predictions are sometimes quite large [46].

The models described above consist exclusively of algebraic equations that can only capture the steady-state behavior through restricting assumptions regarding coupling between the included metabolic pools. Typical for these models is that they first attempt to describe the relationship between steady-state supply and demand fluxes. Although they were extended to describe the behavior of the chloroplast and/or the whole leaf, they cannot be used to study transient changes of the relationship between photosynthesis and photorespiration at varying light intensities and concentrations of CO_2 and O_2. Moreover, due to their RuBisCO focus, their applications in investigating various metabolic engineering strategies are severely limited.

Gross *et al.* [177] restricted the model to the components deemed responsible for the dynamic responses of whole leaf photosynthesis to variations in PFD, namely light activation and dark deactivation of enzymes in the CBC, stomatal responses to varying PFD, variation in the internal leaf $[CO_2]$ and variations in the pool R, representing RuBP and all intermediate components leading to RuBP, and pool G, denoting components of the GLC and GA pathways. This resulted in six non-linear differential equations, of which three model the change in R, G and v_c over time, and the remaining three are taken from an existing model of stomatal conductance [178]. Although this model facilitates steady-state analysis as well as analysis of transient behavior, the functional form for some of the fluxes in the model remains questionable. On the other hand, Lawlor and Pearlman [179] established a relationship between steady-state carbon fluxes in a model containing RuBP, PGA, *Gly* and *Ser*, as well as the storage and sucrose pools, which were in turn used to simulate the dynamics of incorporating ^{14}C in the aforementioned pools.

4.3.3. Models coupling photosynthetic and photorespiratory metabolisms

A description of the dynamics of the pools involved in photosynthetic and photorespiratory metabolism was included in the model of Hahn (1987) [180], a simpler version of which is analyzed in Hahn (1991) [181], as well the model of Zhu *et al.* [51]. All of these have modeled this interplay based on kinetics and have partially considered compartmentalization. A quite different approach was pursued by Young *et al.* [182], who assembled a structural model of the underlying biochemical reactions to estimate fluxes from the associated carbon maps. Here, we briefly review the modeling assumptions in the four models together with the solution strategies.

The model of Hahn (1987) [180] is an extension to his mathematical model of C$_3$ leaf carbon metabolism [80, 183] to consider the oxygenation reaction as well as the GLC and GA pathways (as described in Tolbert [184]). The model is not mass-balanced regarding H$_2$O and O$_2$ and does not incorporate coenzymes such as nicotinamide adenine nucleotide (NAD). To reduce the number of variables, the chloroplast, peroxisome and mitochondrion are regarded as a single pool, referred to as the extended chloroplast. In addition, the model includes pools in the cytosol, vacuole, intercellular air spaces and mesophyll tissue. The model includes 33 state variables whose dynamic changes are described through a system of ordinary differential equations (ODEs) with mass action kinetics. Due to the choice of the kinetic law, this modeling approach does not consider regulatory mechanisms.

Inspection of the included biochemical reactions indicates that the GLC oxidase and the *Ser*:GOX transaminase are lumped with the catalase and the GA dehydrogenase, respectively, while the three reactions of the GCS and the *Gly* hydroxymethyltransferase are lumped into one. The other reactions are modeled as presented in Table 4.1. To find steady-state solutions, Hahn (1987) [180] provided two approaches: one for the state variables and the other for the rate constants. The first approach is based on the six conservation laws regarding: (1) concentrations of ATP and ADP, (2) concentrations of uridine tri- and di-phosphate (UTP and UDP), (3) Pi concentration in the extended chloroplast and (4) in the cytosol, (5) nitrogen concentration in the extended chloroplast and (6) concentrations of *Glu*, glutamine (*Gln*) and KG. One can then specify six of the state variables arbitrarily, which can be used to express the remaining variables as functions thereof. In the second approach, Hahn (1987) [180] specified two external parameters, the gross photosynthetic rate and the photorespiration rate per unit area of leaf tissue, in such a way that the photorespiration rate is 20% of the value of the gross photosynthetic rate. Therefore, the relationship between photorespiration and photosynthesis is a priori fixed and is not elucidated from the model itself. Based on these assumptions, the model includes 28 parameters describing the mass action kinetic constants, dark respiration rate constant and the diffusion rate constants for CO$_2$ and O$_2$ between the atmosphere and intercellular air spaces. Subsequently, the values for these parameters are obtained in dependence on the concentration of the considered pools and a trial-and-error procedure. We point out that with this solution strategy, the obtained values for the mass action kinetic constants are apparent and may reflect regulatory mechanisms in place, although mass action law is employed. This is probably the explanation for why, in a recent ranking of models for the CBC, a highly regulated process in itself, the model of Hahn (1986) [80] performs relatively well [46].

Hahn used the model to simulate the induction of photosynthesis after a period in absence

of light, with the assumption that the concentration for some of the CBC intermediates is zero. The predictions of the model for increased $[CO_2]$ and decreased $[O_2]$ implies that the inhibition of photosynthesis by photorespiration is eliminated, where the compensation point is assumed to be $\Gamma = 0.05$. However, the model results in higher GLC levels for lower O_2 concentrations, which, as Hahn (1987) [180] argued, might be a result of excluding some pathways in which this metabolite participates. Until today, no such pathways have been discovered, which would imply that the model of Hahn requires modifications leading to possibly different findings and implications.

To demonstrate that photorespiration has a stabilizing effect on photosynthesis, in the sense that without photorespiration a model of the CBC cannot exhibit a stable steady state, Hahn (1987) [180] simplified the model presented above to mathematically describe dynamic changes in three pools, namely RuBP, PGA and TP. These pools are transformed in six lumped stoichiometrically balanced biochemical reactions, whose fluxes are again mathematically described via mass action kinetics. The model was subsequently reduced to consider only two variables, *i.e.*, the pools of RuBP and PGA, with which one could still demonstrate the induction stages of photosynthesis and the inhibition of photosynthesis by photorespiration. While the findings from these models qualitatively matched observations from the medium-size kinetic model, the quantitative agreement and rigorous explorations of the entire phase space for the simplified models have not yet been undertaken.

The model of Zhu *et al.* [51] mathematically captures the state-of-the-art regarding knowledge of the CBC and endproduct processes, namely, starch and sucrose syntheses, as well as its interplay with photorespiration. It captures 28 state variables, including the pools of TP, hexose phosphates and pentose phosphates. Within the 36 reactions, the pool metabolites are used separately, which requires that the biochemical reactions interconverting such metabolites are given additionally as equilibrium approximations. The remaining reactions are modeled either via Michaelis-Menten kinetics, whereby, surprisingly, the rate of reversible reactions includes the denominators pertaining to irreversible reactions, or special kinetics taken from other models. As in the model of Hahn (1987) [180], GCS and the *Gly* transformation are lumped. However, in contrast to Hahn's model [180], GA dehydrogenase is included separately, while catalase is not considered. Interestingly, the concentrations of O_2, *Glu*, KG, and oxidized and reduced NAD and NADH are fixed, while Pi and ADP are bound by conservation relationships. The majority of model parameters are obtained from a large literature review. Recent analyses of models for the CBC have shown that by integrating required corrections [46, 185], the predictions from this model are in relatively good agreement with experimental measurements of metabolite concentrations [46].

As mentioned above, the model of Young *et al.* [182] uses the atom transitions within photoautotrophic metabolism of *Synechocystis* to estimate carbon fluxes from ^{13}C labeling data. In contrast to the previously described models, this is based on a structural modeling approach. Structural models are typically used to estimate the flux distribution for a given cellular scenario. To this end, the considered reactions coherently exclude coenzymes and by-products, and are partially merged. For instance, the seven reactions of photorespiration from *Gly* transaminase to GA dehydrogenase are modeled as one reaction. Such a modeling approach precludes detailed investigations of a particular pathway. Nevertheless, structural modeling provides the possibility to expand boundaries of the modeled metabolic processes.

Moreover, it addresses the limitations of kinetic modeling, which requires substantial knowledge about parameter values.

4.4. Genome-scale metabolic models – the perspective for modeling photorespiratory metabolism

Characterization of gene functions, especially for model species, provides the basis for developing large-scale models of metabolism covering multiple compartmentalized and highly coupled biochemical reactions. Accurate genome annotation is the first step in establishing the correct stoichiometry, which can in turn be used in structural modeling techniques, such as flux balance analysis (for a recent review, see Llaneras and Picó [29]). The number of reactions is usually much larger compared to the number of metabolites and, thus, determining the steady state fluxes results in solving an underdetermined system of linear equations. This system usually has an infinite number of solutions. To restrict the solutions, one usually constrains the space of feasible flux distributions by imposing flux boundaries, and assumes that the systems operate towards optimization of one or a combination of several objectives (expressed as linear combination or ratios of fluxes).

Here, we examined the accuracy of the core photorespiration pathway in the existing genome-scale metabolic networks in the model C_3 plant *Arabidopsis thaliana* (Arabidopsis), the alga *Chlamydomonas reinhardtii* (Chlamydomonas) as well as the cyanobacteria *Synechocystis* sp. PCC 6803 and *Cyanothece* sp. ATCC 51142. To this end, we use as reference the compartment and reversibility information of Arabidopsis from PlantCyc 6.0 [168] for each of the 12 reactions presented in Table 4.1. Finally, for comparison with C_4 species, we use two genome-scale metabolic models of *Zea mays*. The numbers of reactions and metabolites, as well as their source, are presented in Table 4.2 (for models for which there was no Systems Biology Markup Language (SBML) file available, the information from the original paper is used). In the following, we emphasize the reactions for which the direction is inaccurately assigned compared to that presented in Table 4.1. This is in agreement with the idea that a reversible reaction may assume a particular direction based on further constraints derived from data.

The model of Poolman *et al.* [104] contains 1406 reactions and 1297 metabolites, and does not consider compartmentalization and mass balancing with respect to H_2O. The Catalase is not included in the model, while the reactions of the GCS are given as one summary reaction. The AraGEM model of de Oliveira Dal'Molin *et al.* (2010) [186] contains 1625 reactions and 1767 metabolites, and considers compartmentalization. As in the model of Poolman *et al.* [104], the GCS is represented as one summary reaction. The compartmentalization of the reactions follows the reference given in Table 4.1. The model of Saha *et al.* [187] is slightly larger compared to AraGEM [186] and contains 1798 reactions and 1820 metabolites. Considering the lumping of reactions belonging to the photorespiration pathway, compartmentalization and reversibility, this model matches AraGEM [186]. The model of Mintz-Oron *et al.* [188] is the largest existing Arabidopsis model, comprising 3508 reactions and 2930 metabolites. It is the only genome-scale metabolic network model that covers secondary metabolism, *e.g.*, fatty acid synthesis. While the lumping of reactions belonging to the photorespiration pathway and their compartmentalization matches that of AraGEM [186], the lumped reaction is constrained to operate in the opposite direction

Table 4.2.: Photorespiration reactions included in genome-scale models. The arrows indicate the direction of the modeled reactions and the compartment in which the reaction takes place are denoted by : [c] – cytosol, [h] – chloroplast, [m] – mitochondrion, [p] – peroxisome. The oxidation of GLC and the reduction of HPR are here represented only by 1.1.3.15 (a) and 1.1.1.29 (b), respectively (Table 4.1).

Reaction (EC)	Arabidopsis				Chlamydomonas			Cyanobacteria			Zea	
	Poolman	AraGEM	Saha	Mintz-Oron	Boyle	AlgaGEM	Chang	Nogales	Knoop	Vu	C4GEM	Saha
	[104]	[186]	[187]	[188]	[189]	[190]	[191]	[192]	[193]	[194]	[195]	[187]
4.1.1.39	↑	[h]↑	[h]↑	[h]↑	[h]↑	[h]↑	[h]↑	↑	↑	↑	[h]↑	[h]↑
3.1.3.18	↑	[h]↑	[h]↑	[h]↑	[c]↑	[h]↑	[h]↑	↑	↑	↑	[h]↑	[h]↑
1.1.3.15	↑	[p]↑	[p]↑	[p]↑	[c]⇅	[p,m]↑	[p,m]↑	↑	↑	↑	[p]↑	[p]↑
1.11.1.6	–	[p]⇅	[p]↑	[p]↑	[p]↑	[p]↑	–	↑	↑	↑	[p]⇅	[p]⇅
2.6.1.4	⇅	[p]⇅	[p]⇅	[p]⇅	[c,m]⇅	[m]⇅	[m]↑	–	⇅	–	[p]⇅	[p]⇅
1.4.4.2					–		[m]↑					
2.1.2.10	↑	[m]⇅	[m]⇅	[m]↓	–	[m]⇅	[m]⇅	↑	↑	↑	[m]⇅	[m]⇅
1.8.1.4					–		[m]↑					
2.1.2.1	↑	[m]⇅	[m]⇅	[m]⇅	[m]⇅	[m]⇅	[m]⇅	⇅	⇅	⇅	[m]⇅	[m]⇅
2.6.1.45	⇅	[p]↑	[p]↑	[p]⇅	[c]↑	[p,m]↑	[m]⇅	⇅	⇅	⇅	[p]↑	[p]↑
1.1.1.29	→,⇌	[p]⇅	[p]⇅	[p]⇅	[c]⇅	[m]⇅	[c,m]↑	↑	⇅	⇅	[p]↑	[h]↑
2.7.1.31	↑	[h]⇅	[h]⇅	[h]⇅	[h]↑	[h]⇅	[h]↑	↑	↑	–	[h]↑	[h]↑
reactions	1406	1625	1798	3508*	484	1718	2191	863	380**	719	1243	1985
species	1297	1767	1820	2930*	458	1890	1706	795	291**	689	1432	2129

blue – reaction is not H_2O balanced, red – reaction direction or the given compartment is not in accordance with Table 4.1; slanted – number deviates from paper; * – taken from the compartmentalized model; ** – taken from COBRA toolbox version

53

compared to the reference in Table 4.1.

In Chlamydomonas, the photorespiratory pathway deviates from that in higher plants [169]. For instance, the oxidation of GLC to GOX is catalyzed by glycolate dehydrogenase (EC 1.1.99.14); together with all other peroxisomal reactions from the plant pathway, it is thought to be displaced to the mitochondrion. As the GLC dehydrogenase does not produce H_2O_2, the catalase is not needed for the photorespiration. The smallest of the genome-scale models for Chlamydomonas is that of Boyle and Morgan [189], including 484 reactions and 458 metabolites compartmentalized in the cytosol, chloroplast and mitochondrion. This model is not mass balanced with respect to H_2O, and includes neither the GLC dehydrogenase nor the plant counterpart. Moreover, it does not include the GCS. The 2PG phosphatase, the *Ser*:GOX transaminase, and the HPR reductase are inaccurately located in the cytosol. The larger model of Nogales *et al.* [192] consists of 863 reactions and 795 metabolites, but does not consider compartmentalization. Here, the reactions of the GCS are again lumped into one reaction. The AlgaGEM model of de Oliveira Dal'Molin *et al.* (2011) [190] is a compartmentalized model, including the chloroplast, peroxisome, mitochondrion and cytosol, and consists of 1718 reactions and 1890 metabolites. It captures current knowledge of the Chlamydomonas photorespiratory pathway, as well as some of the plant-specific reactions. As in almost all models, the GCS is modeled as a single reaction, while the reversibility follows that of AraGEM [186], since AlgaGEM [190] is its extension. The model of Chang *et al.* [191] is the largest compartmentalized model of Chlamydomonas, consisting of 2191 reactions and 1706 metabolites. It is the only model in which the GCS is modeled as three separate reactions. Furthermore, it correctly covers photorespiration regarding compartmentalization and reversibility.

The photorespiration of cyanobacteria is closer to that in Chlamydomonas than to plants, as it comprises exactly the same biochemical reactions as in Chlamydomonas. Of course, cyanobacteria are not compartmentalized as they have only a single compartment. The existing models of *Synechocystis* sp. PCC 6803 of Knoop *et al.* [193] and *Cyanothece* sp. ATCC 51142 of Vu *et al.* [194] are comparatively small, containing 380 and 719 reactions, and 291 and 689 metabolites, respectively. The reactions of the GCS are lumped in both models. Moreover, the model of Vu *et al.* [194] does not include the *Gly* transaminase and GA kinase.

For comparison, the two models of the C_4 species *Zea mays*, of de Oliveira Dal'Molin *et al.* (2010) [195] (C4GEM) and Saha *et al.* [187], contain 1243 and 1985 reactions, and 1432 and 2129 metabolites, respectively. Both models are compartmentalized, consider the GCS as one reaction, and follow the reversibility of AraGEM [195]. This is expected as both are extensions of the AraGEM model [186], and the photorespiration of *Zea mays* is almost identically annotated. The only exception is the GCS, which has not been assigned as a combination of several reactions in PlantCyc 6.0 [168]. However, in the model of Saha *et al.* [187], GA dehydrogenase is inaccurately localized in the chloroplast, while the reference in Table 4.1 indicates localization in the peroxisome.

This brief review indicates that all considered genome-scale models include almost all components of photorespiratory metabolism. Clearly, consideration of the GLC oxidase and the catalase for plants provides the entry point for investigating the role of photorespiration in modulating signals related to cellular energy metabolism. Through the GLC oxidase, photorespiration assumes the role of one of the fastest H_2O_2-producing systems under

different environmental conditions [196]. H_2O_2 has been recognized as a major signaling molecule in plants, notably during hormonal signaling to control plant growth and in stress responses [197]. Therefore, the photorespiration-dependent production of H_2O_2 in the peroxisome could be the first step for signal transduction from carbon assimilation in the chloroplast to the rest of the cell. The impact of changes in the rate and nature of producing of H_2O_2 on metabolism can therefore be captured by providing integrative modeling of signaling and metabolic pathways, which has already been attempted in unicellular organisms [198–200]. In addition, photorespiratory metabolism is thought to provide, through *Ser*, the main substrates to C_1 metabolism, supplying one-carbon units needed to synthesize proteins, nucleic acids and many methylated molecules [201], leading to important processes including DNA and histone methylation. It is expected that large-scale optimization approaches may provide some insights into the interrelations between the aforementioned processes.

While all these models describe the carbon metabolic interplay of photorespiration and photosynthesis, the connection to nitrogen metabolism is neglected. It is known that elevated levels of CO_2 inhibit photorespiration, increase photosynthetic activity and, therefore, accelerate plant growth. Moreover, it has been experimentally demonstrated that after a few days or weeks the rate of plant growth notably falls, and levels of leaf nitrogen and total protein ultimately decrease. There are several hypotheses explaining these coupled effects, known as CO_2 acclimation hypotheses. Beside the carbohydrate sink limitation [202] and the progressive nitrogen limitation theories [203], there is a third hypothesis, provided by Bloom [204], which considers the role of photorespiration. Bloom [204] pointed out that in C_3 plants, nitrate (NO_3) assimilation depends on photorespiration. Therefore, any inhibition of photorespiration directly affects NO_3 assimilation. Consequently, for plants using NO_3 as major nitrogen source, the level of organic nitrogen compounds, such as proteins, declines, resulting in a reduced biomass yield. As an explanation, Bloom [204] has provided three physiological mechanisms. First, reduction of NO_3 to nitrite (NO_2) consumes NADH, which among others, is produced through photorespiration [205]. Second, the conversion of NO_2 to ammonium requires transport from cytosol into chloroplast, which, in turn, requires the chloroplast to be more alkaline than the cytosol. Elevated CO_2 concentrations acidify the chloroplast stroma which then inhibits the abovementioned transport [206]. Third, NO_2 reduction depends on reduced ferredoxin, an electron donor that is also involved in several other reactions. The involved electron acceptors have different affinities for electrons; H_2O_2 reduction exhibits by far the highest affinity, while reduction of NO_2 and CO_2 fixation compete for the ferredoxin at high NO_2–CO_2 ratios [207].

There are a few models covering nitrogen metabolism [208–212] that can serve as a starting point for extending the carbon-centric approaches. All but one of these models are kinetic, comprising six to 17 ODEs and capturing very different aspects of nitrogen metabolism, such as true protein- and non-protein-related nitrogen conversion in the context of digestion [208], ammonia detoxification in liver lobuli [212] and, of course, the integration of inorganic nitrogen components at different levels of detail. While van Riel *et al.* [209] focused on five pools corresponding to the fixation of ammonia to synthesize protein via *Gln* and *Glu* only, Rabouille *et al.* [210] considered different carbon and nitrogen pools for fixation, storage, catabolism and growth. The most promising candidate as a starting point is the model of Resendis-Antonio *et al.* [211]. It is a structural genome-scale metabolic

network of *Rhizobium etli* comprising 387 reactions and 371 metabolites. As Rhizobia are nitrogen-fixing bacteria, the model captures complete nitrogen metabolism, which is largely consistent to that in plants, and related carbon metabolism, such as the citric acid cycle. As a first step, the two reaction lists of the carbon-centric plant and the *Rhizobium* metabolic networks could be merged and manually curated by removing *Rhizobium*-specific reactions and adding those that are plant-specific and nitrogen-related.

Once a complete metabolic network of carbon and nitrogen metabolism is assembled, the effects of the interplay of these two forms of metabolism can be readily tested. For instance, through perturbing nitrogen-specific reactions, the effects on carbon-specific reactions as well as effects on the whole system can be evaluated. This will provide new insights and opportunities for understanding photorespiratory influences on a plant system that may have evolved towards maximizing yield [213].

4.5. Conclusions

Mounting experimental evidence suggests that the role of photorespiration extends beyond the mere supply of the CBC with organic carbon. Photorespiratory metabolism interacts with other important metabolic pathways governing the interplay between carbon and nitrogen metabolism. Moreover, it provides the metabolites and metabolic precursors, such as H_2O_2 and methyl group donors, necessary for initiation of many crucial signal transduction processes. Furthermore, the systemic impact that photorespiratory metabolism may have is tightly linked with the division of its components into multiple compartments. In light of these novel insights, understanding fine tuning of the levels for photorespiratory intermediates and resulting fluxes to maintain a system that robustly responds to perpetually changing conditions is thus of major importance to modern plant biology.

Here, we have argued that existing attempts to establish kinetic models of photorespiration are solely carbon-centric, neglected the complex role of photorespiration, and have mostly accounted for only a handful of its intermediates. We have noted that existing models considered the coupling of simplified photorespiratory and photosynthetic metabolism, and have been used to determine some effects of photorespiration on the CBC. However, the scale of photorespiratory metabolism and the system levels affected by it suggest that accurate kinetic modeling, with the aim of revealing the control points and devising metabolic engineering strategies, would require massive efforts to experimentally determine the staggering number of parameters (even where mass action kinetics is assumed).

Constraint-based modeling, *e.g.*, flux balance analysis, coupled with high-throughput data provide an alternative to detailed kinetic modeling while accounting for the large number of participating reactions and metabolites. Recent extensions to this approach allow the inclusion of regulatory and signaling events that have so far been largely unrecognized in modeling photorespiratory metabolism. We believe that these approaches may provide the basis not only for generation of novel testable hypotheses, but also for synthetic engineering of pathways related to photorespiratory metabolism in order to optimize plant yield.

Chapter 5.

Bottom-up metabolic reconstruction of *Arabidopsis thaliana* and its application to determining the metabolic costs of enzyme production[8]

Large-scale modeling of plant metabolism provides the possibility to compare and contrast different cellular and environmental scenarios with the ultimate aim of identifying the components underlying the respective plant behavior. The existing models of *Arabidopsis thaliana* (Arabidopsis) are top-down assembled whereby the starting point is the annotated genome, in particular, the metabolic genes. Hence, dead-end metabolites and blocked reactions can arise which are subsequently addressed by using gap-filling algorithms in combination with species-unspecific genes. Here we present a bottom-up assembled, large-scale model which relies solely on Arabidopsis-specific annotations and results in inclusion only of manually curated reactions. While the existing models are largely condition-unspecific by employing a single biomass reaction, we provide three biomass compositions which pertain to realistic and frequently examined scenarios: carbon-limiting, nitrogen-limiting and optimal growth conditions. The comparative analysis indicates that the proposed Arabidopsis core model exhibits comparable efficiency in carbon utilization and flexibility to the existing network alternatives. Moreover, the model is utilized to quantify the energy demand of amino acid and enzyme *de novo* synthesis in photoautotrophic growth conditions. Illustrated on the case of the most abundant protein in the world, ribulose-1,5-bisphosphate carboxylase oxygenase, we determine its synthesis cost in terms of adenosine triphosphate (ATP) requirements. This, in turn, allows us to explore the trade-off between protein synthesis and growth in Arabidopsis. Altogether, the model provides a solid basis for completely species-specific integration of high-throughput data, such as gene expression levels, and for condition-specific investigations of *in silico* metabolic engineering strategies.

5.1. Background

Understanding plant responses to changing environmental conditions provides the opportunity of identifying and modifying the components involved in the underlying cellular

---

[8]This chapter is based on the manuscript of [38]. The assembled Arabidopsis core model and the supplemental files are provided on the attached CD.

mechanisms acting on gene regulation, signaling, and metabolism [214, 215]. Constraint-based modeling offers the means for predicting the behavior of plants and, thus, their responses, in different environments based on the stoichiometry of the considered biochemical reactions [216]. Therefore, recent research efforts have been aimed at assembling large-scale models of plant metabolism at different levels of cell-type and compartment resolution [186, 188]. Nevertheless, applications of these models lag behind their equivalents in the microbial kingdom, which have been successfully employed in simulating and predicting behavior of unicellular organisms under various internal (*e.g.*, genetic modifications) and external (*e.g.*, environmental stimuli) perturbations [213, 217].

The existing models have been assembled by following a top-down approach, whereby the list of partial plant-specific, but often not species-specific, annotations is augmented via gap-filling algorithms to achieve a functional network of biochemical reactions [218]. For simulating plant responses, an additional assumption is made regarding the optimization of a biomass reaction, reflecting the plant composition arising under a particular condition [186, 219]. Therefore, any predicted solutions, often resulting from subsequent optimization of secondary objectives (*e.g.*, photon usage efficiency [186], total flux minimization [40]), pertain to the sole scenario captured in the biomass reaction. Here we take a bottom-up approach to reconstruct a large-scale metabolic network of Arabidopsis which relies solely on Arabidopsis-specific annotations and results in the inclusion of only manually curated reactions (thus, avoiding the need of using gap-filling algorithms). In addition, by employing high-throughput data, we have assembled and validated biomass reactions representing carbon-limiting, nitrogen-limiting, and optimal growth conditions for Arabidopsis.

Enzymes are the driving force of biochemical reactions by enhancing the reaction rates, and, accordingly, form the link between proteome and metabolome. Understanding the costs of enzyme production, in terms of energy demand, provides the opportunity to estimate the expenditures for metabolism. In analogy to Craig and Weber [99], we define metabolic costs as the amount of energy sacrificed by diverting it to enzyme synthesis instead of utilizing it for ATP production/formation from adenosine diphosphate (ADP) and inorganic phosphate (Pi). Besides the metabolite interconversion, enzyme synthesis constitutes a large part of the metabolic energy demand. To this end, we ask the question of quantifying the cost of enzymes partitioned into costs of amino acid biosynthesis and protein assembly. While the energy demand of protein assembly can be well approximated, as shown in the analysis, the cost of amino acid biosynthesis requires the consideration of the underlying metabolic pathways. A careful curated metabolic model provides the opportunity for determining biochemically meaningful costs by incorporating the complexity of the synthesis pathways [101–103] (see Amino acid costs section). To this end, we make the distinction between physical (*e.g.*, light) energy and metabolic equivalents, including ATP and nicotinamide adenine nucleotide (phosphate; NAD(P)H). Moreover, as NAD(P)H can be converted into ATP via cellular respiration, metabolic costs can be solely expressed in terms of ATP equivalents. Finally, and most importantly, the costs will be biochemically meaningful for photoautotrophic growth conditions if the enzyme synthesis is only based on low-energetic inorganic precursors (*i.e.*, water, carbon dioxide (CO_2), Pi, nitrate/ammonium (NO_3/NH_4) and sulfate/hydrogen sulfide (SO_4/H_2S)). Once quantified, the enzyme costs are used to explore the trade-off between protein synthesis and growth, which is the ultimate application of the assembled model as illustrated on the case of ribulose-1,5-bisphosphate

carboxylase oxygenase (RuBisCO).

5.1.1. Succinct review of existing models of Arabidopsis metabolism

Large-scale modeling enables the investigation of specific pathways (*e.g.*, amino acid synthesis) or combinations thereof (*e.g.*, carbon metabolism) within their metabolic context. As metabolic pathways are tightly interconnected, their functionalities, regarded as objectives, usually depend on the remaining network. In case of Arabidopsis, there already exist a handful of large-scale metabolic models [104, 186, 188, 220] which assemble a wide range of Arabidopsis' genome-metabolome associations [104], incorporate compartmentalization [186, 188, 220] and, in the case of one of the most recent, accounts for pathways of the secondary metabolism [188]. However, these models are, unfortunately, unsuitable or incomplete for the purpose of determining conditions-specific cellular behavior as well as metabolic costs of enzyme production. The model of [104] reflects heterotrophic environmental conditions, neglects subcellular compartmentalization (and, thus, also transport reactions), and is not mass-balanced regarding oxygen and protons. While the updated and improved version of this model, proposed in [220] incorporates subcellular compartmentalization, the law of mass conservation is still violated. The subsequent models, including those of [186] and of [188] tackle also this issue and claim that they are capable to additionally simulate photoautotrophic scenarios. However, the model of [186] cannot produce all 20 amino acids providing SO_4, more precisely *Arg* and *His*. While the latest model from [188] resolve this drawback, there are other issues arising. On the one hand, the model violates the law of energy conservation in such a way that it can produce, for instance, biomass without any supply of energy, neither in form of light nor as high-energetic precursors. On the other hand, the gene-protein-reaction (GPR) associations for enzyme complexes comprising isoforms are inaccurate: For all 19 unique complexes, the logical operators declaring whether genes are encoding different subunits ('AND') or isoforms ('OR') are given in such a form that the 'OR' operation is of higher precedence than the 'AND' operation. Consequently, for a gene deletion of the large subunit of RuBisCO (RbcL), the model would predict no effect since the small subunit of RuBisCO (RbcS1A, RbcS2B and RbcS3B) is considered as a complete isozyme. Beyond these issues, the model of [188] additionally covers the secondary metabolism which can be disregarded for our purposes, *i.e.*, examining the central carbon metabolism and estimating enzyme costs. For these reasons, we have assembled a novel metabolic network capable of producing all amino acids by providing only inorganic compounds, while benefiting from the inclusion of verified knowledge only.

5.2. Results and Discussion

The starting point for the novel metabolic model was the photoautotrophic conditions whereby only the import of light, water, CO_2, Pi, NO_3 and/or NH_4, and SO_4 and/or H_2S is allowed. Consequently, pathways enabling the synthesis of all amino acids solely from these inorganic compounds were considered for inclusion. To this end, we first identified the metabolic precursors of the amino acids and their pathways, *e.g.*, Calvin-Benson cycle, glycolysis and citric acid cycle (Figure 5.1; Appendix Table B.10). We then extended the list

of required pathways to all associated pathways of the plant central carbon metabolism, such as light reactions and respiration. To assemble a metabolic model that can be employed to simulate various scenarios, we additionally included the synthesis pathways of the remaining known major cell components, or at least their precursors. As a result, the first draft of the model comprised twenty pathways and can produce sink and source sugars (*i.e.*, fructose, glucose (Glc), sucrose), a cell wall precursor (*i.e.*, uridine diphosphate glucose a representative of cellulose), a fatty acid precursor (*i.e.*, malonyl-CoA) and a signaling precursor (*i.e.*, trehalose).
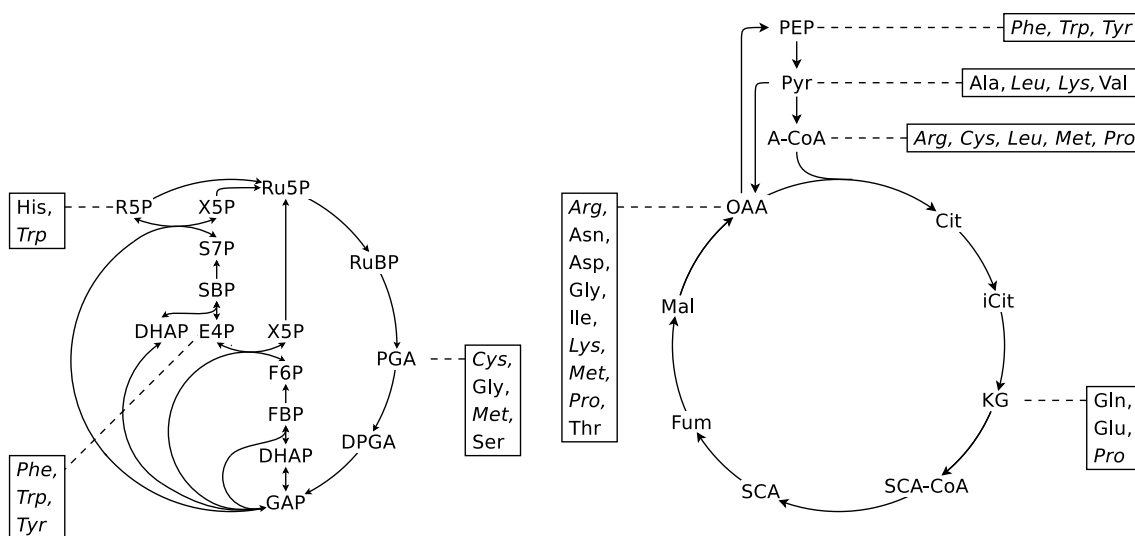

Figure 5.1.: Schematic representation of the most basic biochemical pathways comprising the metabolic precursors for the carbon skeletons of all amino acids. The list of pathways comprising the direct precursors is provided in Appendix Table B.10. Amino acids highlighted in italics have more than one precursor.

The reconstruction process is almost the same as for a genome-scale metabolic network and, therefore, we followed the protocol from [221] in all steps except the initial. In contrast to using genome-mapping as in previous top-down reconstructions, here we started with the essential pathways of the central carbon metabolism and identified, first, the underlying reactions, their corresponding Enzyme Commission (EC) numbers and, in the end, the annotated genes (Figure 5.2; Methods). Therefore, we denote our approach as bottom-up, resulting in what we term the *Arabidopsis core model* (Supplemental File 1 – SBML model).

The Arabidopsis core model represents a photoautotrophically growing Arabidopsis leaf cell, but is also capable to simulate heterotrophic scenarios by adapting the energy source and disabling the light reactions. Due to three experimentally determined biomass functions (see Methods and Appendix B.1.4), defined as the fractional contribution of known cell components to the overall biomass, the model can be utilized to simulate three realistic and frequently examined environmental settings, namely carbon-limiting, nitrogen-limiting and optimal growth conditions. Altogether, it comprises 236 unique metabolites, 345 unique reactions related to 61 unique subsystems which comprises 87 unique internal transport
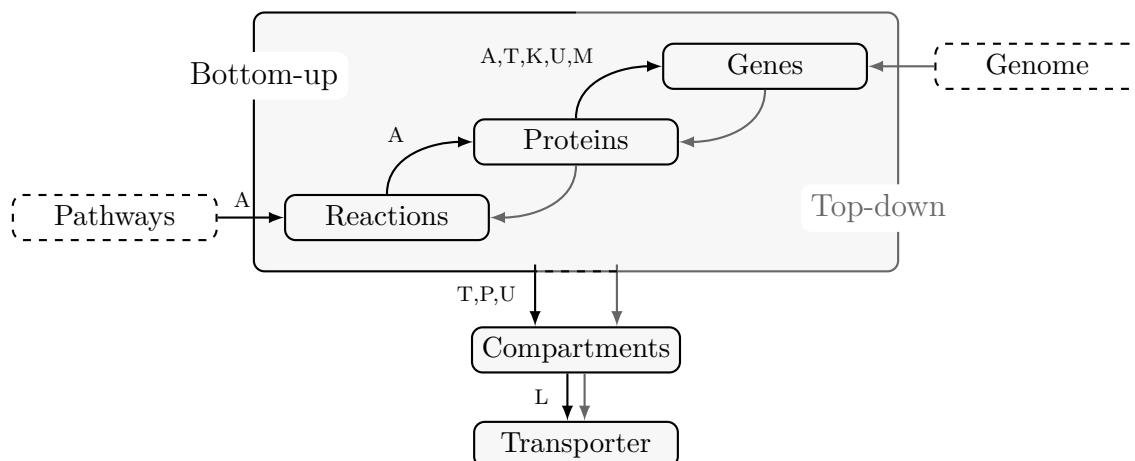
Figure 5.2.: Reconstruction workflow for the Arabidopsis core model. The arrow labels represent the used databases: A – AraCyc 11.5 [32], T – TAIR [222], K – KEGG [223], U – UniProt [224], M – MapMan [225], P – PPDB [226], L – Literature. Black arrows refer to the workflow of the bottom-up approach, while gray arrows denote the steps of the top-down approach.

reactions (Figure 5.3; Supplemental File 2 – Reaction list). The subsystems capture common biochemical pathways and functional groups such as transport. Moreover, the Arabidopsis core model is accurate with respect to mass and energy conservation, a prerequisite to ensure an optimal nutrient utilization and biochemical soundness of the predictions. By means of balancing the atoms of the left-hand and right-hand side of all internal reactions, the model utilizes exactly the amount of each precursor to produce one unit of biomass as the sum of molecular formula it specifies (Appendix Table B.6).

5.2.1. Model comparison

The Arabidopsis core model differs in several issues from the existing Arabidopsis models [104, 186, 188], not only due to the particular reconstruction approach employed. The photoautotrophic scenario permits only the uptake of inorganic compounds such that nitrogen, sulfur and phosphorus are taken up as ions, a process mostly associated with a proton uptake. In addition, this is the first model that includes the corresponding additional ATP costs to account for the proton secretion necessary for maintaining the intercellular pH value (Appendix B.1.2). Further, we incorporated the electron transport chains, namely light reactions and oxidative phosphorylation, as sequences of reactions with each reaction representing one of the five respective complexes. In doing so, we included the linkage of photosynthesis and nitrate assimilation via the competition for the reduced ferredoxin, and facilitated the consideration of the 14-fold symmetry of the ATP synthases subunit C, the proton-powered turbine [227].

Another substantial difference is the consideration of the GPR associations. As mentioned above, we considered only organism-specific annotations such that we accomplished a functional metabolic network without support of gap-filling algorithms, unlike the existing Arabidopsis models [104, 186, 188]. To this end, for only four biochemical reactions of the
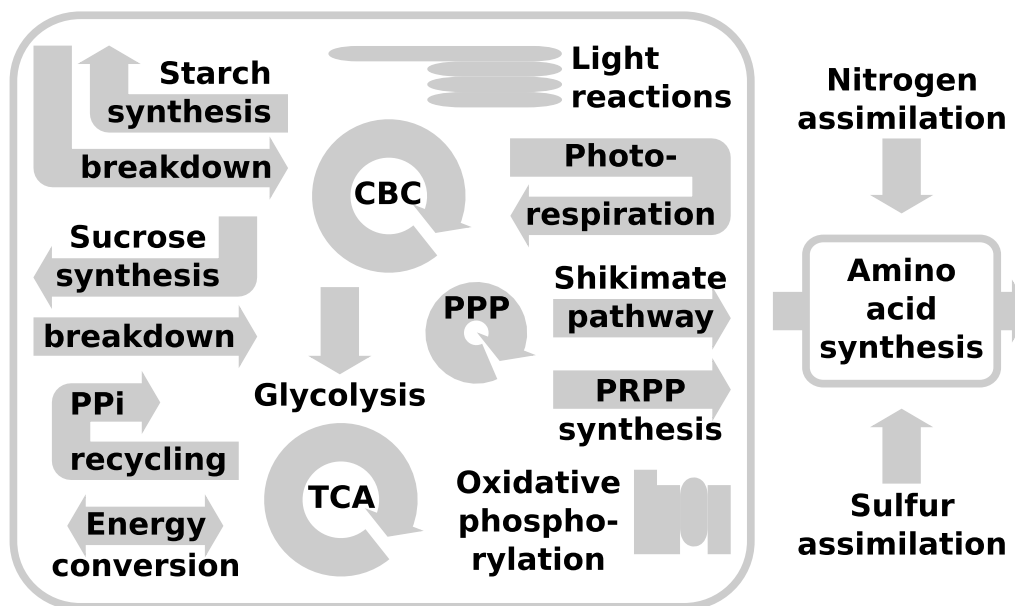
Figure 5.3.: Schematic overview of the major subsystems of the Arabidopsis core model.

Arabidopsis core model are no gene annotations provided, so-called spontaneous reactions $(4/(549 - 229) \approx 1.25 \%$; Table 5.1). In addition, only reactions involved in maintenance, transport, import and export processes, and reactions related to biomass production are not annotated (229 inapplicable; Table 5.1). In contrast, for the models of [186] and [188] 21 % and 37 % of these potentially annotated reactions, respectively, have no support for appearance in Arabidopsis. Importantly, spontaneous reactions are not excluded as there is not enough information provided for their precise identification. Nevertheless, for the Arabidopsis core model, the four nonannotated reactions are precisely such spontaneous reactions. Moreover, we did not provide only the common GPR associations, but also rendered the assignments compartment-specific where possible. Additionally, we collected the complete enzyme complex structure with respect to the stoichiometry (Appendix B.1.5; Supplemental File 2 – Reaction list). Together with the organism-specific annotations, this enables a reliable incorporation of Arabidopsis high-throughput data, such as gene expression levels, and, accordingly, plausible gene knockout as well as gene expression studies.

There are some aspects in which the Arabidopsis core model performs differently in comparison to the existing models. First, we tested the biomass simulated by the models through maximizing the production of biomass units under the same photoautotrophic conditions (Appendix Table B.8) and by using the biomass reaction of [186]. Technically, this maximization problem is solved by flux balance analysis (FBA; Appendix B.2.1). Based on 1000 molecules of CO_2 the Arabidopsis core model synthesizes 60.2845 units of biomass, while the models of [186] and [188] yield 55.0119 and 55.0212 biomass units, respectively (Table 5.1). This shows that the Arabidopsis core model predicts more efficient conversion of CO_2 into biomass. Nevertheless, for this purpose it assimilates more photons than the model of [186]. We would like to stress that it is not possible to make such a statement for

the model of [188] as, unfortunately, the import of photons is not necessary to produce biomass under photoautotrophic conditions in this model. Since no high-energetic, organic compounds are imported, in the respective model, light is the only external energy source. Therefore, the model of [188] violates the law of energy conservation such that it produces energy out of nothing within the system.

Furthermore, we tested the models for their flexibility by examining the flux variability, the reaction-specific range of feasible flux values ensuring optimal biomass production (Appendix B.2.2), and the flux coupling, the pairwise dependencies of feasible reaction fluxes (Appendix B.2.3). As expected, the Arabidopsis core model has the lowest flux variability frequency and the highest flux coupling frequency (Table 5.1) implicating lower flexibility in comparison to the other models considered. Intuitively, this is caused by the difference in its reconstruction, bottom-up instead of top-down, and the resulting smaller number of reactions.

Nevertheless, the comparative analyses are based only on the functional fraction of the models to ensure a consistent framework. To this end, we first removed all blocked reactions and, subsequently, all import and export reactions which were rendered redundant under photoautotrophic conditions along with the resulting blocked reactions. In this way, we achieved a considerably smaller but completely functional metabolic network specifically for the photoautotrophic conditions. While for the Arabidopsis core model and that of [188] approximately 80 % of the original reactions are functional, only 40 % of the reactions of the model of [186] are operational, *i.e.*, are capable of carrying a nonzero flux. The resulting functional network sizes are concordant with the respective flexibility, *i.e.*, the large functional model of [188] has a higher flux variability frequency and a lower flux coupling frequency than the others. This might result from the multitude of transport reactions enabling the cell-wide transport of for instance ADP, ATP, NAD(P) NAD(P)H and Pi in this model. Moreover, the flexibilities of the two models capturing only the primary metabolism, namely the Arabidopsis core model and the model of [186], are similar. This demonstrates that, in fact, a bottom-up reconstructed model can achieve a similar functional network size as a genome-scale metabolic network, assembled by a top-down procedure.

5.2.2. Cell performance with respect to different cellular scenarios

The performance of a cell is usually described in terms of the properties of a specific biochemical process, regarded as an objective. For instance, energy efficiency, expressed through ATP consumption, is often assumed and used as an objective [103, 228]. To this end, one determines the number of required ATP to produce a metabolite of interest or a set of metabolic precursors representing a biochemical process, such as sucrose or biomass. Accordingly, the optimal cell performance can be characterized by the minimum amount of required energy.

The validity of the cell performance depends on a clearly defined cellular state such as developmental stage, trophic status of the cell, and cell type. Juvenile and mature leaf cells differ in the respective predominantly active biochemical processes, *e.g.*, photosynthesis to enhance growth and to synthesize sucrose, respectively. Therefore, the appropriate choice of the biochemical process of interest is very important. Furthermore, while under heterotrophic conditions high-energetic organic precursors are provided to the cell, these

Table 5.1.: Model comparison regarding network properties pertaining to gene annotation, functionality, biomass production and flexibility, *i.e.*, flux variability and flux coupling. Reactions which are inapplicable for gene-annotations cover reactions involved in maintenance, transport, import or export processes nor related to biomass production (*). The flexibility analysis is performed for the photoautotrophic scenario (**; Table B.8).

Property	Arabidopsis core model	de Oliveira [186]	Mintz-Oron [188]
Total number of reactions	549	1601	3508
Annotated	316	1177	1696
Inapplicable*	229	111	811
Annotation-deficient	4	313	1001
Blocked	0	928	0
Redundant importer and exporter**	88	9	92
Additional blocked**	35	21	586
Functional**	426	643	2830
Maximal biomass (1000 CO_2)	60.2845	55.0119	55.0212
Flux variability frequency ($> 1\%$)	0.1423	0.2581	0.9132
Flux coupling frequency	0.0919	0.0412	0.0015
Fully	0.0047	0.0197	0.0009
Partially	0.0145	0.0002	0.0000
Directionally	0.0727	0.0212	0.0006

precursors first have to be synthesized from low-energetic inorganic substances under autotrophic conditions. Consequently, the energy efficiency would at least differ in the amount of ATP in excess, starting from the high-energetic precursors, and the amount of required ATP to synthesize them. For this study, we focused on autotrophic, juvenile leaf cells and assumed that growth enhancement as well as energy efficiency are appropriate objectives. As a further objective, we accounted for the optimal resource allocation by determining the minimal precursor requirements for a functional network.

Altogether, we investigated three different environmental conditions, *i.e.*, cellular scenarios, namely carbon-limiting, nitrogen-limiting and ambient growth conditions. These scenarios are represented by the three experimentally determined biomass compositions each reflecting 1 g dry weight (DW) of Arabidopsis leaf material (Appendix B.1.4; Supplemental Data 1 and 2). The ambient, or often called optimal, growth conditions are, if at all, limited by the availability of energy which corresponds to light limitation under autotrophic conditions. The experimental setup for the nitrogen limitation is based on a protocol from [229] that results in a mild but sustained restriction of growth. In accordance to the authors, we would like to stress that this differs from earlier nitrogen deficiency experiments where strong nitrogen limitations were obtained. In contrast, the carbon limitation is experimentally realized via short-day conditions (8:16 light-dark cycle) [97, 230–232] which affect the starch accumulation during the day and results in a restriction of carbon availability (during nights) [97, 232].

From the modeling perspective, the limiting environmental settings can be implemented by restricting the import of the respective nutrient source. Thus, carbon limitation is realized via restricted CO_2 import and nitrogen limitation by constraining NO_3 and/or NH_4 import. For the optimal growth scenario, we acted on the assumption of light limitation, since unlimited resource availability is not applicable. Nonetheless, as light absorption of a plant is limited by its overall leaf surface, this assumption is biochemically justified. Energy efficiency and minimal precursor requirements are determined by means of FBA (Appendix B.2.1). Here, the modeling characteristic of the photoautotrophic scenario is the the lack of overlap in the supply of nitrogen, carbon, sulfur and phosphorus, *i.e.*, each imported metabolite comprises only one of these chemical elements. This substantially simplifies the analysis as the required amount of precursors can be readily determined.

As expected, the cell performance varies across the cellular scenarios. The carbon-limiting scenario requires the least amount of each precursor as well as of the energy equivalents to produce one unit of biomass, namely 1 g DW (Table 5.2, columns 2–7). This indicates that the biomass contains less organic material in that it comprises more inorganic compounds or water. On the one hand, this is in accordance with the calculated biomass coverage (Appendix Table B.7). On the other hand, it points out that the nominal values of the required precursors are inappropriate to compare the different scenarios. Moreover, regarding the ATP requirement of the system, we would like to emphasize that the biomass compositions of the Arabidopsis core model do not include maintenance costs representing the energy demand necessary for cell replication, *e.g.*, macromolecular synthesis (growth-associated maintenance), and cell maintenance, *e.g.*, turgor pressure (nongrowth-associated maintenance; see Methods section). These costs are highly condition-specific [220] which would result in distinct, higher overall ATP demands and, accordingly, affected photon requirements. As a consequence, we considered the respective ratios of utilized CO_2 and NO_3 (Table 5.2, columns 8) which is notably smaller for the nitrogen-limiting conditions implying that the nitrogen incorporation and accumulation is increased. Experimentally, it has been confirmed that Arabidopsis growing under low NO_3 concentrations contain similar levels of proteins and higher levels of free amino acids [229]. In contrast, the cell nitrate content is reduced more than 10-fold which, nevertheless, results in a small decrease of the total nitrogen concentration. As the biomass reaction does not comprise NO_3, only the increase in the organic nitrogen can be examined. Moreover, it has been experimentally shown that the reduced overall nitrogen availability can almost be compensated by an elevated NO_3 assimilation [232]. In support of this claim, the gene expression of the low-affinity NO_3 transport system is shown to be induced under low NO_3 concentrations [233]. This demonstrates that the biomass function used in our modeling also reflects cell storage.

5.2.3. Estimation of enzyme costs

Metabolic costs can be considered as the amount of energy, in terms of ATP, sacrificed by diverting it to the synthesis of a building block instead of utilizing it for ATP production/formation from ADP and inorganic phosphate (Pi) [99]. The metabolic costs of an enzyme are mainly determined by its *de novo* synthesis costs related to its turnover and recycling. The half-life time, and thus the turnover, of a protein can be approximated by means of the N-terminal amino-acid residue, the so-called N-end rule [234]. For *E. coli*, *S. cerevisiae* and mammalian cells there exist quite precise protein turnover for the

Table 5.2.: Minimum requirements of precursors and energy equivalents to produce 1 mg
biomass of the representative cellular scenarios. The gray shaded entries denote
the respective limiting precursor.

Cellular scenario	Minimal precursor consumption $[\frac{mmol}{d}]$						Precursor ratio		
	$h\nu$	CO_2	NO_3	Pi	SO_4	ATP	$h\nu : CO_2$	$h\nu : NO_3$	$CO_2 : NO_3$
Optimal growth	196.862	20.621	2.595	0.013	0.065	186.316	9.547	75.875	7.948
Carbon limitation	182.639	19.192	2.350	0.013	0.058	172.854	9.517	77.714	8.166
Nitrogen limitation	259.648	26.737	4.104	0.014	0.099	245.738	9.711	63.260	6.514

different N-terminal residues [235, 236]. For plants, in contrast, while the stability order of chloroplastic and cytosolic proteins have been resolved [237, 238] determining the turnover times requires further investigations. Nevertheless, the key aspect for a proper estimation of enzyme costs is adequate modeling of the protein synthesis. This process actually comprises three parts: (1) the biosynthesis of the single amino acids, (2) the composition of the amino-acid sequence, and (3) the protein maturation. The costs of each part are separately determined and are then summed.

The costs for producing amino acids were approximated based on the assembled model. For this purpose, we relied on the assumption that the plant uses the synthesis pathway requiring the least amount of energy, in terms of ATP. The underlying optimization is solved using FBA (Appendix B.2.1). Since light energy is provided in the photoautotrophic scenario, we were able to directly restrict the consumption of energy instead of falling back on the utilization of metabolic precursors such as Glc. Subsequently, the minimal amount of required photons was converted into, on the one hand, the amount of ATP required to synthesize the amino acid of interest, and, on the other hand, the remaining number of ATP available for other processes (Appendix B.2.1). The result of these calculations is the minimal metabolic energy costs in terms of ATP (Figure 5.4; Appendix Table B.10) and provides one possible steady-state flux distribution, *i.e.*, a possible flux through the system which permits the synthesis of the amino acid without any accumulation or depletion of other metabolites.

To establish a cost measure for the protein synthesis in terms of assembling the amino-acid sequence, a representative amino-acid sequence for the respective enzyme has to be determined. This is in almost all cases not unique as enzymes can be multimers, can have isoforms, and/or different splicing forms can exist. Unfortunately, the complex structure of enzymes is only seldom available, *e.g.*, predominantly from the Braunschweig Enzyme Database (BRENDA) [239] and the AraCyc [32] database. Moreover, the structure of some enzymes varies across species and/or environmental conditions such as for ATP synthase (EC 3.6.3.14). Its proton-powered turbine, encoded by subunit III, is predicted to consist of 12 subunits in *E. coli*, 10 subunits in *S. cerevisiae* and 14 subunits for plants [227]. Furthermore, there is evidence that the size of this proton-powered turbine changes under stress conditions indicating a modification in the subunit stoichiometry [240]. As a

consequence, we accounted only for plant data and collected all verified compositions. We followed a similar procedure for the identification of subunit isoforms or whole isozymes as well as for different splicing forms of the encoding genes, respectively. Finally, we assembled all different combinations of enzyme compositions and, thereby, yield for RuBisCO 3003 potential amino-acid sequences.

To assign protein assembly costs, the next task consists of identifying the energy demanding steps which are fourfold: (1) the amino-acid activation with an estimated cost of two molecules ATP per amino acid, and the three stages of the translation process: (2) initiation, (3) elongation, and (4) termination (Appendix Table 3.5). Thereby, the formation of the initiation complex requires approximately three molecules ATP per polypeptide and the elongation of the amino-acid sequence consumes two molecules ATP per cycle and, thus, per amino acid. The final release of the new polypeptide during the termination stage costs another molecule of ATP per polypeptide.

The third component of the enzyme costs is related to the energy requirements for protein maturation. This comprises costs for error correction and maintenance of the biosynthesis apparatus, synthesis of signal sequences and post-translational processing such as methylation and phosphorylation [106, 108]. Therefore, one can approximate another molecule of ATP per amino acid for protein maturation (Appendix Table 3.5). Altogether, the simplified costs of protein synthesis and maturation can be estimated by five molecules of ATP per amino acid residue, respectively.

Due to the multitude of previously determined amino-acid sequences, the resulting costs of an enzyme may lie in a large range. As the enzyme costs are based on the minimal amino-acid costs, they are underestimated. The minimal cost value serves as a lower boundary for the estimation.

The range of costs permits insights of the extent of variation in actual enzyme costs. The comparison of different enzyme costs may be of specific interest to identify enzymes with highly variable costs. These enzymes may serve as a target for improving the systems efficiency, in terms of metabolic engineering.

Costs of amino acids

Amino acids are the basic components of all proteins. In metabolic modeling they are often even used as protein representatives as the protein synthesis itself is too elaborate to be modeled accurately. The various and intricate synthesis pathways of the different amino acids result in such a highly complex network that even their minimal costs lie in a quite wide range of 22.5 and 117.4 molecules ATP (Figure 5.4; Appendix Table B.10). The costs largely comply with the size of the corresponding amino acids. This compliance becomes higher by further considering other chemical features such as the number of amine groups or the incidence of sulfur. It seems that the incorporation of additional amine groups (*e.g.*, *Arg* and *Lys*) as well as the incorporation of sulfur (as for *Cys* and *Met*) is expensive. Apparently, also the formation of cyclic, especially aromatic, structures is very expensive. This is the case for *Trp*, *Tyr* and *Phe*, the aromatic amino acids, as well as for *His* which additionally contains three amine groups. Altogether, the calculated minimal costs coincide rather well with the complexity of the corresponding amino acids.

The comparison with existing cost measures facilitates the evaluation of our cost estimation (Appendix Table B.10). The existing approaches are threefold: (1) based on
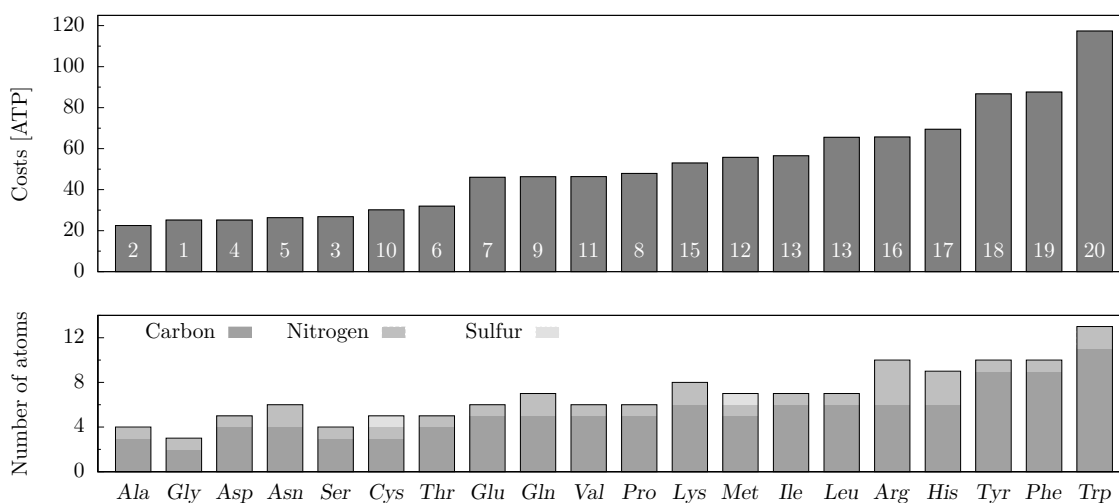
Figure 5.4.: Amino acid costs based on the Arabidopsis core model under optimal growth conditions. The white numbers at the bottom of the bars denote the average ranking over the amino acid cost estimation of [99], [241],[242], [243], [102], [101] and that of the Arabidopsis core model (see Appendix Table B.10). For comparison, the amino acid composition with respect to the size of the carbon-skeleton, and the number of incorporated amine groups and sulfur is given in the lower bar plot.

physiochemical properties of amino acids, (2) relying on selected metabolic pathways, and (3) based on genome-scale models. [242] employed the former approach using the molecular weight of the amino acids as a approximation of their synthesis costs. This enables a valid estimation across all species as the molecular weight is constant (Appendix Table B.10). In contrast, the remaining two approaches incorporate organism-specific information. For instance, [99], [241] as well as [243] followed the second approach relying on glycolysis, citric acid cycle and pentose phosphate pathway. Intermediates of these pathways are provided as precursors for the amino acid synthesis and the ATP equivalents directly invested into amino acid synthesis are counted. These studies are conducted for *E. coli* and *S. cerevisiae*. With the existing of genome-scale metabolic models the determination of amino acids costs in the metabolic context has been enabled for full genome sequenced organisms, such that first plant-specific costs were calculated. [102], [101], and [103] determined the costs for *S. cerevisiae*, Arabidopsis and *E. coli*, respectively, and compared their results to calculations based on central metabolic precursors. [101] provided an estimation specifically for day and night scenarios, whereby the minimal number of required ATP equivalents is determined with Glc provided to the system.

While all these studies describe heterotrophic scenarios, our estimations are determined for autotrophic conditions. As, moreover, only the approximation of [101] is plant-specific, we only consider the ranking of the amino acids for comparison. By ordering the amino acids regarding their average rank over all measures, they can be divided into four groups coinciding with our cost ranking: the five cheapest amino acids (*Ala, Gly, Ser, Asp, Asn*), two groups of six and four moderate costly ones, respectively (*Thr, Pro, Cys, Glu, Gln,*

Val, and *Met*, *Ile*, *Lys*, *Leu*), and the five most expensive ones (*Arg*, *His*, *Phe*, *Tyr* and *Trp*; Figure 5.4; Appendix Table B.10). Particularly for the last group the ranking is fully consistent as even the order within the group is the same.

Costs of RuBisCO

RuBisCO is the most abundant protein in plant cells and, moreover, the most abundant protein in the world [244]. It is one of the key enzymes of photosynthesis, more precisely it is the initial enzyme of the Calvin-Benson cycle. This attributed key position can be explained, on the one hand, by its very low catalytic rate of 3.4 to 3.7 fixed molecules carbon dioxide per second [110] and, on the other hand, it is justified by the substrate competition. Approximately each fourth reaction proceeds with oxygen instead of carbon dioxide [245] whereby the respective catalytic rate is even lower, 0.42 fixed molecules oxygen per second [246].

A single RuBisCO complex comprises 16 subunits, eight large and eight small ones. The large subunits are encoded by a single gene in Arabidopsis, namely ATCG00490. In contrast, the small subunits can be derived from four different genes, AT1G67090, AT5G38430, AT5G38420 and AT5G38410, which, furthermore, occur in two, one, one and three splicing forms, respectively. In order to determine the costs for RuBisCO, all possible 3003 amino acid compositions have to be analyzed. The resulting range of costs account for 243,287.9 to 269,133.9 molecules ATP per complex (Appendix Table B.11). For the same amount of ATP between 2508.1 to 2774.6 molecules of sucrose can be synthesized. To obtain a better impression of this huge amount of ATP, we converted the number of ATP molecules in the standard unit of physical energy. Using the Gibbs free energy of ATP synthase, one can easily establish that 36 kJ mol^{-1} ATP [247] results in 1.454 to 1.609 $\times 10^{-14}$ J per complex or 8.758 to 9.689 J nmol^{-1} RuBisCO. In comparison, 1 nmol of RuBisCO has a mass of 0.55 mg and the human heart and brain consume roughly 2 and 20 J s^{-1} [248–250], respectively.

Probably of greatest interest is the amortization of the production costs of RuBisCO. To this end, one can consider the resource allocation trade-off between growth and protein synthesis, as two contending tasks: increasing the leaf area allows a higher light absorption, while enhancing the protein synthesis stimulates and accelerates the metabolism. By means of the biomass composition representing optimal growth conditions, we were able to determine the potential growth enhancement, in terms of produced biomass, if resources are not utilized towards synthesizing RuBisCO. Based on the assumption that one unit of biomass represents 1 g DW of an Arabidopsis rosette and, accordingly, instead of synthesizing 1 nmol of RuBisCO the plant could gain 1.3058 to 1.4445 mg DW. The plants that were used to assemble the optimal growth biomass composition had an average mass of 124.36 mg FW [232], approximately 10.94 mg DW, which implicates an increase of 12 to 13 % DW. Certainly, this is only an approximation as, on the one hand, *de novo* synthesis of RuBisCO is considered and, on the other hand, cell maintenance costs are not incorporated (see Method section). However, these values seems to be in a physiologically plausible range considering a half-life of RuBisCO of approximately seven days in mature leaves [251] and an absolute growth rate of 0.9 mg DW d^{-1} [252]. According to this, Arabidopsis approximately gains 6.3 mg DW while one complex of RuBisCO has to be re-synthesized.

5.2.4. Perspectives of estimating enzyme costs

The presented enzyme cost estimation takes into account only the required energy for *de novo* synthesizing a complex of RuBisCO. However, in homeostasis synthesis costs are only of relevance if the degradation process is also considered. Accordingly, only the amount of degraded enzyme has to be synthesized whereby the breakdown process itself can involve additional energy requirements. Overall, the actual enzyme costs are highly affected by the protein degradation (rate) which varies with the environmental scenarios.

   A possible approach to assess the protein degradation is the examination of the relative costs per reaction. One complex of RuBisCO can catalyze in parallel eight carboxylation or oxygenation reactions, respectively. Moreover, as mentioned above RuBisCO has a half-life of approximately seven days [251] which results in a total number of approximately 1.6 to 1.74 million reactions per complex. Relative to each ongoing reaction, the costs for a complex of RuBisCO are between 0.14 and 0.17 parts of a molecule ATP per RuBisCO reaction. These additional costs represent an essential part of the overall metabolic costs that should be incorporated for the involved enzymes. Consequently, the cost estimations for all soluble metabolites, amino acids as well as for the enzymes have to be updated. This will improve the current estimations and lead to new insights regarding the metabolic relevance of the enzymes as the enzymatic influences can be taken into account.

5.3. Conclusion

In this study, we presented a bottom-up reconstruction of the extended central carbon metabolism of a young Arabidopsis leaf cell, and demonstrated that the resulting metabolic network model can be used to effectively simulate photoautotrophic conditions. Due to the bottom-up reconstruction and a careful manual curation, we resolved the shortcomings of the existing genome-scale models of Arabidopsis such as: (1) the demand for using gap-filling algorithms and, accordingly, a low gene-annotation coverage of the reactions [186, 188], (2) the occurrence of dead-end metabolites and blocked reactions [104, 186], (3) potential violation of mass and energy conservation [104, 188], and/or (4) incorrect GPR associations [188]. Combined with the, here presented, extended GPR associations, *i.e.*, the incorporation of the complex stoichiometry, the Arabidopsis core model can be readily employed for the integration of high-throughput data, *e.g.*, gene expression studies, and for investigations of other *in silico* metabolic engineering scenarios, *e.g.*, codon usage optimization.

   In comparison with the existing compartmentalized Arabidopsis models [186, 188], the presented Arabidopsis core model shows slightly higher efficiency in carbon utilization, and flexibility similar to that of [186]. In contrast to these models, the Arabidopsis core model comprises three biomass reactions which pertain to frequently examined scenarios: carbon-limiting, nitrogen-limiting and optimal growth conditions. The underlying compositions are determined experimentally and enable biologically reliable condition-specific analyses of the young leaf metabolism and the plant response.

   Moreover, we demonstrated the application of the Arabidopsis core model to estimate the energy demand of amino acid and enzyme synthesis in photoautotrophic conditions. Using the example of RuBisCO, the predominant protein of the world, we specify the

costs of the protein *de novo* synthesis, in terms of ATP requirements, and provide an approximation for the costs per catalyzed reaction. Finally, we used the Arabidopsis core model to explore the trade-off between protein synthesis and growth to quantify how much energetic resources, in terms of ATP, are sacrificed on synthesizing proteins instead of dedicating them to biomass production and vice versa.

5.4. Methods

The reconstruction of the Arabidopsis core model is in accordance with the protocol from [221], except for the initial step. We started from well-documented and essential biochemical pathways and identified the underlying reactions, their corresponding EC numbers and, in the end, assigned the annotated genes (Figure 5.2). This first, automated draft was assembled by means of the AraCyc 11.5 [32], The Arabidopsis Information Resource (TAIR) [222], Kyoto Encyclopedia of Genes and Genomes (KEGG) [223], UniProt [224] and MapMan [225] databases. Thereby, we only considered organism-specific data, particularly for the gene identification.

In the next manual curation stage, every entry of the draft was examined critically and the information was corrected, improved, or completed. This comprised the assignment of the correct metabolite formulas in dependence of the occurring pH value. As the model should be adaptable for different cell scenarios and the pH value of the cell varies with the environmental conditions, we enabled the assignment of pH-dependent charged formulas. Thereby, given the pH value of a compartment, the charge and, consequently, the metabolite formula was determined by employing the International Chemical Identifier (InChI) keys and the resulting pK_a values.

Moreover, we assigned information about the subcellular localization of a reaction extracted from TAIR [222], Plant Proteome Database (PPDB) [226] and UniProt [224] databases (Figure 5.2). We took into account four subcellular compartments – the chloroplast, the mitochondrion, the peroxisome and the cytosol which represents all remaining cell compartments. If contradictory or no information was available, we followed these three rules of thumb: (1) assign the compartment of the reactions in the vicinity of the pathway, (2) allow spontaneous reactions to be assigned in all compartments, and otherwise (3) restrict the reaction to take place only in the cytosol.

The compartmentalization not only requires but also permits the incorporation of subcellular transport reactions, representing active and passive transport processes across compartment boundaries (also termed transporters). To avoid futile cycles and to achieve a biological reliable network, we attempted to minimize the number of internal transporters and included primarily verified transport reactions. A thorough literature scan resulted in 87 different suitable transport reactions including diffusion processes of small, hydrophilic components as well as reactions realized by uniporters, symporters and antiporters. Transport reactions affecting the peroxisome were modeled as diffusion since the case is still not fully resolved whether active transporters are needed. Notably, experimentally verified transporters exist only for a small portion of metabolites. To render the model functional, in terms of ensuring that all reactions are functional (*i.e.*, can carry flux), we had to consider incorporating unconfirmed or even unreported transport reactions whose existence is speculated by experimentalists [253].

Biologically probably the most important step is the specification of a biomass function, defined as the fractional contribution of known cell components to the overall biomass. As the biomass varies with the cellular scenario, we have to account for the developmental stage of the cell and the cell type, as well as the environmental conditions. While growing cells produce the required components for cell division and expansion, mature leaf cells mainly synthesize the transport sugar, sucrose [254, 255]. In contrast, mature stem cells predominantly serve as mechanical support and transport system and, usually, convert the transport sugars into cell wall components. In addition, the overall biomass composition under ambient conditions greatly differs from the composition under stress conditions [256]. For these purposes, we assembled three different biomass functions for the model focusing on growing leaf cells and, therefore, comprising sugars, amino acids, nucleotides, and precursors for cell wall, fatty acids and signaling pathways (Table 5.3, Cell performance section). The proposed biomass compositions used in the presented model are based on refer to rosette fresh weight measurements of autotrophically grown plants which are converted into dry weight for modeling consistency (Appendix Text B.1.4). In contrast, the biomass compositions of the existing genome-scale Arabidopsis models were either experimentally determined heterotrophic cell culture compositions [104, 219, 257] or estimated leaf cell compositions [186].

It is noteworthy to emphasize that the biomass compositions of the Arabidopsis core model do not cover maintenance costs of a cell representing the energy demand necessary for cell replication, *e.g.*, macromolecular synthesis (growth-associated maintenance), and cell maintenance, *e.g.*, turgor pressure (nongrowth-associated maintenance). The reasons for this are twofold: (1) the established underlying experimental determination [220] is, unfortunately, not applicable for photoautotrophic scenario, and (2) a general approximation is biologically implausible as the maintenance energy demand is highly condition-specific. [220] could approximate the difference between energy expenditure of the cell, measured in terms of Glc uptake, and the requirements for synthesizing biomass *in silico* for heterotrophic Arabidopsis cell cultures. An analogous measurement for the photoautotrophic scenario would require the determination of the respective energy source, *e.g.*, photons. Furthermore, in the study of [220] it has been shown that the maintenance costs are highly condition-specific, between 13 and 79 % of total ATP produced. Consequently, the approximation of a general cost value for the three different growth conditions in the photoautotrophic scenario based on the heterotrophic cell culture data would be biologically unacceptable.

After collecting and reviewing, the resulting reconstruction was converted into a mathematical model using the COBRA toolbox. Thereby, we would like to emphasize that this single model can account for all three cellular scenarios by adjusting the respective model parameters. To facilitate easy access and wide usability, we provide these models into Systems Biology MarkUp Language (SBML) format (Supplemental File 1 – SBML model).

Table 5.3.: Major biomass components of an photoautotrophic Arabidopsis leaf cell and their approximated composition with respect to dry weight (DW). The detailed composite is given in Appendix B.1.4.

Component	$\frac{\mu\text{mol}}{\text{g DW}}$	References
Cell wall	363.88[a]	[258]
Protein	2911.09[b]	[232, 259]
Soluble metabolites	214.66[b]	[232]
Lipid	779.97[b]	[260]
Starch	294.86[a,b]	[232]
DNA, RNA	4.25[b]	[261]

[a] values for Glc dimer; [b] values of optimal growth conditions

Chapter 6.

Effects of varying nitrogen sources on amino acid synthesis costs in *Arabidopsis thaliana* under different light and carbon-source conditions[9]

The efficient allocation of resources in terms of matter and energy towards individual processes, like growth and reproduction, is essential to guarantee survival of plant species. The quantification of the costs of metabolite synthesis is a prerequisite to understand trade-offs arising from energetic limitation. Here, we define metabolic costs of a building block as the energy required for its *de novo* synthesis in terms of adenosine triphosphate (ATP), as a universal cellular energy currency. Amino acids synthesis is of accentuated interest, since it does not only bridge the energy demanding carbon and nitrogen metabolisms, but also provides the building blocks of proteins that enable metabolic functionality including resource acquisition. We quantify amino acid synthesis costs in *Arabidopsis thaliana* (Arabidopsis) utilizing flux balance analysis of three distinct state-of-the-art metabolic reconstructions to ensure robust results. As metabolism depends on and adjusts to the environmental conditions, we consider the day-night alternation, which is known to be highly influential, as well as photoautotrophic and heterotrophic growth conditions. In particular, we show that the nitrogen supply in the form of nitrate and/or ammonium is one of the key determinants of amino acid synthesis costs, in agreement with evidence from experiments. The association of costs with experimentally observed growth patterns suggests that metabolic costs are involved in shaping regulation of plant growth. Furthermore, we derive a purely stoichiometry-based explanation for the preferred simultaneous uptake of both sources (if provided) which has been observed in several plant species.

6.1. Background

Metabolism can be regarded as an integrated network of biochemical reactions through which simple chemical building blocks are assembled into larger molecules supporting cellular tasks which ensure viability and reproduction. As a result, metabolism does not operate in isolation from the other levels of cellular organization, and its state is tightly regulated via the participating proteins that enable the underlying reactions [262]. In addition, metabolism responds timely to changes in the environment or internal cues

---

[9]This chapter is based on the manuscript of [49]. The supplemental data is provided on the attached CD.

communicated via regulatory events and cascades [263]. The metabolic state is, therefore, ultimately modified by reallocating the utilization of the available resources, namely, the acquired simple precursors, the enzymatic composition, and the levels of intermediate building blocks.

In a simplified manner, metabolism can be viewed as a complex supply and demand network whose operation arguably and most critically depends on the environmental conditions. The metabolic costs of a building block can be defined by the resource requirement of its (*de novo*) synthesis, as these resources cannot be allocated to other metabolic processes. Clearly, metabolic costs will only affect the metabolic state if the respective resource is limiting (in terms of matter or energy). Metabolic costs in terms of energy have already been studied, as energy availability often limits growth [101, 264]. Thereby, the energy equivalent ATP can be utilized as a universal energy currency since it drives the majority of all endothermic cellular reactions. In plants, the availability of energy, provided by sunlight or carbohydrates, determine the energy budget of the cell. Therefore, understanding the (re)allocation of the cellular energy budget under varying conditions would shed light on the cost-related regulation of plant metabolism.

Amino acids bridge carbon and nitrogen metabolism in plants (as they are composed from both nutrients) and provide the building blocks of enzymes that enable metabolic functionality. Therefore, they are of particular interest when addressing the relation between metabolic costs and metabolic states, giving insight into resource allocation as a design principle. In line with this, it has been experimentally shown that, in Arabidopsis, the nitrogen source, predominantly nitrate (NO_3) and/or ammonium (NH_4), influences the amino acid and protein levels [265] as well as the growth rate [265, 266]. Therefore, we ask to what extent the *in silico* predictions of amino acid biosynthesis costs depend on the nitrogen source and whether or not they are in accordance with evidence from experiments.

Here, we employed large-scale modeling to examine system-wide metabolic responses to changing environmental conditions. However, the resulting predictions are expected to highly depend on (the quality of) the underlying metabolic model. To reduce the possibility that the results are mere artifacts from the used model, we conducted a comparative analysis of three metabolic networks of Arabidopsis, namely, the models of Poolman *et al.* [104], de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38]. In addition, we examined the influence of changing environmental and cellular scenarios on amino acid biosynthesis costs. Thereby, we accounted for day and night scenarios, as probably the most frequently examined with respect to effects on plant metabolism, under photoautotrophic and heterotrophic growth conditions. The associated alternation between photosynthetic carbon fixation and cellular respiration during night and the resulting transitory levels of starch [97] give rise to the assumption that these conditions likewise affect amino acid biosynthesis. Finally and most importantly, we elucidated the influence of nitrogen supply on amino acid synthesis costs by examining four different scenarios: sole NO_3 and NH_4 supply, respectively, as well as supply of both nitrogen sources with an additional constraint for equal uptake (termed, "equal nitrogen uptake") and without (termed, "arbitrary nitrogen uptake"). The latter results in a self-adjusting uptake ratio towards optimal energy source utilization.

Altogether, our predictions are in accordance with the experimental data and are supported by the comparative analysis of state-of-the-art large-scale model reconstructions

of Arabidopsis. Furthermore, the findings confirm the prevailing observation of preferred simultaneous acquisition of nitrogen in both forms, as NO_3 and NH_4, for optimal plant growth [267, 268].

6.1.1. Cost estimation of amino acid synthesis

We quantify metabolic costs of amino acids in terms of the metabolic energy equivalent ATP. The energetic equivalency of ATP is given by the Gibbs free energy which is consumed during ATP formation from adenosine diphosphate (ADP) and inorganic phosphate (Pi) or released via ATP hydrolysis. For simplicity, we refer to these processes as "production of ATP" and "consumption of ATP", respectively. As ATP hydrolysis can drive various endothermic reactions by providing free energy and because other energy equivalents (*e.g.*, the redox equivalent nicotinamide adenine nucleotide (phosphate; NAD(P)H)) can be utilized for ATP production, we refer to ATP as the universal cellular energy currency.

We define the cost of an amino acid as the energy required for its *de novo* synthesis in terms of ATP. In analogy to Craig and Weber [99], cost is quantified as the ATP sacrificed by diverting energy source to amino acid synthesis instead of ATP production. This definition differs from that used in Kaleta *et al.* [103], based on the net ATP usage, which implies that allocation of resources to amino acid synthesis does not impose a burden on remaining metabolism. We distinguish between different carbon, nitrogen, and energy sources pertaining to the environmental and cellular scenarios of interest. As in the approach of Sajitz-Hermstein and Nikoloski [101], we utilize available metabolic reconstructions together with flux balance analysis [28] to determine the corresponding environment- and organism-specific costs. We assume that an organism employs the synthesis pathway requiring minimum energy. Therefore, we provide lower bounds of the metabolic costs.

The amino acid synthesis costs are determined in three optimization steps (utilizing flux balance analysis), resulting in: (i) maximum of the ATP production per energy source uptake, termed ATP production efficiency, (ii) minimum energy uptake per unit amino acid synthesis in terms of ATP, termed total ATP cost, and (iii) maximum surplus ATP production while synthesizing the amino acid, denoted by surplus ATP cost. The total ATP costs are described as the product of the minimum energy source uptake per unit amino acid synthesis and the ATP production efficiency. The total ATP costs can be higher than the amount of ATP actually utilized in the synthesis of the amino acid. On the one hand, the energy source may also serve as the carbon source, *e.g.*, glucose (Glc) in heterotrophic cells. In this case, it can occur that more Glc is taken up than needed for providing energy to drive amino acid synthesis, because the demand for building the carbon skeletons exceeds the energetic demand. On the other hand, inflexibilities of the system could render it incapable to optimally utilize the energy source, *e.g.*, due to fixed stoichiometries in the generation of ATP and NAD(P)H (explored in greater detail below). In both cases, while synthesizing the amino acid with minimum energy source uptake, a surplus of ATP can be generated and diverted to other processes. Altogether, the energetic costs of amino acid synthesis are calculated as the difference of the total and the surplus ATP costs. A detailed description of the three steps of the computational procedure is given in the Methods section.

6.2. Results and Discussion

In accordance with earlier studies [99, 103, 241, 242], the costs of the different amino acids fall in a wide range from 11 to 129.2 ATP per amino acid. Nevertheless, the costs largely differ across the used models, examined environmental and cellular scenarios, as well as the different nitrogen sources. First, we examined how the amino acid costs are affected by the used metabolic network. Second, we explored the impact of day and night as well as of cellular conditions. In particular, we examined the two major cell types with respect to their trophic level: carbon-fixing cells using the light energy, photoautotrophs, and cells utilizing organic compounds as both carbon and energy source, namely chemoheterotrophs. In the following, we refer to these as autotrophic and heterotrophic cells, respectively. Obviously, the night scenario is also heterotrophic. In the analysis, we considered the sole NH_4 supply as a nitrogen reference condition. Third, we analyzed the influence of distinct nitrogen sources and their combination on the amino acid costs.

6.2.1. Model comparison

The used metabolic models of Poolman *et al.* [104], de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38] represent the state-of-the-art modeling of Arabidopsis' primary metabolism. The model of Poolman *et al.* [104] describes heterotrophic leaf cells, whereas the models of de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38] are capable to simulate both heterotrophic and autotrophic growth. The models of Poolman *et al.* [104] and de Oliveira Dal'Molin *et al.* [186] are both top-down reconstructions, starting from the genome annotations. In contrast, the model of Arnold and Nikoloski [38] is reconstructed in bottom-up fashion, whereby the operability of the incorporated reactions and metabolites is ensured by starting from well-documented and vital biochemical pathways. Therefore, the models differ in the total number of reactions and metabolites. Nevertheless, the size of the underlying operational network, which excludes blocked reactions and dead-end metabolites, is comparable (Table 6.1). Here, unique reactions and metabolites correspond to the numbers of different reactions and metabolites across all compartments. As the model of Poolman *et al.* [104] comprises equivalence reactions, *e.g.*, interconverting two species of fumarate ('FumEquiv'), in this case we do not provide the number for unique reactions and metabolites.

Furthermore, the model of Poolman *et al.* [104] is uncompartmentalized, although it contains a pseudo-compartment including a small number of segregated metabolites without any corresponding reactions. In the model of Poolman *et al.* [104], the corresponding metabolites are labeled by 'mit' and comprise NAD, NADH and protons which are participating in the citric acid cycle and the oxidative phosphorylation. In contrast, the models of both de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38] contain the four major cell compartments: cytosol, chloroplast, mitochondrion and peroxisome. In addition, the model of de Oliveira Dal'Molin *et al.* [186] includes three pseudo compartments, namely the vacuole, an accumulation compartment, and a biomass compartment, to enable specific export. The model of Arnold and Nikoloski [38] contains the lumen in the chloroplast and the intermembrane space in the mitochondrion, as two subcompartments to enable the generation of the proton motive force for ATP formation. This pseudo compartmentalization may cause the occurrence of futile cycles, as in the case of the model of Poolman *et al.*

Table 6.1.: Comparison of the models of Poolman *et al.* [104], de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38] with respect to the network properties pertaining to functionality.

Property	Poolman		de Oliveira		Arnold	
Reactions (unique)	1406		1601	(1472)	549	(345)
Blocked reactions	712		956		0	
Functional reactions (unique)	**694**		**645**	**(578)**	**549**	**(345)**
Importer and exporter	39	(39)	18	(18)	97	(37)
Metabolites (unique)	1253		1736	(1508)	407	(236)
Dead-end metabolites	428		797		0	
Functional metabolites (unique)	**825**		**939**	**(752)**	**407**	**(236)**
Compartments (pseudo)	1	(1)	4	(3)	4	(2)
Transporter (unique)	–		80	(42)	125	(87)
Blocked transporter	–		15		0	
Functional transporter (unique)	–		65	(36)	125	(87)

[104], which prohibits the analysis of ATP production in the original setup. To resolve this issue and allow a reasonable comparison across the different models, minor modifications of the models are applied (see Methods).

The interpretation of metabolite synthesis costs in terms of energy might depend on the overall energetic expenditure of an organism into individual metabolite synthesis. Therefore, instead of considering only the synthesis costs, we examined the ratios of total energetic investment into amino acids of Arabidopsis bound in proteins to obtain biologically reliable results. To this end, we determined the amino acid frequencies, denoted by f, in Arabidopsis leaf proteins [259] (see Supplemental Data 3), which we combine with amino acid costs. It has to be noted that we cannot ascertain the available nitrogen source(s) in the underlying data set. Furthermore, it is unclear if the therein examined cells were strictly autotrophic or heterotrophic, as both cell types had most likely been utilized. Therefore, the analyses of amino acid costs are presented not only for each model but also for the different trophic levels and environmental conditions. We consider amino synthesis costs at day which is in accordance with the setup of the utilized data set.

We calculated first the average weighted amino acid costs, C_w^{av} (Equation 6.1)

$$C_w^{av} = \sum_{i=1}^{n} f(aa_i)\, C(aa_i)\,. \tag{6.1}$$

The model of Poolman *et al.* [104] yields lowest average costs for each examined trophic and nitrogen uptake scenario compared to the costs based on the other models (Table 6.2). This is most likely due to the absence of compartments in this model, whereby the synthesis pathways are simplified, such that some metabolite interconversions, requiring additional ATP, do not take place. For instance, according to AraCyc [32], guanosine monophosphate is synthesized in the cytosol starting from inosine monophosphate (IMP) which, in contrast, is synthesized in the chloroplast. The IMP itself cannot be transported across the membranes separating these two compartments. Therefore, IMP has first to be converted to adenosine

monophosphate (AMP) under consumption of energy. After crossing the membrane, AMP is converted back to IMP. Obviously, such conversions do not take place in the model of Poolman *et al.* [104]. In contrast, the models of de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38] are both compartmentalized and, accordingly, show higher costs for each scenario.

Next, we determined the average weighted distance of amino acid costs, δ_w^{av}, which quantifies the similarity of costs obtained by utilizing two different metabolic networks (Equation 6.2)

$$\delta_w^{av} = \sum_{i=1}^{n} f(aa_i)\,|C_1(aa_i) - C_2(aa_i)\,|.$$ (6.2)

The costs based on the models of de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38] are more similar to each other than to the costs based on the model of Poolman *et al.* [104]. This is corroborated by the Kendall rank correlation coefficient (see Table 6.2). These qualitative findings hold irrespectively of the trophic level and the nitrogen source uptake scenario, and may indicate that the costs obtained by the models of de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38] are more reliable.

Table 6.2.: Model comparison of Poolman *et al.* [104], de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38] with respect to weighted average amino acid cost, C_w^{av}, and average weighted distance of amino acid costs, δ_w^{av} as well as Kendall rank correlation τ of the product of frequencies and costs (p-value < 0.5 in all cases). The trophic level scenarios, heterotrophic and autotrophic, are denoted by Het and Aut, respectively. The examined nitrogen source scenarios are sole NH_4 supply, equal uptake of NH_4 and NO_3 (50:50), sole NO_3 supply, and arbitrary nitrogen uptake (arb).

| C_w^{av}	Poolman				de Oliveira				Arnold			
	NH_4	50:50	NO_3	arb	NH_4	50:50	NO_3	arb	NH_4	50:50	NO_3	arb
Het	28.19	28.19	28.19	28.19	31.41	42.85	54.89	31.41	31.13	38.83	46.8	34.44
Aut	–	–	–	–	43.76	54.92	66.09	43.76	42.06	50.6	60.51	46.75
δ_w^{av}	Poolman : Arnold				de Oliveira : Poolman				Arnold : de Oliveira			
	NH_4	50:50	NO_3	arb	NH_4	50:50	NO_3	arb	NH_4	50:50	NO_3	arb
Het	5.61	10.82	18.61	7.92	5.01	14.67	26.7	5.01	2.44	5	8.09	4.87
Aut	–	–	–	–	–	–	–	–	2.54	4.57	5.58	4.39
τ	Poolman : Arnold				de Oliveira : Poolman				Arnold : de Oliveira			
	NH_4	50:50	NO_3	arb	NH_4	50:50	NO_3	arb	NH_4	50:50	NO_3	arb
Het	0.67	0.68	0.72	0.6	0.75	0.73	0.69	0.75	0.88	0.87	0.96	0.77
Aut	–	–	–	–	–	–	–	–	0.93	0.89	0.94	0.89

6.2.2. Comparison of day and night scenarios under autotrophic and heterotrophic growth conditions

Probably the most influential environmental factor investigated in plant research is the alternation of day and night which causes extensive changes in plant metabolism, including most likely also adjustments in amino acid biosynthesis [97, 269–272]. As another important factor, we examined the impact of the trophic level of the cell, more precisely, the influence of autotrophic and heterotrophic scenarios. As the model of Poolman *et al.* [104] does not comprise light reactions, the comparison of heterotrophic and autotrophic day conditions can be conducted only for the models of de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38].Furthermore, we performed the analyses for all relevant inorganic nitrogen supply settings to arrive at conclusions which hold irrespectively of the nitrogen source.

In general, the costs of amino acids under heterotrophic day and (heterotrophic) night conditions are more similar to each other compared to autotrophic day conditions. While the deviation of absolute amino acid costs is distinctly smaller between heterotrophic day and night conditions (see Supplemental Data 3), the rank correlations are very large for all models and scenarios. Moreover, the correlations are stronger for heterotrophic conditions versus night conditions (Table 6.3; the only exception is the setting of equal uptake of nitrogen sources for the model of Arnold and Nikoloski [38]). These costs are similarly downscaled compared to autotrophic conditions. The reason for the two findings is the equal energy and carbon source usage, namely Glc, in both scenarios. For the models of de Oliveira Dal'Molin *et al.* [186], as well as of Arnold and Nikoloski [38], the synthesis costs pertaining to night conditions of almost all amino acids are higher than those for heterotrophic day conditions (Figure 6.1). We observed an increase of Glc import accompanied by increased ATP surplus production at night indicating that the night-specific restrictions on metabolism result in a more inflexible system. In contrast, the costs based on the model of Poolman *et al.* [104] are smaller at night than during the day. This is caused by the decreased capability of the system to produce ATP from Glc (ATP production efficiency with respect to Glc) at night, namely, 28 ATP/Glc compared to 31 ATP/Glc at day. Hence, utilizing a specific amount of Glc for amino acid synthesis leads to a smaller sacrifice of potential ATP production at night compared to utilizing the same amount at day. This effect likely outweighs the increase in costs due to additional inflexibilities at night. The change in the ATP production efficiency with respect to Glc could not be justified by biologically meaningful interpretation and renders the costs based on the model of Poolman *et al.* [104] questionable. Such discrepancy of ATP production efficiency was not detected for the other two models.

The amino acid synthesis under autotrophic conditions requires, in general, more ATP than under heterotrophic conditions (Figure 6.2). The only exception is glycine synthesis which is marginally smaller in the model of Arnold and Nikoloski [38] for arbitrary nitrogen uptake. The reason for the increased costs under autotrophic conditions is the type of carbon source: starting from inorganic CO_2 instead of utilizing high-energetic Glc requires additional, energy-demanding assimilation/fixation reactions to synthesize common precursors of the amino acids, *e.g.*, 3-phosphoglycerate or pyruvate. In addition, the amount of ATP which is released by Glc breakdown to arrive at these precursors results in lower costs.

Importantly, the impact of costs can differ across scenarios, in particular for autotrophic
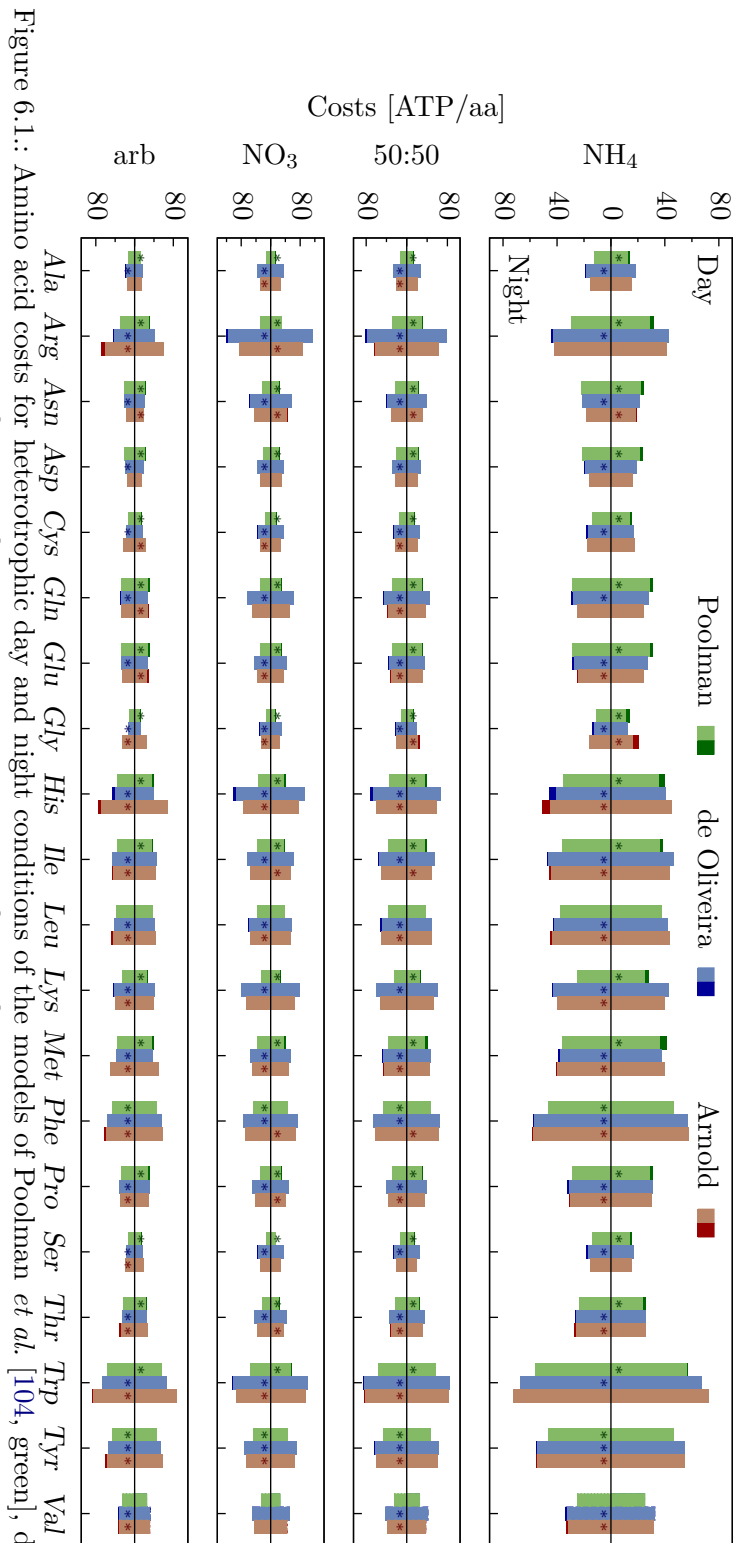
Figure 6.1.: Amino acid costs for heterotrophic day and night conditions of the models of Poolman *et al.* [104, green], de Oliveira Dal'Molin *et al.* [186, blue], and Arnold and Nikoloski [38, red]. The light-colored bars represent the shared cost fraction of day and night, whereas the dark bars denote the cost difference. The scenario with the higher costs is additionally denoted by the star above the bar.
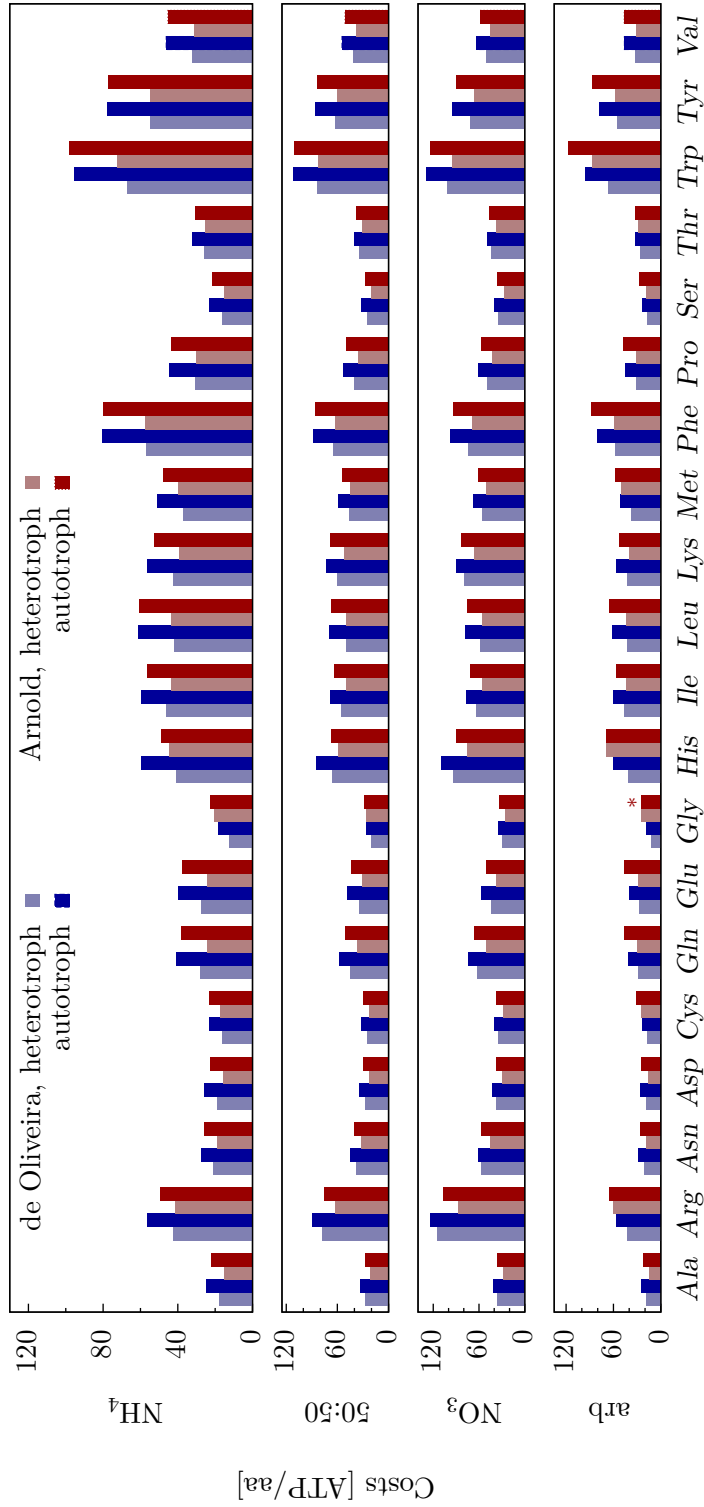
Figure 6.2.: Amino acid costs for heterotrophic (light) and autotrophic day (dark) conditions of the models of de Oliveira Dal'Molin *et al.* [186, blue], and Arnold and Nikoloski [38, red]. The star denotes the exceptions for which autotrophic costs are higher than the corresponding heterotrophic costs.

Table 6.3.: Association of amino acid costs pertaining to the individual models with respect to day-night shift and trophic level determined by Kendall rank correlation τ (p-value < 0.05 in all cases). In particular, the scenarios are heterotrophic and autotrophic day (Het and Aut), and night (Nig). The examined nitrogen source scenarios are sole NH_4 supply, equal uptake of NH_4 and NO_3 (50:50), sole NO_3 supply, and arbitrary nitrogen uptake (arb).

τ	Poolman				de Oliveira				Arnold			
	NH_4	50:50	NO_3	arb	NH_4	50:50	NO_3	arb	NH_4	50:50	NO_3	arb
Het:Nig	0.94	0.94	0.94	0.94	0.96	1	1	0.96	0.97	0.91	0.98	0.95
Aut:Het	–	–	–	–	0.94	0.96	0.92	0.94	0.87	0.92	0.96	0.89
Nig:Aut	–	–	–	–	0.96	0.96	0.92	0.96	0.88	0.93	0.96	0.86

and heterotrophic cells. A shift in resource availability, *e.g.*, of the respective energy source, may be accompanied by a considerable change of the total cellular ATP production, which, in turn, may outweigh the effects of higher costs. For instance, while in autotrophic cells the accessible amount of energy source per cell is physiologically restricted by the leaf surface, which also limits the total cellular ATP budget, the utilization of Glc in heterotrophic cells is limited by different mechanisms, which might affect the ATP budget. Therefore, conclusions cannot be drawn generally without accounting for specifics of the investigated condition.

6.2.3. Effects of different nitrogen sources

The importance of an adequate nitrogen supply in the form of NO_3 and/or NH_4 is well-documented for optimal plant growth [229, 265]. For instance, Masakapalli *et al.* [265] have shown experimentally that amino acid and protein levels vary with the nitrogen source, although the measured net depletion of nitrogen from the medium was found to be of almost the same rate. To examine whether the effects of nitrogen source on the absolute amino acid levels could be due to supply-dependent differences in synthesis costs, we set up and investigate four nitrogen scenarios by adjusting the nitrogen uptake: (i) sole NO_3 supply, (ii) sole NH_4 supply, (iii) equal uptake of the two nitrogen sources, NO_3 and NH_4, and (iv) arbitrary uptake of the two nitrogen sources, which results in a self-adjusting uptake ratio towards optimal energy source utilization.

The comparison of the respective amino acid synthesis costs demonstrates that the scenarios where nitrogen is provided exclusively as NO_3 or NH_4 correspond to the extremes of costs. It is important to point out that the costs for the model of Poolman *et al.* [104] do not change across the nitrogen uptake scenarios (Tables 6.2 and 6.3). Consequently, the following comparison is based on the models of de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38]. The sole supply of NH_4 results in the lowest amino acid synthesis costs across all scenarios. In contrast, omitting NH_4 in favor of NO_3 requires the largest amount of ATP for most conditions. Only the costs of glycine synthesis in the case of sole NO_3 supply for the model of Arnold and Nikoloski [38] are marginally smaller compared to the equal uptake of both nitrogen sources (Supplemental Data 3). The experimental finding that decreased levels of amino acid and protein levels are detected for

sole NO_3 supply compared to mixed nitrogen supply [265] gives rise to the hypothesis that higher costs of amino acid synthesis result in lower synthesis rates. As mentioned above, the net depletion of nitrogen was observed at almost the same rate for the two scenarios, indicating that the nitrogen taken up by the plant is stored in another form, most likely as free nitrate in the vacuole [229]. Accordingly, the assimilated carbon (or Glc) which is not allocated to amino acid synthesis can be invested into other growth-related synthesis processes including, amongst others, the synthesis of sugars and organic acids. This is in accordance with the observed increased growth rate and elevated levels of sugars and organic acids under sole NO_3 supply [265, 266].

To verify and further analyze the relation between (minimum) amino acid costs and nitrogen supply, we next examined the costs of amino acid synthesis for arbitrary nitrogen uptake. For the model of de Oliveira Dal'Molin *et al.* [186], as expected, the costs and the corresponding nitrogen uptake match those of the minimum cost scenario, namely with sole NH_4 supply. Interestingly, this is not the case for all amino acids in the model of Arnold and Nikoloski [38] (see Supplemental Data 3). More specifically, the amino acids whose synthesis exceeds the minimum costs were, in fact, those whose synthesis exhibits the utilization of both NO_3 and NH_4. This is due to the minimization of the respective energy source uptake in the evaluation of amino acid synthesis costs (see Methods). In the case of the autotrophic scenario, the light reactions enable the utilization of light energy to generate NADPH and ATP. Thereby, both energy equivalents are produced in a fixed ratio, which is 7:9 (NADPH:ATP) for the model of Arnold and Nikoloski [38]. The final transfer of the electrons to NADP is catalyzed by ferredoxin-NADP reductase (FNR; EC 1.18.1.2) which oxidizes ferredoxin. Well-established (but still highly discussed) is the hypothesis of the competition of FNR and nitrite reductase (NiR; EC 1.7.7.1) for reduced ferredoxin [273]. Accordingly, the assimilation of NO_3 would relax the fixed NADPH:ATP ratio, since the reduced ferredoxin produced by photosystem I (EC 1.97.1.12) can be diverted to nitrite reduction via NiR. Following this line of argumentation, the assimilation of NO_3, ultimately, allows a more efficient usage of the energy source, light, but at higher costs, in terms of ATP. Similar argumentation holds for heterotrophic day and night conditions with respect to the fixed NADH:ATP ratio from Glc breakdown. Here, the FNR operates in the reverse direction such that NADPH reduces ferredoxin which is diverted to nitrite reduction. In accordance with the hypothesis, this also results in a relaxation of the NADH:ATP ratio, since NADH can be easily converted to NADPH and can result in more efficient utilization of the energy source Glc.

In contrast, the light reactions in the model of de Oliveira Dal'Molin *et al.* [186] are lumped to an artificial reaction which, as stated by the authors, describes how "ATP and NADPH [are] generate[d] from light (overall reaction in chloroplast)". However, to ensure the assimilation of NO_3, the FNR reaction is required separately, thus rendering the light reactions and NO_3 assimilation independent. Accordingly, the above discussed scenario of activating NO_3 assimilation to relax the fix NADPH:ATP ratio has no effect for the model of de Oliveira Dal'Molin *et al.* [186]. On the other hand, several studies have shown that optimal plant growth is usually achieved if nitrogen can be acquired in both forms, NO_3 and NH_4, from the soil [267, 268]. Therefore, our findings based on the model of Arnold and Nikoloski [38] for arbitrary nitrogen uptake have experimental support. It is noteworthy to emphasize that these findings were obtained by only exploiting the stoichiometry of

Arabidopsis' metabolism.

As mentioned above, the costs of amino acid synthesis increase with the NO_3 fraction of nitrogen uptake. Interestingly, this relation seems not to occur in a strictly proportional manner. For the majority of amino acids the deviation from the averaged nitrogen costs, $D_N^{av} = \frac{1}{2}(C_{NO_3} + C_{NH_4}) - C_{50:50}$ differs from zero (Figure 6.3). The reason behind is the change of the nitrogen uptake ratio of NO_3 and NH_4 which affects predominantly the operation of nitrate reductase (NR; EC 1.7.1.1) and NiR. Therefore, we expected that reactions linked via the other reactants of NR and NiR, namely ferredoxin and NAD(P)H, are also strongly affected. Indeed, we found a reciprocal relation of NiR to the competing FNR such that less NADPH is produced if NO_3 is assimilated, as discussed above. In addition, NADPH (or its unphosphorylated equivalent, NADH) is needed as substrate of the NR such that NO_3 assimilation induces further deficiency of NADPH. Nevertheless, the complexity of the metabolic networks precludes addressing the exact implications on the individual costs. The exceptions to the nonzero deviation include the costs under autotrophic conditions of the model of de Oliveira Dal'Molin *et al.* [186] which are, as stated above, incapable to reveal/point out the competition of carbon fixation and NO_3 assimilation.
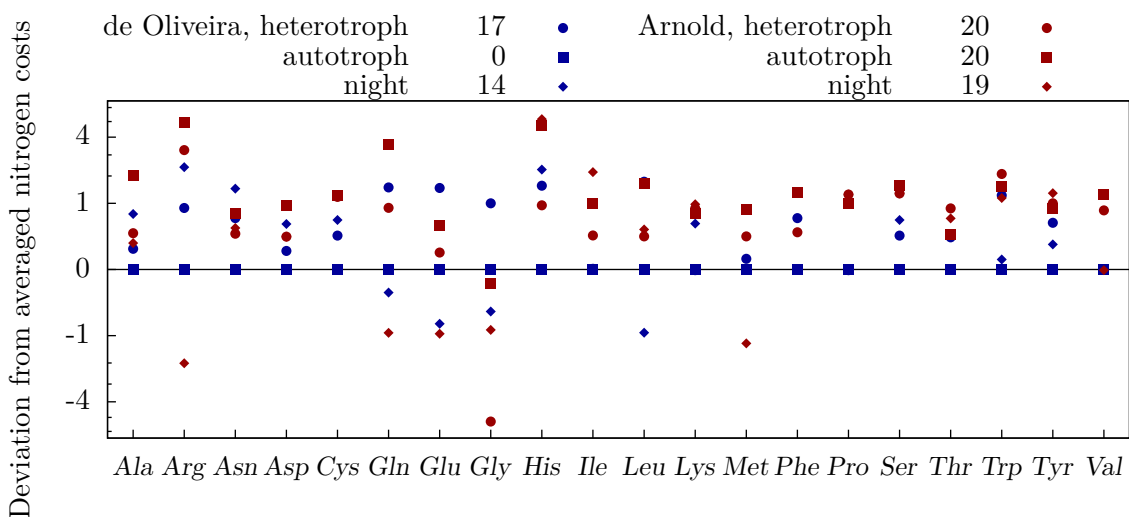

Figure 6.3.: Deviation from the averaged nitrogen costs for the models of de Oliveira Dal'Molin *et al.* [186, blue], and Arnold and Nikoloski [38, red]. The values given in the legend denote the numbers of amino acids whose deviation differs from zero.

6.2.4. Conclusions and implications

While, by definition, a model is never an exact reproduction of reality, the applied assumptions should affect the properties of interest in the least possible way. The quality of a metabolic reconstruction strongly influences the model predictions since model artifacts and errors could bias the results. By examining the impact of three published models of primary

Arabidopsis metabolism [47, 104, 186], we ascertained the model of Poolman *et al.* [104] to be unsuitable for the determination of amino acid costs. The lack of compartmentalization, amongst others, is a major source of bias in that case.

To obtain biologically meaningful results, the employed modeling approach should take into account the relevant environmental conditions. Here, we focused on the influence of day and night, of the trophic level, as well as of the nitrogen supply, which have been shown to strongly affect amino acid synthesis costs.

The majority of amino acid costs for night conditions are higher than those under heterotrophic day conditions. In addition, the costs under heterotrophic day conditions are distinctly exceeded by the amount of required ATP for synthesis under autotrophic conditions (Figure 6.4). Nevertheless, it is worth pointing out that positing hypotheses about resource reallocation in plants based on comparison of these costs may be misleading. The impact of metabolic costs in terms of energy depends on the availability of the respective resources (in the sense of environmental supply, *e.g.*, light intensity, as well as the ability to use this supply, *e.g.*, leaf surface). A limitation of resource availability imposes constraints on the energy budget. The effect of metabolic costs then depends on their ratio to the energy budget. Consequently, an environmental shift accompanied with a change in resource limitation may not only alter the metabolic costs but also the total energy budget (and, hence, the effect of the costs).
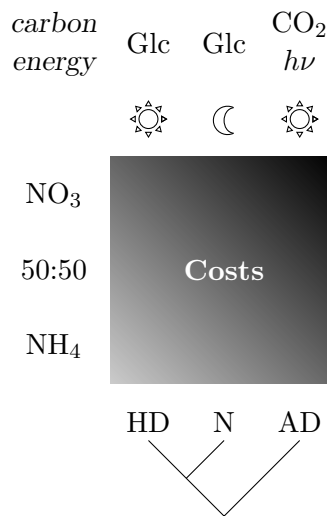

Figure 6.4.: Schematic representation of the different investigated scenarios, namely, autotrophic and heterotrophic day (AD and HD) as well as night (N), and the trend with respect to the corresponding amino acid synthesis costs. The gradient represents the costs trend whereby lighter gray denotes low costs.

Understanding the patterns of resource reallocation as a result of a shift in conditions requires the calculation of metabolic costs and, in addition, the determination of changes to the limiting resource budget. Such an analysis would undoubtedly require a clear analysis of time intervals which a resource budget may account for. Shifts between autotrophic and heterotrophic day as well as night conditions are likely to be accompanied by considerable changes of the energy budget of individual cells due to the switch of energy source.

Nevertheless, in this case, a constant energy budget (per cell) may still exist as a result of a coordinated balance between energy source utilization and growth rate. This seems to be in contrast to the fact that amino acids appear to be predominantly synthesized during day [274] which indicates that the energy budget may be larger compared to night. In general, the comparison of costs across trophic levels and species warrants caution and careful biological interpretation.

The supplied nitrogen source extensively affects amino acids costs, exhibiting lowest costs for sole NH_4 and highest costs for sole NO_3 supply, in the models of de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38] (Figure 6.4). Our results, which are based only on the stoichiometric information of Arabidopsis' metabolism, are in accordance with experimental findings. The increased costs of amino acid synthesis for sole NO_3 supply compared to sole NH_4 supply offer an explanation for the decreased amino acid and protein levels observed in plants grown under these conditions [265]. This is in line with former calculations showing increased production of enzymes for plants grown with NH_4 as sole nitrogen source compared to NO_3 as sole nitrogen source [100]. Furthermore, the elevated efficiency in energy usage under mixed nitrogen supply suggests a stoichiometric explanation of mixed uptake of NH_4 and NO_3 for optimal plant growth which has been observed in many plant species if both nitrogen sources are supplied [267, 268]. Certainly, our *in silico* analysis may not provide the complete, biological explanation but provides starting points for further experimental studies.

The utilization of resources often limits growth, even under optimal conditions [275]. Thus, particular resource allocation towards individual processes point at cellular strategies to ensure plant survival and reproduction. As protein synthesis, and therefore amino acid synthesis, consumes a major part of a plant's available energetic resources [251], it is expected to play a considerable role in trade-offs arising from energetic limitation. Biological reasonable attribution of metabolic costs is a prerequisite to understand these trade-offs and their implications for plant growth. As it has also been recently argued, metabolic costs (in terms of ATP) of ribulose 1,5-bisphosphate carboxylase oxygenase turnover represent a major investment in plants Arnold and Nikoloski [38]. Therefore, the determination of condition-specific amino acid synthesis costs may serve as a starting point for metabolic engineering in plants towards improved growth and yield in agricultural applications. This could be achieved by manipulation of amino acid sequences of individual enzymes as well as of utilization of whole pathways towards higher efficiency of energetic investments and, therefore, improved growth. However, understanding the energetic trade-offs may not only be exploited by using individually cheaper or more efficient pathways, but also by reorganization of metabolism. We believe that understanding these trade-offs is a first step towards such holistic manipulations in plants.

6.3. Methods

6.3.1. Model modifications

To enable a reliable determination of amino acid costs, minor modifications for each model had to be made. To ensure a comparable framework, we disabled all importer and exporter as well as biomass reactions in the models and enabled only the necessary

importer and exporter in the examination of the individual scenarios (see 6.3.2). In the case of the models of Poolman *et al.* [104] and de Oliveira Dal'Molin *et al.* [186], additional modifications were required to enable the determination of the ATP production efficiency. While de Oliveira Dal'Molin *et al.* [186] provide an updated version of their model (`http://web.aibn.uq.edu.au/cssb/resources/Genomes.html`) which resolves the issue (Table 6.4), no improved version is provided for the model of Poolman *et al.* [104]. In that case, the authors have stated that energy can be generated out of nothing which lead us to the assumption that there exists at least one futile cycle including the ATPase. To identify a futile cycle, we constrained all fluxes to be less than or equal to an arbitrary number, *e.g.*, 1000, and determined the maximum flux through the ATPase. We then consecutively knocked out each reaction of the network to identify those which contribute most and are, accordingly, part of a futile cycle (Algorithm 6.1). We manually gathered the participating reactions, defined a modification to cancel the futile cycle and integrated it in the succeeding iterations of this algorithm (Table 6.4 and 6.5). Altogether, we could identify all futile cycles affecting ATP production in three iterations, whereby analysis of single knockouts was sufficient.

Table 6.4.: Modified model constraints compared to the published model versions of Poolman *et al.* [104] de Oliveira Dal'Molin *et al.* [186] which are valid for all scenarios. Lower and upper boundaries are denoted by v^{min} and v^{max}, respectively.

Model	Reaction ID	v^{min}	v^{min}
Poolman	`reac_1321`	0	Inf
	`reac_1331`	0	Inf
	`MalDH`	0	`v(AlphaKGDH)`
de Oliveira	`TCM22`	0	0
	`TCP27`	0	0
	`TCX14`	- Inf	Inf
	`R00086_c`	0	Inf
	`R00093_p`	0	0
	`R00114_c`	0	0
	`R00149_p`	0	0
	`R00243_c`	0	Inf
	`R00243_m`	0	Inf
	`R00253_m`	0	Inf
	`R00253_p`	0	Inf
	`R00343_m`	0	0
	`R00472_x`	0	Inf
	`R00945_c`	0	0
	`R01221_m`	0	0
	`R01398_c`	0	0
	`R01752_c`	0	0
	`R05875_p`	0	0

For the compartmentalized models, the determined amino acid costs are the minimum costs across all compartments. In the model of Arnold and Nikoloski [38], the amino

---

Algorithm 6.1: Futile cycle detection strategy to disable the unbounded ATP production in the model of Poolman *et al.* [104] (via `ATPase`).

Input:
Metabolic network of Poolman *et al.* [104], including
    stoichiometric matrix, $\mathbf{S}$,
    vector of lower and upper flux boundaries, $\mathbf{v}^{min}$ and $\mathbf{v}^{max}$
Output: Set of candidate reactions for participating in a futile cycle, r

Definition: Let d be the revised reactions of already detected futile cycles, with flux boundaries $v_d^{min} = d^{min}$ and $v_d^{max} = d^{max}$.

Begin
> Restrict the flux boundaries of each reaction except for the maintenance reaction `ATPase`, v_m:
>
> $$v_i^{min} = \begin{cases} 0 & \text{if } i \in \{\text{'irreversible'}, m\} \\ -1000 & \text{otherwise.} \end{cases} \qquad v_i^{max} = \begin{cases} \text{Inf} & \text{if } i = m \\ 1000 & \text{otherwise.} \end{cases}$$
>
> Disable all import and export reactions except for O_2 import and CO_2 export
> Save current reaction boundaries as $cond := (\mathbf{v}^{min}, \mathbf{v}^{max})$
> **For** $j = 1$ **to** *'total number of reactions'* N **do**
> > Reset reaction constraints according to *cond*
> > Assign the flux of Glc import reaction, v_{Glc}, to one,
> > Disable the respective reaction j,
> > Maximize the flux through the maintenance reaction `ATPase`, v_m:
> >
> > $$z = \max \mathbf{c}^T \mathbf{v} \ , \qquad c_i = \begin{cases} 1 & \text{if } i = m \\ 0 & \text{otherwise.} \end{cases}$$
> >
> > $$\begin{aligned} s.t. \quad & \mathbf{S} \cdot \mathbf{v} = 0 \quad &\text{(LP1)}\\ & v_{\text{Glc}}^{min} = v_{\text{Glc}}^{max} = 1 \\ & v_j^{min} = v_j^{max} = 0 \\ & \mathbf{v}^{min} \leq \mathbf{v} \leq \mathbf{v}^{max} \\ & d^{min} \leq v_d \leq d^{max} \end{aligned}$$
>
> **End**
> Determine the set of reactions whose deactivation results in the highest deviation
>
> $$r = \arg \min_j z \ , \qquad r \in \{1, \ldots, 1406\}$$

End

---

acid synthesis pathways are highly compartmentalized and, additionally, only a minimum number of transport reactions across the compartments exist. Therefore, for those amino acids with multiple synthesis pathway localizations, we determined compartment-specific costs. In contrast, for the model of de Oliveira Dal'Molin *et al.* [186], all amino acid synthesis pathways are either completely located in the cytosol or transport reactions into the cytosol exist. Accordingly, by implementing export reactions for each amino acid from the cytosol, the entire set of possible amino acid costs were covered. However, the accuracy of costs, in this case, depends on the biological plausibility of these transport reactions.

In the model of Arnold and Nikoloski [38], there exist three maintenance reactions, one in each compartment (chloroplast, mitochondrion, and cytosol including the peroxisome). We ascertained that the maximum ATP production efficiency for the model can be achieved only if the sum of all maintenance reactions is taken into account. Accordingly, we checked whether similar effects can arise for the model of de Oliveira Dal'Molin *et al.* [186] by incorporating maintenance reactions for each compartment. In doing so, we found the same ATP production efficiency with and without the additional reactions (data not shown). Here, we used the original setup without an additional maintenance reaction. The impact of an additional maintenance reaction for the model of Poolman *et al.* [104] could be disregarded due to the lack of compartmentalization.

6.3.2. Environmental and cellular set-up

A large percentage of plant metabolism is largely affected by the change of day and night, which we expected to similarly affect amino acid biosynthesis. In C_3 plants, the predominant processes during the day include carbon fixation, photorespiration, and starch assimilation, while at night these are starch degradation and cellular respiration (also referred to as dark respiration). The majority of metabolism-related changes between day and night are attributed to the redox regulation of enzymes in the Calvin-Benson cycle, starch synthesis, ATP synthesis, and NADPH export from chloroplasts in response to light [269–272]. As structural metabolic modeling approaches cannot readily incorporate regulatory processes, without making too many simplifying assumptions, we simulate the regulatory effect as an activation and deactivation of the corresponding reactions, respectively. More precisely, we manipulated the ribulose-1,5-bisphosphate carboxylase oxygenase (EC 4.1.1.39), the plastidic glyceraldehyde-3-phosphate dehydrogenases (GAPDH; EC 1.2.1.12 and 1.2.1.13), the plastidic fructose-1,6-bisphosphatase (FBPase; EC 3.1.3.11), sedoheptulose-1,7-bisphosphatase (EC 3.1.3.37), the phosphoribulokinase (EC 2.7.1.19), the ADP-glucose pyrophosphorylase (EC 2.7.7.27), the ATP synthase (EC 3.6.3.14), the NADP-linked malate dehydrogenase (EC 1.1.1.82), and the phenylalanine ammonia-lyase (EC 4.3.1.24), the glucose 6-phosphate dehydrogenase (EC 1.1.1.49), and the FNR (EC 1.18.1.2). The details of the respective condition-specific modifications are given in Table 6.6. In addition, we accounted for the regulatory effect of fructose-2,6-bisphosphate on the inter-conversion of fructose-6-phosphate and fructose-1,6-bisphosphate in the cytosol (Figure 6.5; [276, 277]). From the modeling perspective, this results in a light dependent manipulation of the cytosolic FBPase and the pyrophosphate-dependent phosphofructokinase, respectively (see Table 6.6).

Moreover, we investigated the impact of the trophic level, in particular autotrophic and heterotrophic growth conditions. For autotrophic scenarios, in general, the minimum nutri-
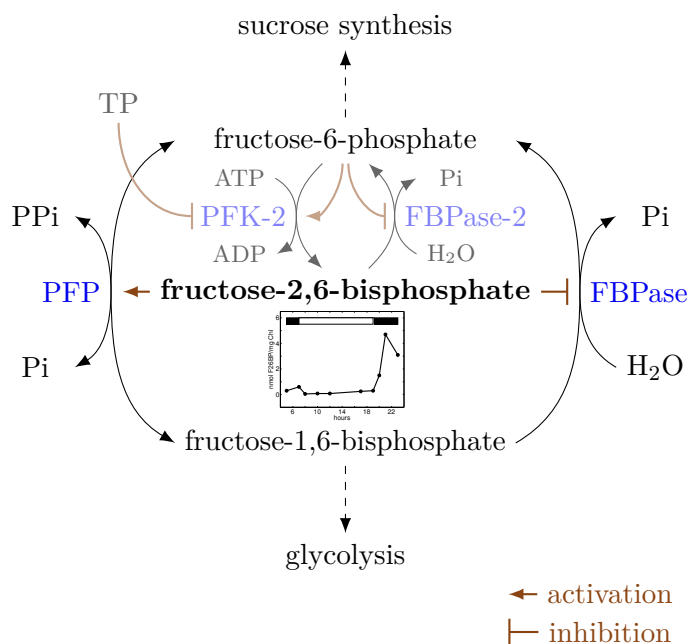
Figure 6.5.: Schematic representation of the regulatory mechanism of fructose-2,6-bisphosphate (F26BP) on the pyrophosphate-dependent phosphofructokinase (PFP) and the cytosolic fructose-1,6-bisphosphatase (FBPase) according to Nielsen *et al.* [276] and Stitt [277]. The regulators of F26BP itself are presented light-colored. The inlay graphic shows the regulatory effects of day-night alternation on the levels of F26BP (after Sicher *et al.* [280]).

ent requirements are the conventional low energy precursors water, CO_2, Pi, NO_3 and/or NH_4, SO_4 and/or H_2S in addition to light as energy source. In contrast, in heterotrophic scenarios the carbon source, here Glc, serves simultaneously as energy source. In that case, to preclude the carbon fixation and utilization of light energy, we disabled the photosystem reactions and deactivated the CO_2 import (Table 6.6). Furthermore, for simulating amino acid biosynthesis, the number of precursors can be further reduced. On the one hand, Pi is not required and, on the other hand, the model de Oliveira Dal'Molin *et al.* [186] cannot handle SO_4 such that the respective importers are disabled. As a result, to enable a fair comparison, for the model of Poolman *et al.* [104] the supply of H_2S had to be allowed by implementing the respective importer (Table 6.6).

6.3.3. Amino acid cost calculation

The calculation of amino acid synthesis costs is based on flux balance analysis, an approach to determine flux distributions optimizing a specific metabolic function, *e.g.*, biomass production, at steady state. The optimal flux distributions are determined by linear programming and have usually to obey specific limitations of nutrient uptake [28]. The cost calculation requires importers of nutrients and exporters of amino acids. Additionally, our approach requires a maintenance reaction (*i.e.*, a generic ATP hydrolysis reaction,

ATP → ADP + Pi) which accounts for generic energy consuming processes. For the models in which any of these reactions is not present, it was additionally included. The model of Arnold and Nikoloski [38] comprises a maintenance reaction for every compartment. Therefore, the flux through the maintenance reaction is in fact the sum of fluxes carried by all three compartment-specific maintenance reactions. Constraints according to the examined model, the environmental conditions and the trophic level are also imposed, and are referred to as initial constraints (shown in Table 6.6).

The amino acid cost calculation proceeds in three steps. We assume that nutrient import, amino acid export and the maintenance reaction(s) are such that they carry nonnegative flux. First, we determine the ATP production efficiency, as described in Algorithm 6.2 (LP2): the initial constraints are imposed, the energy source import, v_e, is constrained to 1 and the flux through the maintenance reaction(s) is maximized. The ATP production efficiency gives the maximum number of ATP molecules which can be produced per imported unit of energy source. It has the dimension [ATP molecules/energy source unit]. The ATP production efficiency is a property of the network and the initial constraints.

The amino acid cost calculation is then executed according to the two following steps, which are repeated for every amino acid.

After the initial constraints have been restored, the total ATP cost, C_j^{tot}, is determined, as described in Algorithm 6.2 (LP3): the export of the considered amino acid is set to 1 and the import of the energy source is minimized. Multiplying the minimal energy source import per amino acid export with the ATP production efficiency of the energy source yields the total ATP cost, C_j^{tot}, of amino acid synthesis with dimension [ATP molecules/amino acid molecule]. Therefore, the total ATP cost quantifies the maximum amount of ATP that could be produced from the energy source instead of synthesizing the amino acid using the pathway with minimum energy source requirement.

In the third step, we determine the surplus ATP cost, C_j^{sur}, as described in Algorithm 6.2 (LP4): we additionally fix the (minimum) energy source import and maximize the flux through the maintenance reaction(s). The ratio of the flux through the maintenance reaction(s) and the amino acid import yields the surplus ATP cost, C_j^{sur}, with dimension [ATP molecules/amino acid molecule].

The amino acid synthesis cost then is given as the difference of the total and the surplus ATP cost.

It should be noted that the amino acid costs depend on the energy source uptake or on the amino acid export flux, if nonzero boundaries of other reactions limit the outcome of the optimization in any of the three steps. This is usually not the case in the original setup of metabolic reconstructions (like in this study), but can occur when additional boundary conditions are incorporated, *e.g.*, by utilizing high-throughput data [281].

Algorithm 6.2: Three-step optimization for condition-specific amino acid cost computation.

WLOG:

Import and export reactions are assigned as left-to-right operating such that the transport has positive flux

Input:

Metabolic network, including

stoichiometric matrix, $\mathbf{S}$,

vector of lower and upper flux boundaries, $\mathbf{v}^{min}$ and $\mathbf{v}^{max}$

Initial constraints, *cond* (environment and model specific)

Output: Condition-specific amino acid costs, $\mathbf{C}_{cond}$

Begin

(De)Activate reactions according to *cond*,

Assign the flux of the corresponding energy source import reaction, v_e, to one,

Maximize the flux through the maintenance reaction(s), v_m:

$$z_1 = \max \mathbf{c}^T \mathbf{v} , \qquad c_i = \begin{cases} 1 & \text{if } i \in m \\ 0 & \text{otherwise.} \end{cases}$$

$$s.t. \quad \mathbf{S} \cdot \mathbf{v} = 0 \qquad \qquad \text{(LP2)}$$

$$v_e^{min} = v_e^{max} = 1$$

$$\mathbf{v}^{min} \leq \mathbf{v} \leq \mathbf{v}^{max}$$

Define *ATP production efficiency* as $E_{\text{ATP}} = \dfrac{z_1}{v_e}$

For $j = $ *'Ala'* **to** *'Val'* **do**

Reset reaction constraints according to *cond*,

Assign the flux of the respective amino acid export reaction, v_j, to one,

Minimize the flux through the corresponding energy source import reaction, v_e:

$$z_2 = \min \mathbf{c}^T \mathbf{v} , \qquad c_i = \begin{cases} 1 & \text{if } i = e \\ 0 & \text{otherwise.} \end{cases}$$

$$s.t. \quad \mathbf{S} \cdot \mathbf{v} = 0 \qquad \qquad \text{(LP3)}$$

$$v_j^{min} = v_j^{max} = 1$$

$$\mathbf{v}^{min} \leq \mathbf{v} \leq \mathbf{v}^{max}$$

Define *total ATP cost* as $C_j^{\text{tot}} = \dfrac{z_2}{v_j} \cdot E_{\text{ATP}}$,

Assign the flux of the corresponding energy source import reaction, v_e, to z_2,

Maximize the flux through the maintenance reaction(s), v_m:

$$z_3 = \max \mathbf{c}^T \mathbf{v} , \qquad c_i = \begin{cases} 1 & \text{if } i \in m \\ 0 & \text{otherwise.} \end{cases}$$

$$s.t. \quad \mathbf{S} \cdot \mathbf{v} = 0 \qquad \qquad \text{(LP4)}$$

$$v_j^{min} = v_j^{max} = 1$$

$$v_e^{min} = v_e^{max} = z_2$$

$$\mathbf{v}^{min} \leq \mathbf{v} \leq \mathbf{v}^{max}$$

Define *surplus ATP cost* as $C_j^{\text{sur}} = \dfrac{z_3}{v_j}$,

Set $C_{cond}(j) = C_j^{\text{tot}} - C_j^{\text{sur}}$

End

End

Table 6.5.: Identified futile cycles in the model of Poolman *et al.* [104]. The reactions marked with '>' were detected by the Algorithm 6.1, and the reactions highlighted in gray are the revised reactions (the respective adaption is given in Table 6.4). '∘' denotes reactions which are presented in a reversed manner compared to their implemented annotation (reads now right-to-left). The metabolite annotation is in accordance to that of Poolman *et al.* [104].

Reaction ID	Formula	
reac_374	$2\,\mathrm{GSH} + \mathrm{NADP} \rightarrow \mathrm{NADPH} + \mathrm{GSSG}$	
reac_33	$\mathrm{AMP} + \mathrm{SO}_3 + \mathrm{GSSG} \rightleftharpoons 2\,\mathrm{GSH} + \mathrm{APS}$	∘
reac_1331	$\mathrm{PPi} + \mathrm{APS} \rightleftharpoons \mathrm{SO}_4 + \mathrm{ATP}$	∘
reac_1317	$2\,\mathrm{Cyt}c_\mathrm{red} + \mathrm{SO}_4 \rightarrow 2\,\mathrm{Cyt}c_\mathrm{ox} + \mathrm{SO}_3$	
> reac_267	$2\,\mathrm{Cyt}c_\mathrm{ox} + \mathrm{Lac} \rightleftharpoons 2\,\mathrm{Cyt}c_\mathrm{red} + \mathrm{Pyr}$	
> reac_302	$\mathrm{Pyr} + \mathrm{NADH} \rightleftharpoons \mathrm{NAD} + \mathrm{Lac}$	∘
1.	$\mathrm{AMP} + \mathrm{PPi} + \mathrm{NADH} + \mathrm{NADP} \rightarrow \mathrm{ATP} + \mathrm{NAD} + \mathrm{NADPH}$	
reac_1260	$\mathrm{NADH} + \mathit{Glu} \rightleftharpoons \mathrm{NAD} + \mathrm{P5C}$	∘
> reac_1321	$\mathrm{P5C} \rightleftharpoons \mathit{Glu}\text{-SeA}$	∘
reac_379	$\mathit{Glu}\text{-SeA} + \mathrm{NADP} + \mathrm{Pi} \rightleftharpoons \mathrm{NADPH} + \mathit{Glu}\mathrm{P}$	
reac_378	$\mathit{Glu}\mathrm{P} + \mathrm{ADP} \rightleftharpoons \mathit{Glu} + \mathrm{ATP}$	∘
2.	$\mathrm{ADP} + \mathrm{Pi} + \mathrm{NADH} + \mathrm{NADP} \rightleftharpoons \mathrm{ATP} + \mathrm{NAD} + \mathrm{NADPH}$	
FumEquiv	$\mathrm{FUM} \rightleftharpoons \mathrm{Fum}$	∘
Fumarase	$\mathrm{Fum} + \mathrm{H}_2\mathrm{O} \rightleftharpoons \mathrm{Mal}$	
> MalDH	$\mathrm{Mal} + \mathrm{NAD}_\mathrm{m} \rightleftharpoons \mathrm{OxalAc} + \mathrm{NADH}_\mathrm{m} + \mathrm{H}_\mathrm{m}$	
OAAEquiv	$\mathrm{OxalAc} \rightleftharpoons \mathrm{OAA}$	
reac_197	$\mathrm{OAA} + \mathit{Glu} \rightleftharpoons \mathrm{KG} + \mathit{Asp}$	∘
reac_363	$\mathrm{NH}_3 + \mathrm{KG} + \mathrm{NADH} \rightleftharpoons \mathrm{NAD} + \mathit{Glu}$	∘
reac_158	$\mathrm{GTP} + \mathrm{IMP} + \mathit{Asp} \rightarrow \mathrm{GDP} + \mathrm{DC\text{-}AMP} + \mathrm{Pi}$	
reac_180	$\mathrm{DC\text{-}AMP} \rightarrow \mathrm{FUM} + \mathrm{AMP}$	
reac_178	$\mathrm{AMP} \rightarrow \mathrm{NH}_3 + \mathrm{IMP}$	
reac_347	$\mathrm{GDP} + \mathrm{ATP} \rightleftharpoons \mathrm{GTP} + \mathrm{ADP}$	
	$\mathrm{ATP} + \mathrm{H}_2\mathrm{O} + \mathrm{NADH} + \mathrm{NAD}_\mathrm{m} \rightarrow \mathrm{ADP} + \mathrm{Pi} + \mathrm{NAD} + \mathrm{NADH}_\mathrm{m} + \mathrm{H}_\mathrm{m}$	
Complex_I	$\mathrm{NADH}_\mathrm{m} + \mathrm{Q} + 5\,\mathrm{H}_\mathrm{m} \rightarrow \mathrm{NAD}_\mathrm{m} + \mathrm{QH}_2 + 4\,\mathrm{H}$	
Complex_III	$\mathrm{QH}_2 + 2\,\mathrm{Cyt}c_\mathrm{ox} + 2\,\mathrm{H}_\mathrm{m} \rightarrow \mathrm{Q} + 2\,\mathrm{Cyt}c_\mathrm{red} + 4\,\mathrm{H}$	
> Complex_IV	$4\,\mathrm{Cyt}c_\mathrm{red} + 8\,\mathrm{H}_\mathrm{m} + \mathrm{O}_2 \rightarrow 4\,\mathrm{Cyt}c_\mathrm{ox} + 2\,\mathrm{H}_2\mathrm{O} + 4\,\mathrm{H}$	
> Complex_V	$4\,\mathrm{H} + \mathrm{ADP} + \mathrm{Pi} \rightarrow 4\,\mathrm{H}_\mathrm{m} + \mathrm{ATP}$	
3.	$\frac{3}{2}\,\mathrm{ADP} + \frac{3}{2}\,\mathrm{Pi} + \frac{1}{2}\,\mathrm{O}_2 + \mathrm{NADH} \rightarrow \frac{3}{2}\,\mathrm{ATP} + \mathrm{NAD}$	

Table 6.6.: Condition-specific model boundary constraints. As the model of Poolman *et al.* [104] is only capable to simulate the heterotrophic day and the night scenario, the entries for the autotrophic day scenario pertain exclusively to the models of de Oliveira Dal'Molin *et al.* [186], and Arnold and Nikoloski [38]. Due to the incapability of the model of de Oliveira Dal'Molin *et al.* [186] to utilize SO_4, a H_2S importer is implemented for the model of Poolman *et al.* [104] ([a]), and the SO_4 importer is disabled ([b]). The unnecessary import of Pi for amino acid synthesis is deactivated ([c]).

Reaction		Models						Day		Night	Refs
		Arnold		de Oliveira		Poolman		auto	hetero		
Importer and exporter	Glucose	Ex_Glc	←	Ex13	←	GLC_tx	→	−	+	+	
	Photon	Im_hnu	→	Ex16	→	−		+	+	−	
	CO_2	Im_CO2	⇌	Ex1	⇌			+	−	−	
			←			CO2_tx	← / →	−	+	+	
	H_2O	Im_H2O	⇌	Ex2	⇌	−		+	+	+	
	O_2	Ex_O2	⇌	Ex3	⇌	O2_tx	⇌	+	+	+	
	NO_3	Im_NO3	→	Ex4	→	NO3_tx	→	+	+	+	
	NH_4	Im_NH4	→	Ex5	←	NH3_tx	→	+	+	+	
	H_2S	Im_H2S	→	Ex11	→	Im_H2S [a]	→	+	+	+	
	SO_4 [b]	Im_SO4	→	Ex12	→	SO4_tx	→	−	−	−	
	Pi [c]	Im_Pi	→	Ex18	→	Pi_tx	→	−	−	−	
Photosystems		PSII_h, PSI_h	→	REner01_p	→	−		+	−	−	
FNR		Fd-NADPR_h	→					+	−	−	
			⇌			−		−	+	+	[271, 278]
				R01195_p	⇌			+	+	+	
ATPS		ATPase_h	⇌			−		+	+	−	[272]
				REner01_p	→			+	−	−	
RuBisCO		RBC_h, RBO_h	→	R00024_p, R03140_p	→	reac_621	→	+	+	−	[279]
						reac_1136	→	−	+	+	
GAPDH (plast)		GAPDH1_h	→			reac_17	→	+	+	−	[272, 279]
		GAPDH2_h	⇌	R01061_p, R01063_p	⇌			+	+	−	
FBPase		FBPase_h		R00762_p, R04780_p							[272, 279]
			→		→	reac_318	→	+	+	−	
		FBPase_c		R00762_c, R04780_c							[276, 277]
SBPase		SBPase_h	→	R01845_p	→	reac_1305	→	+	+	−	[272, 279]
PRK		Ru5PK_h	→	R01523_p	→			+	+	−	[272]
						reac_562	→	−	+	+	
AGPase		AGPase_h	→	R00948_p	⇌	reac_354	→	+	+	−	[270]
NADP-MalDH (plast)		MalDH3_h	→	R00343_p	⇌	−		+	+	−	[272, 279]
G6PDH		G6PDH_h	→	R02736_p	→	reac_353	→	−	−	+	[272, 279]
PFP		PPIF6PK_c	⇌	−		reac_77	→	−	−	+	[276, 277]
PAL (cyt)		−		R00697_c	⇌	reac_555	→	+	+	−	[279]

Chapter 7.

Concluding remarks and implications

The ideal scenario in metabolic modeling ultimately corresponds to a detailed description of the dynamic behavior of all metabolic processes in a cell. Despite the great research efforts, large-scale kinetic metabolic networks are of limited number and applications, and necessitate extensive parameter estimation [282–284]. Nevertheless, the current state-of-the-art of metabolic modeling has provided substantial insights into the functionality of metabolism.

The decision of whether to use kinetic or stoichiometric modeling, in particular constraint-based modeling, has the largest effect on the findings and predictions resulting from the applied approach. Both approaches have their advantages and shortcomings, which have important implications for the resulting predictions and interpretations. Briefly, kinetic modeling facilitates the detailed investigation of the behavior of a dynamic system based on largely complex mathematical descriptions of the reaction rates. The required kinetic information of these reaction rate descriptions, in turn, restrict the applicability of these models to small to medium scale at best. In contrast, in stoichiometric modeling the complexity of the reaction rate descriptions is reduced to the topology of the underlying system which, in general, prohibits the examination of transitory behavior. On the other hand, this simplification allows for large-scale modeling and, therefore, accounting for the metabolic context. Aside from this distinction, the quasi-steady state approximation can be used in both modeling approaches to simulate the systemic behavior for homeostatic conditions [37].

Irrespective of the selected modeling approach, the quality of the used model is a crucial determinant of the resulting predictions. Shortcomings in the assembly do, as a general rule, drastically reduce the usability of the model and its predictive value. A careful curation and model adjustments offer the means to resolve this issue. Moreover, the above-mentioned limitations can be amplified if the remaining requirements of the particular approach are only partially fulfilled: The availability of experimental data for the reaction rate parameters and the gene-protein-reaction (GPR) associations can be restrictive, and/or the considered cellular scenarios, postulated by the reaction kinetic and steady-state assumption, might not be met.

In this work, various aspects of modeling photosynthesis and the related metabolic processes were illustrated and carefully examined. In doing so, a wide spectrum of biological questions were addressed to reveal new insights into system's behavior, support established opinions for which mounting experimental evidence exists, and to provide novel hypotheses for further directed large-scale experiments. Consistently, the (above-mentioned) crucial points of modeling are pointed out, and solutions for the detected problems were provided.

7.1. Summary

Starting from detailed kinetic investigations, the scope of examining photosynthesis-related processes was extended to analyze the interplay of metabolic processes and to infer the extent of their influence on metabolic functionality.

The proposed ranking of the existing kinetic models of the Calvin-Benson cycle (CBC), in Chapter 2, can be utilized as a guidance for selecting a photosynthesis model suitable for metabolic engineering. The considered criteria included biological functionality (stability, robustness) and ensure reliability (compliance with data). Moreover, the presented classification of the models can serve as an easy reference to determine suitable model components and/or functional structures for different modeling aspects. For instance, the models of reaction kinetics and types of regulatory effects were characterized whereby in some cases the theoretical foundations of the enzyme actions are still lacking. These findings can be used to assemble a better-performing model, with respect to the proposed ranking framework, comprising the presented model features. In order to enable the comparative analysis, amendments and model adjustments were made to obtain properly functioning models for the same initial cellular conditions.

The stoichiometric model of the CBC and related photosynthetic end-product processes, presented in Chapter 3, illustrates the possibility to integrate stoichiometry with partial information about the dynamic behavior of the system. The incorporation of enzymes as 'internal metabolites', their synthesis and their turnover enable the expansion of the metabolic costs by the maintenance costs of enzyme provision. The analysis is based on a stoichiometric version of the best-performing model with respect to metabolic engineering applicability, that of Poolman *et al.* (2000) [60]. Comparing both approaches, the traditional, which does not incorporate enzymes, and the extended, demonstrates that integrating enzyme action provides a purely stoichiometric explanation for what reason starch rather than sucrose is the transitory carbon storage during the day.

The review of existing models considering the photorespiratory pathway, in Chapter 4, illustrates the diverse modeling attempts which largely follow the structure and complexity of the models for photosynthesis-related processes. However, due to its controversial role, photorespiration is often included as an incidental component in the existing models of carbon-centered metabolism. The carbon-centric model paradigm as well as the bridging feature to nitrogen metabolism across multiple compartments was carefully assessed to reveal the demand for investigating the photorespiratory in a larger metabolic context. In doing so, shortcomings of existing attempts were highlighted to reinforce the importance of proper model selection.

The presented, bottom-up assembled model of the primary metabolism of *Arabidopsis thaliana* (Arabidopsis), Chapter 5, has proven to be as productive and effective in simulating growth enhancement, in terms of biomass synthesis, as the existing genome-scale metabolic Arabidopsis models. Moreover, the Arabidopsis core model provides the means for data integration of high quality as the involved metabolic reactions are completely annotated (as far as it is a question of gene-encoded reactions). The three proposed biomass compositions allow environment-specific analyses of the system's behavior, provided that the metabolic changes, such as (de)activation of reactions or entire pathways, are known. Expanding the GPR associations by the enzyme complex structure allows the determination of enzyme *de*

novo synthesis costs, here, exemplified by ribulose-1,5-bisphosphate carboxylase/oxygenase. This provides the means for understanding the resource allocation trade-off between growth, in terms of biomass synthesis, and protein synthesis. Certainly, proteins are biomass components. However, the resource allocation towards protein synthesis represents the protein fraction exceeding the maintenance. Moreover, it was shown that a careful curation is needed to reveal the inadequacy of existing models for simulating primary anabolic processes, such as amino acid synthesis, under photoautotrophic conditions.

Finally, in Chapter 6, the proposed determination of metabolic costs of amino acid synthesis demonstrates the often preferred mixed nitrogen acquisition. The emerging, purely stoichiometric explanation corroborates the still controversial discussed competition of nitrate assimilation and light reactions for ferredoxin. The (seemingly obvious) impact of the environmental factors on the metabolic costs are highly dependent on the model and, therefore, emphasize the relevance of the model selection. In order to enable the comparative analysis, amendments and model adjustments were made to get the models properly working and to meet the shared cellular initial scenario.

7.2. Encountered problems

In modeling photosynthesis and the related metabolic processes, the acquisition of required knowledge has proven to be one of the most challenging steps in the process of model assembly. Obviously, the type of required information differs with the selected modeling approach. For kinetic modeling primarily knowledge/data of the functional principle of enzymes and the kinetic parameters are required, whereas for stoichiometric modeling, amongst others, the GPR associations are needed. However, the majority of the encountered problems deal with the scale of modeling rather than on the selected modeling approach.

Small-scale models are mostly covering one to a couple of well-established biochemical pathways, often occurring in a single cell compartment, *e.g.*, CBC and starch synthesis. Accordingly, the crucial biochemical information about the catalyzing enzymes (and coding genes) are well-known. This provides the solid basis for analyzing the dynamic behavior of these models, by means of kinetic modeling.

For an accurate kinetic model, the information about the functional principle of the catalyzing enzymes is indispensable, *e.g.*, the order of substrate binding, the occurrence of substrate saturation and the interdependency to regulators. Based on the availability of this information, the kinetic rate law can be selected: a precise form accounting for reversibility and regulatory effects, or a generalized form omitting these effects, which is a common cause of unrealistic behavior [285]. In some cases, a kinetic rate law is plainly assumed or previously used rate laws are applied without checking for justification or consistency (see sections 2.4 and A.1.2).

Furthermore, the kinetic-specific reaction rate parameters, *e.g.*, Michaelis-Menten constant K_m or rate constants k of the law of mass action, are a challenging issue. On the one hand, experimental data are only partially available, mostly for well-investigated reactions, on the other hand, the measurements are commonly conducted *in vitro*. It has been shown that for a large portion of enzymes the *in vitro* kinetics distinctly deviates from the *in vivo* behavior [286, 287]. While *in vivo* enzyme kinetic parameter determination is intrinsically hampered by the cell environment [288], it would allow to account for potential *in vivo*

modifications of the kinetic parameter due to substrate-product competition, competitive product binding, allosteric regulation and post-translational modification.

In large-scale metabolic modeling, the problems one has to cope with are on a different level of cellular organization. One of the major problems arises from cellular compartmentalization. The majority of assignments of reaction localization only relies on predictions based on the peptide sequence of the catalyzing enzymes. These predictions have to be examined carefully since predictors, *e.g.*, TargetP [289] and Predotar [290], can only distinguish between chloroplast, mitochondrion, secretory pathways and other location, usually denoted by 'cytoplasm'. Of course, not all secretory proteins have a signal peptide and the plastidic and mitochondrial gene-encoded proteins do not have a transit and targeting peptide, respectively.

As a consequence of compartmentalization, transport processes across compartment boundaries have to be considered, active transport, such as carrier proteins, and inactive transport, such as diffusion. As the detection of metabolite-specific translocation is difficult, probably only the minority of existing transport processes are identified [291–294]. For instance, amino acid synthesis in plants is highly compartmentalized such that for the protein assembly, almost each amino acid has to be transported into the target compartment. Similarly, for several biochemical pathways the translocation of at least one intermediate metabolite is required based on the reported or predicted localization. However, often the corresponding transport processes are not identified yet [295], entailing the inclusion of several non-annotated transport reactions.

The mapping of the metabolic processes of a cell entails almost unavoidably missing associations of the catalyzing enzyme and/or coding gene(s), resulting in further non-annotated reactions. These can be, furthermore, caused by spontaneous reactions, the incomplete or ambiguous association of Enzyme Commission numbers and reactions, marginally investigated reactions, *e.g.*, of the secondary metabolism which is in parts rather species-specific, or unverified reactions. The latter results from so-called gaps in the metabolism, in the best case a missing reaction in a known biochemical pathway, or produced metabolites for which no further reaction is known to exist in the respective compartment, cell type, or species.

To overcome these dead-ends, thorough literature scanning is needed and gap-filling algorithms can be applied. These algorithms try to obtain missing reactions known to appear in other species [296–300], or by enumerating possible metabolic routes based on computational chemistry methodologies [301]. Thereby, the aim is to determine a minimal set of reactions to ensure the desired metabolic function. Most recently, Vitkin and Shlomi [300] involve functional genomics data for the determination of gap-filling candidates across species, so-called phylogenetic enzyme profiles, and gene expression data. This promises to determine functional orthologs rather than to select a set of possible reactions solely fulfilling the biochemical metabolite interconversion.

However, the usage of such algorithms does not replace careful manual curation. The validation of the candidates specified in this manner for an iteratively improved, well-curated network of *E. coli* shows a precision of 41.9 % and a recall of 24.3 % [300]. This might be improved if the underlying phylogenetic profile would consider explicitly the phylogenetic tree for the respective species. According to this, confident sets of GPR associations are important for the applicability of large-scale models, in particular for data integration.

The integration of experimental data, such as transcriptomics and proteomics, reinforces the biological implications of the models and, ultimately, enables the improvement of their predictive power [302]. A high coverage of enzyme-coding genes and enzymatic proteins promises an expansive data mapping. However, the quality depends on the available information of the GPR associations. Missing or incorrect assignments of genes and/or proteins to the metabolic reactions can lead to data misinterpretation. Apart from this, incomplete knowledge of the GPR associations may cause confounding inferences, *e.g.*, due to compartment-specific enzyme forms, isozymes and enzyme complexes.

Some biochemical reactions, which occur in more than one compartment, are catalyzed by different, compartment-specific forms of the very same enzyme, *e.g.*, glyceraldehyde 3-phosphate dehydrogenase [303]. A mapping neglecting this characteristic may miss compartment-specific differences in the metabolic flux distribution, *e.g.*, among starch and sucrose synthesis pathways.

In addition, a precise classification of assignments, such as enzyme complexes and isozymes, is also required. While isoforms of genes, peptides and/or enzymes are considered in a cumulative way, only the overlap of the essential subunits of an enzyme is taken into account. Inverting the composition may represent marginal or excess abundance of the catalyzing enzyme. Once assigned, such effects are difficult to recognize and allocate without going back to thoroughly scanning literature with respect to GPR associations.

Aside from integrating data for model validation and making predictions, data are also a basic component of a model, irrespectively of the scale of modeling and the selected modeling approach. Environmental conditions, such as temperature and nutrient availability, define thermodynamic and cellular constraints, and cellular conditions, *e.g.*, pH value, levels of metabolites and the biomass composition, describe the cellular state. Ideally, such data should be part of the experimental control, however, the respective measurements are often too elaborated. The accurate quantification of, for instance, cell biomass components involves different high-throughput technologies for assessing the individual components (see Chapter 5). While the expenditure of time is already substantial for a unicellular organism, the measurements are more complex for higher organisms, such as plants.

Although photosynthesis is one of the best-investigated metabolic processes, from the modeling perspective, further, high-resolution data are desirable to improve the predictive power of the models. It is often that studies of plant metabolism focus on the transient behavior or comparative analysis of time-series data from different conditions. Limitations due to the temporal resolution of sampling processes, *e.g.*, enzyme activation measurements, might be resolved by automatation of the measurements. Effortful, quantitative measurements of, *e.g.*, metabolomics and proteomics, require prior knowledge and additional measurements of the identified metabolites and proteins, respectively. To reliably assess the demand for high-resolution measurements, modelers and experimentalists need to set up together the experimental design and the scope of modeling.

7.3. Modeling perspectives

Modeling is considered as an iterative process. Corresponding research findings, even those pertaining purely to modeling, had lead to substantial reinforcement of some and revision of other arguments with respect to the understanding of metabolism. Furthermore,

progress in the experimental knowhow (*i.e.*, technology for data acquisition, data quality) and computing capacity (*e.g.*, for parameter estimation, elementary mode calculation) permit innovations in modeling. Consequently, modeling attempts such as the ranking framework, proposed in Chapter 2, and the Arabidopsis core model, presented in Chapter 5 demand constant revisions and updates. Such improvements and modeling innovations are expected to enable more accurate predictions and advanced applicability for the various fields of biological engineering [304] and for addressing diverse biological questions, also the self-suggesting ones of "How to improve crop yield?" and "How does crop plants adapt to changing climatic conditions?".

Based on the findings of Chapter 3 and 5, the conception of a large-scale model can be envisioned and undertaken by combining two features: the combination of the GPR associations expanded by the enzyme complex structures and the incorporation of the catalyzing enzymes. Integrating enzymes as 'internal metabolites', their synthesis and their turnover (Chapter 3) could be employed to partially account for enzymatic activity. This will admittedly enlarge the number of reactions of the underlying metabolic model by a factor of three to four. However, since the resulting model still only considers stoichiometry, constraint-based analyses can be readily performed. Nevertheless, network-based pathway analyses, such as the determination of elementary modes, are expected to become more computation-intensive due to the implied increase in the network size (both in terms of metabolites and reactions).

Combined with complete protein sequences (*e.g.*, from TAIR database), the incorporation of the enzyme complex structures (Chapter 5) enables a semi-automatic determination of enzyme synthesis costs. As a result, the metabolic costs can be system-wide expanded by the expenditures for maintaining enzyme availability, as demonstrated in Chapter 3. The resulting large-scale model accounts, at least partially, for dynamic behavior, in terms of enzymatic limitations. Such enzymatic limitations can be effected by, for instance, elevated enzyme turnover affecting the enzyme levels and/or enzyme (de)activation influencing the portion of active enzyme. In this way, dynamic responses to changing environmental conditions can be included in the examination. This may set the course for diminishing the shortcomings of the currently existing two distinct approaches of modeling metabolism and, therewith, increase the applicability and the predictive value of a metabolic model.

Furthermore, the explicit inclusion of enzymes allows to consider further enzyme-related processes on other levels of cellular organization. Amongst others, this includes protein-protein interactions and the interrelationship of the system's level regarding regulatory processes. Interacting proteins in metabolic pathways, so-called metabolons, allow passing the metabolic product from one enzyme directly as substrate into the active site of the consecutive enzyme, termed metabolite channeling [305, 306]. Channeling can render a metabolic pathway more efficient as the intermediatory metabolites cannot be diverted to other biochemical processes. From the modeling perspective, metabolons may restrict the dimension of the flux distributions by the sequential reactions. Additionally, the coupling rate is increased as the sequential reactions are fully coupled and the number of elementary modes is decreased. In this manner, metabolic networks can be linked to protein-protein interaction networks and, thereby, also interacting proteins in signaling pathways can be incorporated [307].

The GPR associations in a large-scale metabolic model provide the means for integrating

gene-regulatory networks [198, 308]. Thereby, transcriptional control and, in a lumped way, the subsequent types of gene-expression regulation can be taken into account. By means of incorporating enzymes, the connecting link between genes and reactions, the level of gene-expression regulation can be reflected to a certain extent. For instance, posttranslational control, in terms of regulating enzyme (de)activation and enzyme degradation rates can be straightforwardly included. As a result, this facilitates the separate examination of the multiple levels of gene-expression regulation which together shape the enzyme profile [9] and, accordingly, affect the metabolism.

Altogether, the findings of this thesis provides the first steps for the incorporation of enzymes on a large scale and, accordingly. offers the means for directing future plant modeling attempts towards capturing the entirety of the system's level. In this manner, the metabolic behavior and its interrelationship to the remaining system's level can be elucidated, compliant with the efforts of the systems biology approach.

Appendix A.

A quantitative comparison of Calvin-Benson cycle models[10]

A.1. Model evaluation

A.1.1. Hierarchy of the models

The compendium comprises different approaches of modeling. By considering each model, some of them can be easily identified as initial approaches and some as extensions of others. However, these two classifications are not disjoint, as the threshold for calling a model based on an older one depends on the observer/reader. In our opinion, a significant part of the structure has to be shared to call the newer one an extension (Table A.1).

Table A.1.: Different model approaches distinguish between initial and extensions

Initial approach	Extensions			
Farquhar [73]	Medlyn [74]	Schultz [75]	Sharkey [76]	Damour [77]
Hahn [80]	–			
Pettersson [58]	Woodrow [59]	Poolman [60]	Zhu '07 [51]	Zhu '09 [78]
Laisk '89 [61]	Laisk '06 [56]			
Giersch [79]	–			
Fridlyand [57]	–			

A.1.2. Model versions in the compendium

The models of the compendium have been assembled from the literature as described in Table 1. However, a few changes are conducted to facilitate the analysis (Table A.2).

Furthermore, some changes are necessary for model correctness: For Farquhar *et al.* [73], and all models based on it [74–77], the values of R_p and ϕ are calculated and the parameters are adapted to initial values of unit $[mM]$ by using the subcellular volume for chloroplast stroma [91] and BioNumbers (ID: 105594) A.2. Moreover, the models of Schultz [75] and Sharkey *et al.* [76] describe the velocity of the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) reaction, v, in dependence of the total assimilation rate,

---

[10]This chapter is based on the supplemental material of the publication of Arnold and Nikoloski (2011) [46]. Parts of section A.2 were shifted into Chapter 2.

Table A.2.: Sources of the parameter values for the models within the compendium.

Model	Parameters	Source	Refs
Farquhar	primarily	Table "Symbols, units, normal values" (p. 88)	[73]
	R_p	Equation (19)	
	ϕ	Equation (17)	
	E_t, j	converted into $[mM]$ using Winter *et al.* [91] Table 3 (stroma)	
Medlyn	primarily	Farquhar *et al.* [73], Equation (4), (18)	[74]
	V_{cmax}	average of corresponding values of Table 2	
	α, θ	out of the text (p. 1170)	
	J	additionally $J = 0.9 J_{max}$ (p. 1172)	
	CO_2	$0.7 \cdot 350 \frac{\mu mol}{mol}$ (p. 1172)	
	K_c, K_o, Γ^*	Bernacchi *et al.* [309] Equation (8) and Table 1	
	V_{cmax}, J	converted into $[mM]$ using Winter *et al.* [91] Table 3 (stroma), BioNumbers (ID: 105594)	
Schultz	primarily	Medlyn *et al.* [74], Equation (6)–(12), Table 1	[75]
	α	Table 1, sun leaves (20–30°C)	
	h	= 0.6 due to "relative humidity of > 50%" (p. 675)	
	LPI	= 15 due to "LPI $10 - 20$ (mature leaves)" (p. 677)	
	R_d, Γ^*	additionally legend of Figure 1	
	V_{cmax}, J, TPU, R_d, g_0	converted into $[mM]$ using Winter *et al.* [91] Table 3 (stroma), BioNumbers (ID: 105594)	
Sharkey		Medlyn *et al.* [74], Equation (4)–(6), Table 1	[76]
	V_{cmax}, J, TPU, R_d, g_m	converted into $[mM]$ using Winter *et al.* [91] Table 3 (stroma), BioNumbers (ID: 105594)	
Damour	primarily	Medlyn *et al.* [74], Urban *et al.* [176] Equation (7) and (8), Urban and Alphonsout [310] Equation (3) and C'_{starch} (p. 15-3)	[77]
	α	Urban *et al.* [176] Table 1	
	[starch]	Urban *et al.* [176] Table 2	
	Γ^*	$= 0.5 \frac{O_2}{\tau}$ Urban *et al.* [176] (p. 291), Equation (8), Table 1	
	a	Urban and Alphonsout [310] out of the text (p. 348)	
	V_{cmax}, J_{max}, TPU	Harley *et al.* [311] Equation (9), Table 1	
	K_c, K_o	Bernacchi *et al.* [309] Equation (8) and Table 1	
	V_{cmax}, J, TPU	converted into $[mM]$ using Winter *et al.* [91] Table 3 (stroma), BioNumbers (ID: 105594)	
Fridlyand	V_{max}	Fridlyand *et al.* [85]	[57]
	K	publication and implementation of Zhu *et al.* (2007) [51], Laisk *et al.* (1989) [61], Anderson [87],Leegood [86]	
	q	Zhu *et al.* (2007) [51], Pettersson and Ryde-Pettersson [58]	
Zhu '09	ATP	Zhu *et al.* (2007) [51]	[78]
Giersch	$k1$	Figure 5, 6	[79]
	A_0	Zhu *et al.* (2007) [51]	
Hahn	only initial values of RuBP, PGA, TP, HeP, TPGA (= R5P + X5P), E4P, S7P, Ru5P from Zhu *et al.* (2007) [51]		[80]
Poolman	ATP	Zhu *et al.* (2007) [51]	[60]
	formula for starch degradation from Poolman [83]		
Pettersson	only initial values of RuBP, FBP, SBP, R5P, ATP, c_A, P_i, NADPH, NADP and proton from Zhu *et al.* (2007) [51]		[58]
Woodrow	only initial values of RuBP, FBP, SBP, R5P, ATP, ADP, P_i from Zhu *et al.* (2007) [51]		[59]
Laisk '89	all	Equation (6)–(56) with description	[61]
	K	of sucrose phosphate synthase from Laisk *et al.* (2006) [56]	
Laisk '06	all	implementation; with references to Laisk *et al.*, Laisk *et al.*, Laisk and Edwards (2006, 1989, 2000) [56, 61, 84]	[56]
Zhu '07	primarily	appendix of Zhu *et al.* (2007) [51]	[51]
	V_{max}	implementation	

A (Equations A.1–A.4):

Sharkey *et al.*:
$$A = v\left(1 - \frac{\Gamma^*}{C_i}\right) - R_d = \min\{W_c, W_j, W_p\} - R_d \tag{A.1}$$

$$W_c = V_{c\,max}\left[\frac{C_c - \Gamma^*}{C_c + K_c\left(1 + \frac{O_i}{K_o}\right)}\right] \qquad C_c = C_i - \frac{A}{g_m} \tag{A.2}$$

$$W_j = J\frac{C_c - \Gamma^*}{4C_c + 8\Gamma^*} \tag{A.3}$$

$$W_p = 3TPU \tag{A.4}$$

To resolve this dependence, C_c is replaced and the equation is solved for $A_{c,j,p}$ (Equations A.5–A.7):

$$A_c = \frac{g_m}{2}\left\{C_i + \frac{V_{c\,max} - R_d}{g_m} + K_c\left(1 + \frac{O_i}{K_o}\right) \pm \sqrt{\left(C_i + \frac{V_{c\,max} - R_d}{g_m} + K_c\left(1 + \frac{O_i}{K_o}\right)\right)^2 +}\right.$$
$$\left.\overline{\frac{4}{g_m}\left(R_d\left(C_i + K_c\left(1 + \frac{O_i}{K_o}\right)\right) + V_{c\,max}\left(C_i - \Gamma^*\right)\right)}\right\} \tag{A.5}$$

$$A_j = \frac{g_m}{2}\left\{C_i + \frac{J - 4R_d}{4g_m} + 2\Gamma^* \pm\right.$$
$$\left.\sqrt{\left(C_i + \frac{J - 4R_d}{4g_m} + 2\Gamma^*\right)^2 + \frac{4}{g_m}\left(R_d\left(C_i + 2\Gamma^*\right) + \frac{J}{4}\left(C_i - \Gamma^*\right)\right)}\right\} \tag{A.6}$$

$$A_p = 3\,TPU - R_d \tag{A.7}$$

The models of Pettersson and Ryde-Pettersson [58], Poolman *et al.* [60] and Zhu *et al.* (2007) [51] use a unit-inconsistent velocity describing the ADP-glucose pyrophosphorylase (AGPase) reaction. In accordance with the authors of the initial model [58], we solve this issue by rewriting the velocity (Equation A.8):

$$v = \frac{V_{max} \cdot G1P \cdot ATP}{(G1P + K_{M_1})\left(\left(1 + \frac{ADP}{K_I}\right)\left(ATP + K_{M_2} \cdot \left(1 + \frac{K_{M_2} \cdot Pi}{K_{A_1} \cdot PGA + K_{A_2} \cdot F6P + K_{A_3} \cdot FBP}\right)\right)\right)}. \tag{A.8}$$

A similar problem appears for a velocity describing the transformation and transport from 3-phosphoglycerate (PGA) to sink in the model of Zhu *et al.* (2009) [78]. Also in accordance with the authors' suggestion, we have solved this challenge by changing the velocity (Equation A.9). This improved form is also used by the authors in their recent publication [312], *i.e.*,

$$v = \frac{V_{max} \cdot PGA}{PGA + K_{m5}}. \tag{A.9}$$

Additional changes have to be made to the model of Zhu *et al.* (2007) [51]. While the velocity of ribulose-5-phosphate (Ru5P) kinase, according to the equations in the original

publication, is regulated by glyceraldehyde-3-phosphate (GAP) [51], by using the authors' model implementation, it turns out that PGA is in fact the actual regulator.

For the model of Laisk *et al.* (1989) [61], an evaluation problem arises for physiologically plausible initial values. The complex description of ADP (Equation A.11) results in imaginary concentrations for reliable values of ADP-glucose (ADPG) and adenylphosphate (AdenylP):

$$\text{ADT} = \frac{\text{AdenylP}}{2}, \tag{A.10}$$

$$\text{ADP} = \sqrt{\left(\frac{\text{ADT} - \text{ADPG}}{4k_e - 1}\right)^2 + \frac{\text{AdenylP}\left(2\left(\text{ADT} - \text{ADPG}\right) - \text{AdenylP}\right)}{4k_e - 1}}$$
$$- \frac{\text{ADT} - \text{ADPG}}{4k_e - 1}, \tag{A.11}$$

$$= \sqrt{\left(\frac{\frac{1}{2}\text{AdenylP} - \text{ADPG}}{4k_e - 1}\right)^2 - \frac{\text{AdenylP}\left(2\,\text{ADPG} + \frac{1}{2}\right) - \text{ADPG}}{4k_e - 1}}. \tag{A.12}$$

To check the metabolic range for non-imaginary ADP concentration, we have fixed each metabolite concentration to its initial value provided by Zhu *et al.* (2007) [51] and have calculated the concentration range for the other one (Equations A.13 and A.14):

$$\text{ADPG} = 6.295e\text{-}3 \implies \text{AdenylP} < 0.0122808 \ \lor \ \text{AdenylP} > 17106.2, \tag{A.13}$$

$$\text{AdenylP} = 3.0 \qquad \implies \text{ADPG} < -0.299922 \ \lor \ \text{ADPG} > 41718.3. \tag{A.14}$$

However, even with such unreliable initial values, this description does not allow any temporal analysis of the model and, therefore, it is excluded from any further examination.

This provides the support to integrate an extension of Laisk *et al.* (1989) [61]. The subsequent model published in 2006 [56] also captures the Calvin-Benson cycle (CBC), starch as well as sucrose synthesis and additionally a description of photosystems and the electron transport chain, *i.e.*, parts of the light reactions. Since we investigate carbon fixation in terms of the CBC and its related processes integrating the fixed CO_2 into the metabolism, we have used a reduced version by excluding the parts of the light reactions for this study. Of course, the entire model should also be investigated; however, this requires more extensive data.

A.1.3. Classification

Prior to performing the analysis, we have classified the models regarding the four criteria mentioned in the main text (Table 1). Two of the criteria need to be further described, namely, the involved types of kinetics and the used regulation terms. Here, the extended version of the classification is given for these criteria (Tables A.3 and A.4).

Table A.3.: Complete classification of the involved kinetics and regulatory processes (part I).

Model		Farquhar [73]		Medlyn [74]		Schultz [75]		Sharkey [76]		Damour [77]	
Reaction	limiting											
RuBisCO	RuBisCO	irr M-M	c2	lin. comb. M-M	c2/c1	lin. comb. M-M	c2/c1	lin. comb. M-M	c2/c1	lin. comb. M-M	c2/c1
	RuBP Reg.	spec. func.	–	spec. func.		spec. func.		spec. func.		spec. func.	
	TPU					mass action		mass action		mass action	
CBC Reg.		mass action		mass action		mass action		mass action		mass action	
NADPH prod.		mass action		mass action		mass action		mass action		mass action	

Kinetics

Several different kinetics are used within the compendium. We have categorized them into five types: (i) mass-action, (ii) equilibrium-approximation, (iii) Michaelis-Menten, (iv) Michaelis-Menten-like, and (v) special-functions kinetics. *Mass action* (red) is one of the simplest kinetics which however precludes inclusion of regulatory processes (Equation A.16). The simplified type, *equilibrium approximation* (green), also does not allow the inclusion of regulation terms because it is based on mass action. Here, a rapidly reached steady state is assumed and approximated as a constant velocity of a mass-action kinetics (Equation A.17). On the one hand, this simplification can reduce the system size: the metabolites of such reactions, which are considerably fast compared to the adjacent reactions, can be merged into a metabolic pool and their transient behavior is determined by the dependence to this, usually small, pool (Equation A.20) [51, 56, 61]:

$$\text{reaction:} \qquad \text{GAP} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \text{DHAP}, \tag{A.15}$$
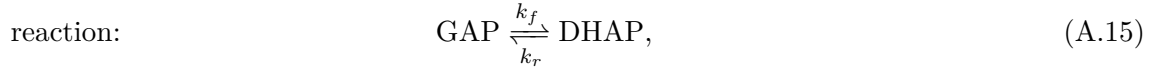
$$\text{mass action:} \qquad v = k_f\,\text{GAP} \ - \ k_r\,\text{DHAP}, \tag{A.16}$$

$$\text{equilibrium approx.:} \qquad \text{GAP} \ = \ k_{eq}\,\text{DHAP} \qquad \left(k_{eq} = \tfrac{k_r}{k_f}\right), \tag{A.17}$$

$$\text{TP pool formation:} \qquad (1 + k_{eq})\,\text{GAP} \ = \ k_{eq}\,(\text{GAP} + \text{DHAP}), \tag{A.18}$$

$$(1 + k_{eq})\,\text{GAP} \ = \ k_{eq}\,\text{TP}, \tag{A.19}$$

$$GAP \ = \ \frac{k_{eq}\,\text{TP}}{1 + k_{eq}}. \tag{A.20}$$

On the other hand, if the so-described reactions do not form a disjoint (separate), small pool, the underlying steady-state assumption reduce, or even disable, the evaluation of the system in the time domain. Additional assumptions for each reaction could again enable the temporal analysis of the system. For instance, the additional knowledge of at least one of the two reaction constants defining the equilibrium constant of a reaction, $k_{eq} = \frac{k_r}{k_f}$, allows the retransformation to mass-action kinetics and, therefore the evaluation in time domain.

The third type, *Michaelis-Menten* (M-M, blue), is also based on mass action but as a composition of metabolite-enzyme (E^*) steps with the possibility to integrate regulation

Table A.4.: Complete classification of the involved kinetics and regulatory processes (part II).

Model / Reaction	Fridlyand [57]	Zhu '09 [78]	Giersch [79]	Hahn [80]	Poolman [60]	Pettersson [58]	Woodrow [59]	Laisk '89 [61]	Laisk '06 [56]	Zhu '07 [51]
RuBisCO (CO_2)	irr M-M	—	h.e. M-M	irr M-M	irr M-M, p3,c4,c6,c14,c15	irr M-M, p3,c4,c6,c14,c15	irr h.e. M-M, c2	equil., approx.	irr M-M, p13	irr M-M, c2,p3,c4,c6,c14,c15
PGAK	rev M-M	irr M-M		mass action	mass action	equil.	equil., approx.	rev M-M	rev M-M, like	r-i M-M
GAPDH	rev M-M	irr M-M	rev M-M, like	mass action	mass action	equil., approx.	rev M-M	rev M-M *	rev M-M, like **	irr M-M
TPI	rev M-M	—		mass action	mass action	equil.	equil.	equil.	equil.	—
FBPA	rev M-M	—		mass action	mass action	equil.	equil., approx.	equil., approx.	rev M-M	rev M-M
FBPase	irr M-M	—		mass action	irr M-M, p5,p14	irr M-M, p5,p14	rev M-M, p14	rev M-M	rev M-M, like	r-i M-M, p5,p14
SBPA	rev M-M	—		mass action	mass action	equil.	equil.	rev M-M	rev M-M	rev M-M
SBPase	irr M-M	irr M-M	irr M-M	mass action	irr M-M, c14	irr M-M, p14	irr M-M, c14	rev M-M	rev M-M	r-i M-M, c14
TK (F6P)	rev M-M	—		mass action	mass action	equil.	equil.	spec. func.	spec. func.	r-i M-M
TK (S7P)	rev M-M	—	irr M-M	mass action ***	mass action	equil.	spec. func.	spec. func.	spec. func.	spec. func.
PRI	rev M-M	—		mass action	mass action	equil.	equil.	equil.	equil.	equil.
RPE	rev M-M	—		mass action	mass action	equil., approx.	equil., approx.	equil., approx.	equil., approx.	equil., approx.
PRK	rev M-M	irr M-M	irr M-M, like, p14	mass action	irr M-M, c3,p7,m13,p14	irr M-M, c3,p7,p13,p14	irr M-M, c3,c4,p7,m13,p14	rev M-M	rev M-M	r-i M-M, c3,p7,m13,p14
AGPase				mass action	irr M-M, n13,c14,a1,a2,a3	irr M-M, n13,c14,a1,a2,a3	irr M-M, like, a1	rev M-M, c14,a1	rev M-M, x14,a1	irr M-M, n13,c14,a1,a2,a3
PGM				mass action	mass action	equil., approx.	equil., approx.	approx.	approx.	approx.
PGI				mass action	mass action	equil., approx.	equil., approx.	equil.	equil.	equil., approx.
StS				mass action	irr M-M, p8			rev M-M	rev M-M	rev M-M
StP				mass action	irr M-M					
ATP synthase	irr M-M	—	rev Giersch	mass action	irr M-M	irr M-M	irr M-M	rev M-M	rev M-M	rev M-M
TPT	rev Giersch	Giersch	Giersch	mass action	Giersch	irr Giersch	Giersch	rev Giersch	rev Giersch	irr Giersch

* – PGAK, GAPDH, TPI merged; additionally TPI to produce GAP   ** – equilibrated with NADPH/NADP ratio (considered as constants)   *** – TK, PRI, RPE rearranged; introducing thiamine pyrophosphate glycoaldehyde

Model Reaction	Fridlyand [57]	Zhu '09 [78]	Giersch [79]	Hahn [80]	Poolman [60]	Pettersson [58]	Woodrow [59]	Laisk '89 [61]	Laisk '06 [56]	Zhu '07 [51]
cyt TPI								equil. approx.	equil. approx.	equil. approx.
cyt FBPA						–		rev M-M	rev M-M –	rev M-M –
cyt FBPase				mass action				spec. func.	spec. func.	r-i M-M p5,p14
cyt PGI								equil. approx.	equil. approx.	equil. approx.
cyt PGM								equil. approx.	equil. approx.	equil. approx.
UGPase						–		rev M-M	rev M-M –	rev M-M –
SPS				mass action				spec. func.	spec. func.	spec. func.
SPP						–		rev M-M	rev M-M –	r-i M-M p16
PF2K								spec. func.	spec. func.	irr M-M n5,n14
F2,6BPase								spec. func.	spec. func.	r-i M-M p16
NDK				mass action						rev M-M –
RuBisCO (O_2)							irr h.e. M-M c1			irr M-M c1,p3,c4, c6,c14,c15
PGPase										irr M-M p9,p14
GK										r-i M-M p3
GPT										irr M-M c9/c10/ c11/c12
GO										irr M-M –
GGAT										r-i M-M p17
GDC										irr M-M c18
SGAT										r-i M-M p17
GDH										r-i M-M c19
sink (PGA)		irr M-M –								
sink (GAP)		irr M-M –								

abbreviations:

x1 – CO_2 x5 – F6P x9 – GCA x13 – ADP x17 – GLYc a1 – PGA
x2 – O_2 x6 – SBP x10 – GCAc x14 – P_i x18 – SERc a2 – F6P
x3 – PGA x7 – RuBP x11 – GCEA x15 – NADPH x19 – HPRc a3 – FBP
x4 – FBP x8 – G1P x12 – GCEAc x16 – SUCc

x ∈ {c – compet. inh., p – compet. product inh., n – non-compet. inh., m – mixed-type inh., x – unidentified inh.}

terms (Equation A.27):

reaction:
$$\text{GAP} + E^* \underset{k_{r_1}}{\overset{k_{f_1}}{\rightleftharpoons}} \mathbf{C} \xrightarrow{k_{f_2}} \text{DHAP}, \tag{A.21}$$
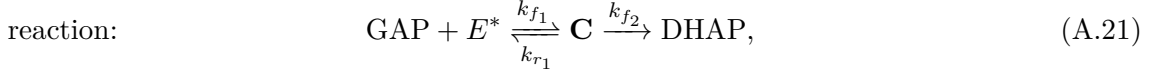
complex $\mathbf{C}$:
$$v_{\mathbf{C}} = k_{f_1} \, \text{GAP} \cdot E^* - (k_{r_1} + k_{f_2}) \, \mathbf{C}, \tag{A.22}$$

$$0 = k_{f_1} \, \text{GAP} \, (E_t^* - \mathbf{C}) - (k_{r_1} + k_{f_2}) \, \mathbf{C}, \tag{A.23}$$

$$k_{f_1} \, \text{GAP} \cdot E_t^* = (k_{f_1} \, \text{GAP} + k_{r_1} + k_{f_2}) \, \mathbf{C}, \tag{A.24}$$

$$\text{GAP} \cdot E_t^* = (\text{GAP} + K_M) \, \mathbf{C} \qquad \left(K_M = \tfrac{k_{r_1} + k_{f_2}}{k_{f_1}} \right), \tag{A.25}$$

$$\mathbf{C} = E_t^* \frac{\text{GAP}}{\text{GAP} + K_M}, \tag{A.26}$$

product:
$$v = k_{f_2} \, \mathbf{C} = V_{\max} \frac{\text{GAP}}{\text{GAP} + K_M} \qquad (V_{\max} = k_{f_2} \, E_t). \tag{A.27}$$

For the formula and derivations of reversible M-M forms, we refer the reader to Segel [313, chap. 2].

Moreover, there exist several extensions of M-M summarized as *Michaelis-Menten like* (violet). A common example is high enzyme M-M or the combination of reversible M-M nominator and irreversible M-M denominator [51]. All kinetics which cannot be categorized into these groups form the group *special functions* (black). For instance, this implies the kinetics of Giersch especially applied for the triose-phosphate translocator [62] and the one for transketolase reactions [51, 56, 61].

Spatiotemporal properties

Concentrations of metabolites change with respect to both space and time, provided that the system is far from the assumption of random mixing. There exist two ways in which spatiotemporal dynamics can be included in any biochemical model (and thus, in a model of the CBC): The first consists of considering the cellular compartmentalization and the effects of partitioning the metabolic pools among the compartments, as we have done in this review. The second, which is pursued in classical studies on spatiotemporal dynamics in biochemical systems [81], includes the study of the diffusion of metabolites.

In plants, the chloroplast thylakoid membrane is the site of light-dependent photosynthetic reactions coupled to ATP synthesis. Recent experimental studies have confirmed that (i) the chloroplast thylakoid ATP/ADP carrier (TAAC) supplies the thylakoid lumen with stromal ATP in exchange for ADP and (ii) increased concentrations of ATP in the stroma cause activation of RuBisCO activase. Since ATP synthesis is essential for the light-independent photosynthetic reactions, the effects of ATP diffusion on the existence of inhomogeneous steady states was investigated in Grimbs *et al.* [81] in order to possibly settle the question for the possibility of existence of bistability. While the choice of ATP is an essential candidate for explaining the existence of inhomogeneous steady states, we point out that analogous analysis can be carried out for the other metabolites in a model. However, at present, there is no experimental evidence for direct effects on photosynthesis from the diffusion of other metabolites.

Regulatory processes

The regulatory processes of the CBC are well investigated experimentally but, unfortunately, not that well integrated into the CBC models. To provide a well-founded, systematical set of regulatory processes for the CBC, first we check the *status quo*. For the CBC and the related processes, we can focus on metabolic regulators which are introduced into the pathway (external) or part of the pathway itself and related ones (internal), respectively. Furthermore, we can classify the regulations terms of M-M and M-M like kinetics: Besides the two types, activation and inhibition, we can distinguish between five (sub)types of inhibition, a previously unidentified one and the following four (Equations A.28–A.31):

competitive:
$$v = V_{max} \frac{[S]}{[S] + K_M \left(1 + \frac{[I]}{K_I}\right)}, \tag{A.28}$$

uncompetitive:
$$v = V_{max} \frac{[S]}{[S] \left(1 + \frac{[I]}{K_I}\right) + K_M}, \tag{A.29}$$

non-competitive:
$$v = V_{max} \frac{[S]}{([S] + K_M) \left(1 + \frac{[I]}{K_I}\right)}, \tag{A.30}$$

mixed:
$$v = V_{max} \frac{[S]}{[S] \left(1 + \frac{[I]}{K_{I_1}}\right) + K_M \left(1 + \frac{[I]}{K_{I_2}}\right)}. \tag{A.31}$$

Most of the reversible inhibitors are modeled using *competitive* inhibition (Equation A.28, Tables 1 column 6, A.3 and A.4), *e.g.*, the CO_2–O_2 competition at RuBisCO [51, 59, 73–77]. In both carboxylation and oxygenation reactions, each gaseous substrate, as an external regulator, competes for the branching-point intermediate, ribulose-1,5-bisphosphate, and, therefore, they determine the ratio of CBC and photorespiration [66].

The second most applied form is the *competitive product* inhibition as a subtype of the competitive inhibition. For instance, the fructose-1,6-bisphosphatase reaction is inhibited by its product fructose-6-phosphate (F6P) [71]. For *uncompetitive*, *non-competitive* and *mixed-type* inhibition such distinction of the regulator is unusual and, therefore, we have only considered the general form (Equations A.29–A.31). While non-competitive and mixed inhibition are in almost all cases applied for ADP inhibition, uncompetitive inhibition is not used at all within the compendium. A common example for modeled and experimentally proven inhibition by ADP is the Ru5P kinase reaction [314]. However, this reaction is also an example for incorporating unproven regulators: in five of the fifteen models, GAP is assumed to act as an inhibitor. Such implausible statements may lead to unrealistic models and unreliable (or even wrong) predictions.

Within the classification, we have only noted regulators for M-M and M-M like kinetics. For the special functions, a classification of the regulatory processes is not trivial. For instance, transketolase catalyzes two reversible reactions both using GAP as a substrate and yield xylulose-5-phosphate as a product (Equation A.32):

$$\begin{matrix} \text{F6P} \\ \text{S7P} \end{matrix} \;+\; \text{GAP} \rightleftharpoons \text{X5P} \;+\; \begin{matrix} \text{E4P} \\ \text{R5P} \end{matrix}. \tag{A.32}$$

Certainly, these reactions affect each other and, consequently, the kinetics should reflect this dependence. However, we cannot distinguish whether this effect is a regulation term or a part of the kinetics [51, 56, 61].

A.1.4. SBML implementation

One of our major goals is to provide an easy reproducibility of the examined models as well as the presented results. Consequently, we have implemented the models in Systems Biology Markup Language (SBML), as far as possible (Table A.5). We choose *xml*-files of level 2 version 1 (L2V1) which allow the usage of computational tools such as: COPASI and SBtoolbox2 (allows at most L2V1). The models will be available on BioModels Database (contact the authors to obtain the respective URLs).

Table A.5.: Status of the implementation of the fifteen CBC models.

Model	SBML-implemented		Compliance	Stability/ Robustness	Sensitivity	Refs
Farquhar	✓		✓	✓		[73]
Medlyn	✓		✓	✓		[74]
Schultz	✓		✓	✓		[75]
Sharkey	✓		✓	✓		[76]
Damour	✓		✓	✓		[77]
Fridlyand	✗	missing parameters				[57]
Zhu '09	✓		✓	✓	✓	[78]
Giersch	✓		✓	✓	✓	[79]
Hahn	✓		✓	✓	✓	[80]
Poolman	✓		✓	✓	✓	[60]
Pettersson	✗	steady state assump.	✓			[58]
Woodrow	✗	steady state assump.	✓			[59]
Laisk '89	✗	imaginary ADP conc.				[61]
Laisk '06	✓		✓	✓	✓	[56]
Zhu '07	✓		✓	✓	✓	[51]
#	13		13	11	6	

The models will be available on the BioModels Database (respective URLs can be obtained by contacting the authors).

A.2. Results and supplementary material of the different analyses

The complete results and the supplementary material of the different analyses are provided in the Figures A.1 and A.2, and the Tables A.6–A.13. They are in order of appearance in the manuscript: steady state detection, stability and robustness analyses, compliance with data, and sensitivity analysis.

Table A.6.: Steady state concentrations of the involved metabolites for all models within the compendium.

Model / Metabolites	Farquhar [73]	Medlyn [74]	Schultz [75]	Sharkey [76]	Damour [77]	Zhu '09 [78]	Giersch [79]	Hahn [80]	Poolman [60]	Pettersson [58]	Woodrow [59]	Laisk '06 [56]	Zhu '07 [51]
PGA	5.452	5.994	33.661	215.042	9.073	0.147	9.704	2.295	2.187	1.703	9.390	2.216	6.437
DPGA						0.645			0.066	0.001			0.005
GAP (TP)						8.503	0.230	0.092	0.005	0.065	0.068	8.171	3.021
DHAP (TP)									0.108	1.441	1.434		
FBP									0.004	0.670	0.670	0.019	0.029
F6P (HeP)									1.251	0.218	0.571		
G6P (HeP)								1.088	2.878	0.500	1.326	0.479	0.412
G1P (HeP)									0.167	0.029	0.069		
TPGA								0.119					
E4P								0.119	0.025	0.016	0.029	1.967	0.227
SBP									0.036	0.300	0.300	0.009	0.335
S7P								0.109	0.167	0.168	0.302	0.037	0.475
X5P (PeP)									0.020	0.075	0.075		
R5P (PeP)									0.034	0.125	0.125	0.957	0.553
Ru5P						0.044	0.259	0.766	0.014	0.050	0.050		
RuBP	2.000	2.000	2.000	2.000	2.000	0.788	0.689	0.544	0.087	2.000	2.000	3.721	1.452
ER												1.019	
EPP												0.241	
EPG												0.268	
EP												0.815	
E0P												0.443	
ADPG												0.139	
starch								29.887					
CO_2[a]	0.230	0.245	0.350	0.245	0.245			0.310				0.365	0.268
O_2[b]	0.210	0.210	0.210	0.210	0.210							0.212	0.206
ATP						0.680	0.213	4.260	1.485	0.680	0.680	1.378	0.149
ADP						0.820	1.287	0.145	0.015	0.820	0.820	0.122	1.351
UTP								3.964					
UDP								1.521					
Pi						1.261	0.846	7.757	6.275	0.975	0.975	2.232	0.085
NADPH	0.165	0.163	0.043	0.003	0.120				0.210	0.210		0.210	0.210
NADP	0.335	0.337	0.457	0.497	0.380				0.290	0.290		0.290	0.290

[a] – in [μbar]; [b] – in [$mbar$];   – fixed value

115

Appendix A. A quantitative comparison of Calvin-Benson cycle models

Model / Metabolites	Farquhar [73]	Medlyn [74]	Schultz [75]	Sharkey [76]	Damour [77]	Zhu '09 [78]	Giersch [79]	Hahn [80]	Poolman [60]	Pettersson [58]	Woodrow [59]	Laisk '06 [56]	Zhu '07 [51]
PGA_c												1.183	0.103
TP_c							0.200	0.104				7.396	9.905
FBP_c												0.014	0.932
$F26BP_c$												0.0001	0.011
HP_c								0.083				0.647	1.915
$UDPG_c$												3.087	1.555
$SucP$												0.001	0.012
Suc							65.265						0.045
ATP_c												0.360	0.350
ADP_c												0.640	0.650
UTP_c												0.360	0.750
UDP_c												0.640	−0.805
Pi_c							1.400					1.904	0.078
$NADH$								0.735	0.500				0.400
NAD													0.470
$2P$													0.0001
GLC													0.054
GA													0.729
HPR_c													0.002
Ser_c													14.376
GLC_c													0.018
Gly_c													1.310
Glu_c													24.000
KG_c													0.400
GOX_c													0.007
GA_c													0.717
$SucV$								65.265					

▨ – fixed value

Table A.7.: Eigenvalues of the Jacobian matrix evaluated in steady state for Farquhar *et al.* [73], Medlyn *et al.* [74], Schultz [75], Sharkey *et al.* [76], Damour and Urban [77] and Zhu *et al.* (2009) [78].

model	Farquhar	Medlyn	Schultz	Sharkey	Damour	Zhu '09	
eigen-values	-5.9225	-4.8578	-2.9879	-0.0308	-3.5556	-29.3775	-1.2750
	-0.3448	-0.2586	-0.0849	-0.0015	-0.4584	-14.7480	-0.0034
						-1.8583	

Table A.8.: Eigenvalues of the Jacobian matrix evaluated in steady state for Giersch *et al.* [79].

-9.6899	-7.0508	-2.4711	-0.5896	-0.2578	0	$1.6336e-18$

Table A.9.: Eigenvalues of the Jacobian matrix evaluated in steady state for Hahn [80].

-3.1561	$-0.4138\pm0.1867i$	$-2.8424e-04$	$-5.6222e-15$
-2.2793	$-0.2201\pm0.1306i$	$-4.8449e-05$	$-4.5568e-17$
-0.6543	$-0.0607\pm0.1000i$	$-4.5727e-05$	$-8.1063e-19$
-0.2567	$-0.0041\pm0.0045i$	$-7.1036e-14$	

Table A.10.: Eigenvalues of the Jacobian matrix evaluated in steady state for Poolman *et al.* [60].

$-1.3253e+11$	$-1.4569e+09$	$-1.3497e+08$	$-1.3101e+07$	$-519.4952\pm3.2137i$	-10.3246
$-9.1330e+09$	$-7.0430e+08$	$-7.0136e+07$	$-2.0364e+03$	-132.4045	-6.9863
$-2.5276e+09$	$-5.1255e+08$	$-5.4122e+07$			

Table A.11.: Eigenvalues of the Jacobian matrix evaluated in steady state for Laisk *et al.* (2006) [56].

$-1.0352e+03$	-501.2538	-65.8395	-7.2501	$-5.8453\pm1.0905i$	-0.9449	-0.0386
$-1.0727e+03$	-356.8311	-18.4438	-6.6926	$-3.7797\pm0.4502i$	-0.6561	-0.1181
$-1.0809e+03$	-88.3300	-12.3621	-2.6696	-1.9390	-0.5654	-0.1613

Table A.12.: Eigenvalues of the Jacobian matrix evaluated in steady state for Zhu *et al.* (2007) [51].

-1933.3014	-27.6033	-3.5501	-1.5939	-0.4097	$-0.0665\pm0.0975i$	-0.0182
-128.1933	-20.6040	-3.4332	-1.4632	-0.3925	$-0.0556\pm0.0231i$	$-7.8464e-04$
-54.5450	-7.2862	-2.8035	-0.8208	-0.2089	-0.0351	
-60.3986	-6.5389	-2.1641	-0.7207	-0.1716	-0.0275	

Table A.13.: Metabolomics data of *Arabidopsis* under light with 210 ppm CO_2 [88].

| Metabolite | in $\frac{nmol}{g \text{ FW}}$ | | in mM | | | |
| | | | Stroma | | Cytosol | |
	Mean	SD	Relative	Absolute	Relative	Absolute
RuBP	98.6	1.1	1	1.52	0	0
PGA	198	11	0.46	2.13	0.54	2.48
GAP*	2.59	0.47	0.24	0.02	0.76	0.06
DHAP	51.7	6.7	0.24	0.37	0.76	1.15
FBP	23.4	1.8	0.77	0.32	0.23	0.1
F6P	113	7	0.28	0.89	0.72	2.29
G6P	239	16	0.08	0.7	0.92	8.07
G1P**	11	0.7	0.08	0.03	0.92	0.37
ADPG	0.607	0.088	1	0.01	0	0
SBP	32.6	3.2	1	0.5	0	0
S7P	79.9	5	1	1.23	0	0
R5P	3.53	0.75	1	0.05	0	0
X5P+Ru5P	33.3	6.1	1	0.51	0	0
ATP***	117	5	0.22	0.79	0.78	2.73
ADP	25	1.5	0.44	0.26	0.56	0.34
$UDPG_c$	144	12	0.03	0.18	0.97	5.52

* – calculated by using GAP = 0.05 DHAP and allocated like DHAP

** – allocated like G6P

*** – allocation calculated via Stitt *et al.* (1982) [90]

Table A.14.: Categories of reduced and merged MCA matrices.

| Incorporated reactions | | Incorporated substrates | | Cate- |
Names	#	Names	#	gory
RuBisCO, residual CBC	2	PGA	1	s_1r_2
RuBisCO, PGA RED, RuBP REG	3	PGA, TP, Ru5P, RuBP	4	s_4r_3
RuBisCO, PGA RED, RuBP REG, TPT	4	PGA, TP, Ru5P, RuBP	4	s_4r_4
RuBisCO, PGA RED, TP RED, RuBP REG, TPT	5	PGA, TP, Ru5P, RuBP	4	s_4r_5
RuBisCO, PGA RED, RuBP REG, TPT, ATP syn.	5	PGA, TP, Ru5P, RuBP, ATP	5	s_5r_5
CBC^r, TPT, ATP syn., starch syn.	12	CBC^s, ATP	10	$s_{10}r_{12}$
CBC^r, TPT, ATP syn., starch syn., sucrose syn.	19	CBC^s, ATP, sucrose^s, PGAc	17	$s_{17}r_{19}$

CBC^r: RuBisCO, PGAK, SFBA, FBPase, F6P TK, SBPase, S7P TK, PRK

CBC^s: PGA, TP, FBP, HeP, E4P, SBP, S7P, P5P, RuBP

sucrose^s: TP_c, FBP_c, $F26BP_c$, HeP_c, $UDPG_c$, $SucP_c$

Table A.15.: Number and names of models available for each category of reduced and merged MCA matrices.

Cate-gory	#	Model					
		Zhu '09 [78]	Giersch [79]	Hahn [80]	Poolman [60]	Laisk '06 [56]	Zhu '07 [51]
s_4r_3	6	✓	✓	✓	✓	✓	✓
s_4r_4	6	✓	✓	✓	✓	✓	✓
s_4r_5	4	✓	✓		✓		✓
s_5r_5	5		✓	✓	✓	✓	✓
$s_{10}r_{12}$	3				✓	✓	✓
$s_{17}r_{19}$	2					✓	✓

Table A.16.: Kendall coefficients τ of Elasticity Coefficients for the category considering 4 substrates and 3 reactions (s_4r_3).

	Zhu '09 [78]	Giersch [79]	Hahn [80]	Poolman [60]	Laisk '06 [56]	Zhu '07 [51]
Zhu '09 [78]	╲	0.5679	0.9334	0.3679	0.5314	0.2861
Giersch [79]	0.5679	╲	0.5065	0.0529	0.7057	0.0176
Hahn [80]	0.9334	0.5065	╲	0.3935	0.5592	0.3521
Poolman [60]	0.3679	0.0529	0.3935	╲	0.1905	0.7143
Laisk '06 [56]	0.5314	0.7057	0.5592	0.1905	╲	0.2222
Zhu '07 [51]	0.2861	0.0176	0.3521	0.7143	0.2222	╲

Table A.17.: Kendall coefficients τ of Flux Control Coefficients for the category considering 4 substrates and 3 reactions (s_4r_3).

	Zhu '09 [78]	Giersch [79]	Hahn [80]	Poolman [60]	Laisk '06 [56]	Zhu '07 [51]
Zhu '09 [78]	╲	0.3333	0.3333	−0.3333	0.3333	−0.3333
Giersch [79]	0.3333	╲	−0.3015	−0.8660	0.8660	−0.8660
Hahn [80]	0.3333	−0.3015	╲	0.2031	−0.3333	0.3333
Poolman [60]	−0.3333	−0.8660	0.2031	╲	−0.6667	0.7222
Laisk '06 [56]	0.3333	0.8660	−0.3333	−0.6667	╲	−0.6111
Zhu '07 [51]	−0.3333	−0.8660	0.3333	0.7222	−0.6111	╲

Table A.18.: Kendall coefficients τ of Elasticity Coefficients for the category considering 4 substrates and 4 reactions (s_4r_4).

	Zhu '09 [78]	Giersch [79]	Hahn [80]	Poolman [60]	Laisk '06 [56]	Zhu '07 [51]
Zhu '09 [78]	╲	0.5413	0.9136	0.4013	0.2637	0.3096
Giersch [79]	0.5413	╲	0.5681	0.1412	0.6153	0.1210
Hahn [80]	0.9136	0.5681	╲	0.4227	0.3288	0.3757
Poolman [60]	0.4013	0.1412	0.4227	╲	0.1880	0.7949
Laisk '06 [56]	0.2637	0.6153	0.3288	0.1880	╲	0.2222
Zhu '07 [51]	0.3096	0.1210	0.3757	0.7949	0.2222	╲

Table A.19.: Kendall coefficients τ of Flux Control Coefficients for the category considering 4 substrates and 4 reactions (s_4r_4).

Zhu '09 [78]	╲	−0.0471	0.3425	0.4104	0.6498	0.0855
Giersch [79]	−0.0471	╲	−0.4352	−0.5597	0.2110	−0.7615
Hahn [80]	0.3425	−0.4352	╲	0.2284	0.1933	0.5446
Poolman [60]	0.4104	−0.5597	0.2284	╲	0.1500	0.3833
Laisk '06 [56]	0.6498	0.2110	0.1933	0.1500	╲	0.0333
Zhu '07 [51]	0.0855	−0.7615	0.5446	0.3833	0.0333	╲

Table A.20.: Kendall coefficients τ of Elasticity Coefficients for the category considering 4 substrates and 5 reactions (s_4r_5).

	Zhu '09 [78]	Giersch [79]	Poolman [60]	Zhu '07 [51]
Zhu '09 [78]	╲	0.4675	0.3800	0.3254
Giersch [79]	0.4675	╲	0.2008	0.0484
Poolman [60]	0.3800	0.2008	╲	0.7454
Zhu '07 [51]	0.3254	0.0484	0.7454	╲

Table A.21.: Kendall coefficients τ of Flux Control Coefficients for the category considering 4 substrates and 5 reactions (s_4r_5).

	Zhu '09 [78]	Giersch [79]	Poolman [60]	Zhu '07 [51]
Zhu '09 [78]	╲	0.0582	0.1943	0.1318
Giersch [79]	0.0582	╲	−0.6988	−0.4760
Poolman [60]	0.1943	−0.6988	╲	0.2387
Zhu '07 [51]	0.1318	−0.4760	0.2387	╲

Table A.22.: Kendall coefficients τ of Elasticity Coefficients for the category considering 5 substrates and 5 reactions (s_5r_5).

	Giersch [79]	Hahn [80]	Poolman [60]	Laisk '06 [56]	Zhu '07 [51]
Giersch [79]	╲	0.5200	0.1047	0.3791	0.1166
Hahn [80]	0.5200	╲	0.4568	0.3665	0.5131
Poolman [60]	0.1047	0.4568	╲	0.4366	0.9002
Laisk '06 [56]	0.3791	0.3665	0.4366	╲	0.4444
Zhu '07 [51]	0.1166	0.5131	0.9002	0.4444	╲

Table A.23.: Kendall coefficients τ of Flux Control Coefficients for the category considering 5 substrates and 5 reactions (s_5r_5).

Giersch [79]	╲	−0.3859	−0.0357	0.2996	−0.4753
Hahn [80]	−0.3859	╲	0.0748	0.0953	0.4771
Poolman [60]	−0.0357	0.0748	╲	0.2267	0.1770
Laisk '06 [56]	0.2996	0.0953	0.2267	╲	−0.1703
Zhu '07 [51]	−0.4753	0.4771	0.1770	−0.1703	╲

Table A.24.: Kendall coefficients τ of Elasticity Coefficients for the category considering 10 substrates and 12 reactions ($s_{10}r_{12}$).

	Poolman [60]	Laisk '06 [56]	Zhu '07 [51]
Poolman [60]	╲	0.6176	0.7951
Laisk '06 [56]	0.6176	╲	0.6344
Zhu '07 [51]	0.7951	0.6344	╲

Table A.25.: Kendall coefficients τ of Flux Control Coefficients for the category considering 10 substrates and 12 reactions ($s_{10}r_{12}$).

Poolman [60]	╲	0.4042	0.2876
Laisk '06 [56]	0.4042	╲	0.0057
Zhu '07 [51]	0.2876	0.0057	╲

Table A.26.: Kendall coefficients τ for the category considering 17 substrates and 19 reactions ($s_{17}r_{19}$).

	Elasticity		*Flux Control*	
	Laisk '06 [56]	Zhu '07 [51]	Laisk '06 [56]	Zhu '07 [51]
Laisk '06 [56]	╲	0.6140	╲	0.2202
Zhu '07 [51]	0.6140	╲	0.2202	╲
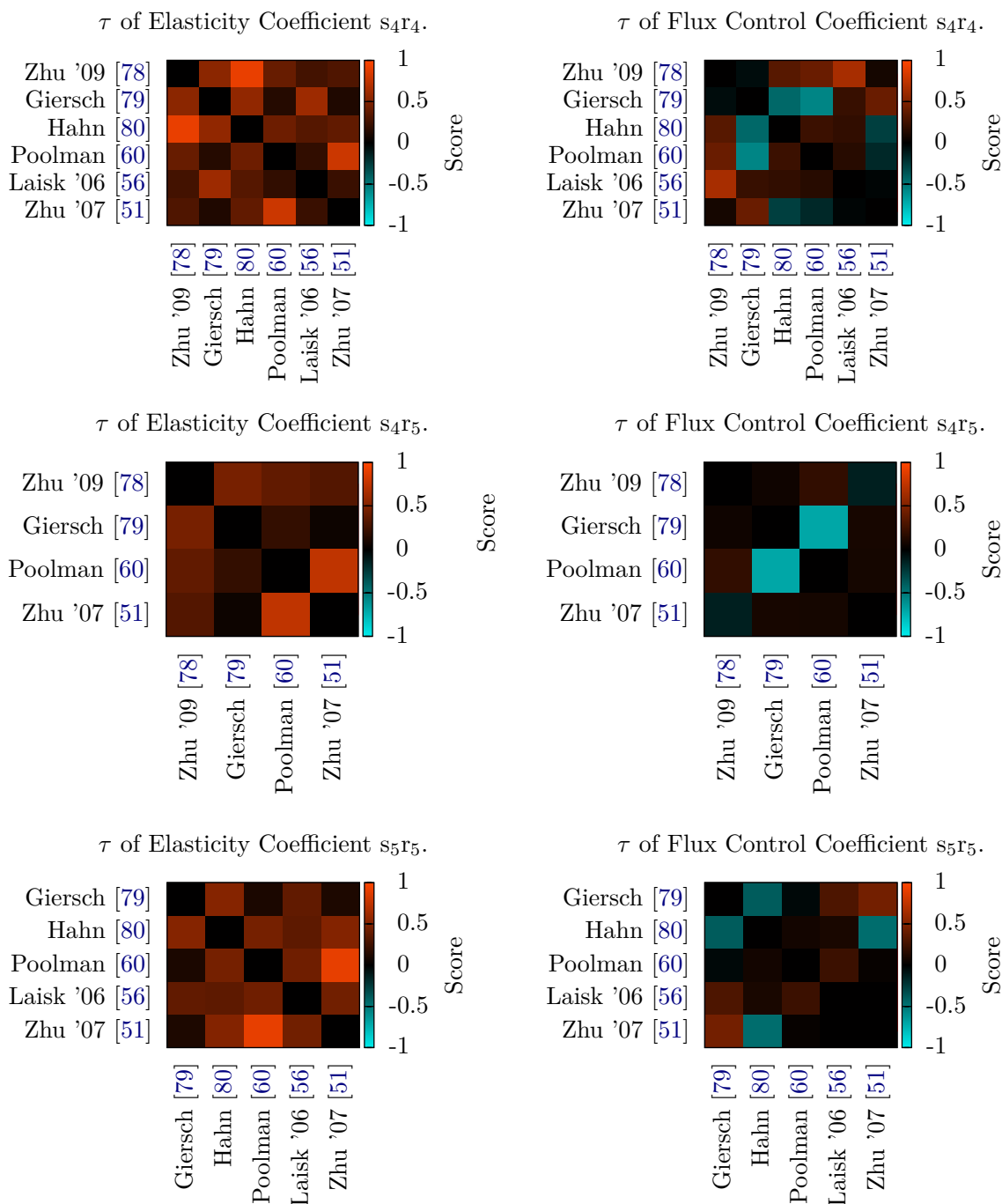
Figure A.1.: Kendall τ's for pairwise comparison where: red indicates very similar, black neutral and turquoise very different dynamic behavior (legend right hand side).

τ of Elasticity Coefficient $s_{10}r_{12}$.

τ of Flux Control Coefficient $s_{10}r_{12}$.

τ of Elasticity Coefficient $s_{17}r_{19}$.

τ of Flux Control Coefficient $s_{17}r_{19}$.
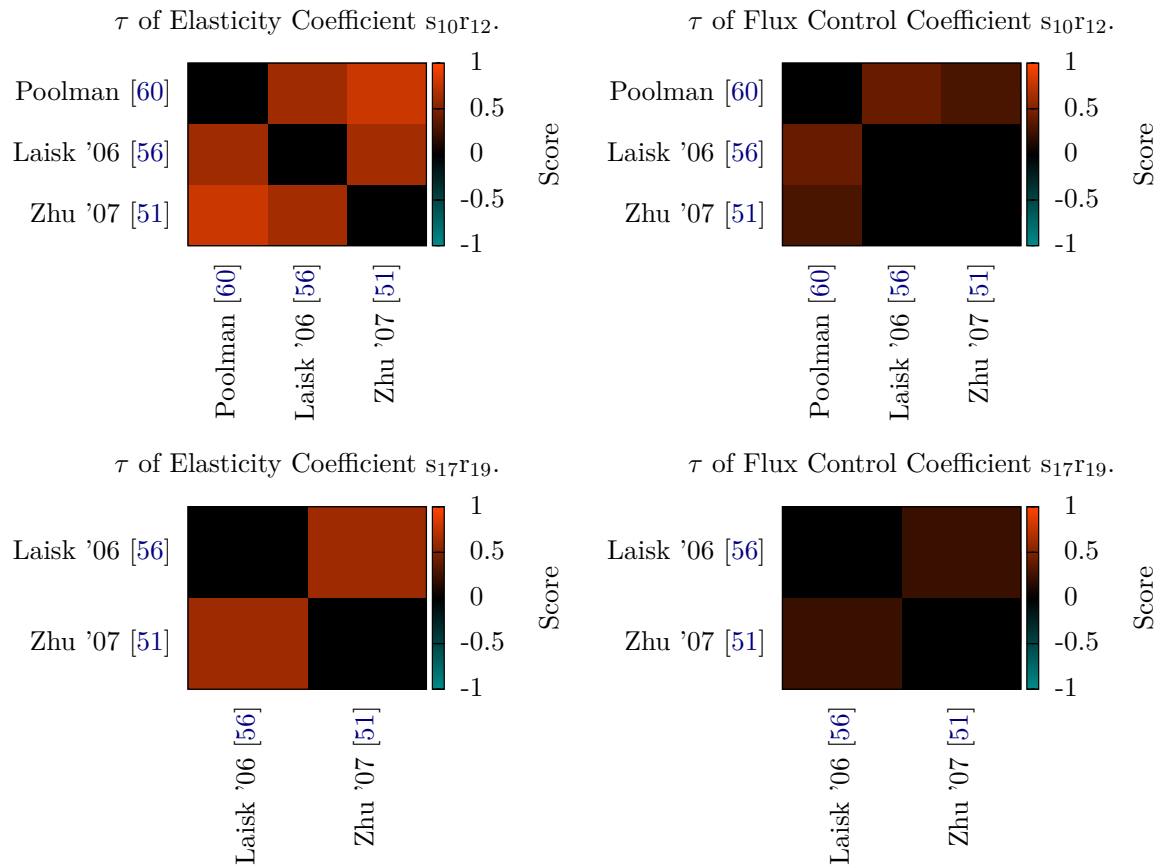
Figure A.2.: Kendall τ's for pairwise comparison where: red indicates very similar, black neutral and turquoise very different dynamic behavior (legend right hand side).

Table A.27.: Averaged Kendall τ of of Elasticity, Flux Control Coefficients and their combination (Equation 2.10).

Averaged Kendall τ of Elasticity Coefficients, $\bar{\tau}_\mathrm{E}$					
	Zhu '09	Giersch	Hahn	Poolman	Laisk '06	Zhu '07
Zhu '09	–	0.53	**0.92**	0.38	0.40	0.31
Giersch	0.53	–	0.53	0.12	**0.57**	0.08
Hahn	**0.92**	0.53	–	0.42	0.42	0.41
Poolman	0.38	0.12	0.42	–	0.36	**0.79**
Laisk '06	0.40	**0.57**	0.42	0.36	–	0.43
Zhu '07	0.31	0.08	0.41	**0.79**	0.43	–

Averaged Kendall τ of Flux Control Coefficients, $\bar{\tau}_\mathrm{F}$					
	Zhu '09	Giersch	Hahn	Poolman	Laisk '06	Zhu '07
Zhu '09	–	0.10	0.28	0.17	**0.48**	0.23
Giersch	0.10	–	-0.37	-0.54	**0.46**	0.31
Hahn	**0.28**	-0.37	–	0.17	-0.05	-0.15
Poolman	**0.17**	-0.54	0.17	–	0.03	-0.15
Laisk '06	**0.48**	0.46	-0.05	0.03	–	-0.03
Zhu '07	0.23	**0.31**	-0.15	-0.15	-0.03	–

Averaged Kendall τ of Elasticity and Flux Control Coefficients, $\bar{\tau}_\mathrm{av}$					
	Zhu '09	Giersch	Hahn	Poolman	Laisk '06	Zhu '07
Zhu '09	–	0.31	**0.60**	0.28	0.44	0.27
Giersch	0.31	–	0.08	-0.21	**0.51**	0.19
Hahn	**0.60**	0.08	–	0.30	0.18	0.13
Poolman	0.28	-0.21	0.30	–	0.19	**0.32**
Laisk '06	0.44	**0.51**	0.18	0.19	–	0.20
Zhu '07	0.27	0.19	0.13	**0.32**	0.20	–

number of samples: ▨ – 2, ▨ – 3, ▨ – 4, ▨ – 5   ▨ – mutually most similar

A.3. Ranking

Based on our classification and analyses, the models of the compendium can be ranked in several ways. By selecting the criteria of interest the constructed compendium just comprises the relevant models. However, we have carried out the ranking of the whole set of models.

First, we have scored the analysis criteria by their relevance: (i) stability, with weighting factor 4, (ii) compliance to data, by factor 2, and (iii) robustness, weighted with 1. Therewith, we have provided the preliminary ranking (Equation A.33, Table A.28 columns 3,5,7 and 9) which is then combined with the fourth part of the analysis—the similarity measure. The similarity values ($\bar{\tau}_{\mathrm{av}}$) of the mutually most similar models are used to weight the preliminary rank position of these model with each other (Equation A.34):

$$\mathrm{Rank}_{\mathrm{pre}}(M) = 4 \cdot \mathrm{Rank}_{\mathrm{stable}}(M) + 2 \cdot \mathrm{Rank}_{\mathrm{RSS}}(M) + \mathrm{Rank}_{\mathrm{robust}}(M) \tag{A.33}$$

$$\mathrm{Rank}_{\mathrm{final}}(M) = \left(1 - \frac{\bar{\tau}_{\mathrm{av}}}{2}\right) \cdot \mathrm{Rank}_{\mathrm{pre}}(M) + \frac{\bar{\tau}_{\mathrm{av}}}{2} \cdot \mathrm{Rank}_{\mathrm{pre}}(\text{most similar to } M) \tag{A.34}$$

Therefore, the similarity value is used as the weighting factor, rescaled to $\left[0, \frac{1}{2}\right]$ whereby $\frac{1}{2}$ corresponds to perfect similarity ($\bar{\tau}_{\mathrm{av}} = 1$) and 0 to no similarity at all ($\bar{\tau}_{\mathrm{av}} = 0$). As the result of this, we have yielded the value for the final ranking (Table A.28).

Appendix A. A quantitative comparison of Calvin-Benson cycle models

Table A.28.: Complete results and model rankings of the CBC models. The results of stability, robustness, RSS and sensitivity analyses are included in columns 2, 4, 6, and 11. Column 10 comprise the most similar model. The single, preliminary and final model rankings are shown in columns 3, 5, 7, 8–9, and 12–13, respectively. Different shadings indicate the importance of the rank and the candidates for metabolic engineering.

Model	Stability		RSS		Robustness		Preliminary ranking		Similarity	Final ranking		Refs
	Result	Score	Result	Score	Result	Score	Unweighted	Weighted		Unweighted	Weighted	
Farquhar	✓	1	11.0248	4	0.41557	4	1	1	—	1	1	[73]
Poolman	✓	1	0.7460	2	0.01873	8	2	1	Zhu '07 0.32	4	2	[60]
Giersch	✓	1	11.8853	5	0.02934	7	3	3	Laisk '06 0.51	5	3	[79]
Medlyn	✓	1	14.9133	9	0.59599	3	3	4		2	4	[74]
Laisk '06	✓	1	14.4639	7	0.01500	9	8	5	Giersch 0.51	7	5	[56]
Damour	✓	1	55.5920	11	1	1	3	5		2	6	[77]
Hahn	✓	1	14.3666	6	0	11	9	5		7	7	[80]
Zhu '07	✓	1	14.5347	8	0.00042	10	10	9	Poolman 0.32	9	8	[51]
Zhu '09	✓	1	25.4759	10	0.06418	5	7	8	Hahn 0.60	8	9	[78]
Schultz	✓	1	994.0520	12	0.66280	2	6	9		6	10	[75]
Sharkey	✓	1	45329	13	0.03036	6	11	11		11	11	[76]
Pettersson	✗	12	0.2811	1	—	12	12	12		12	12	[58]
Woodrow	✗	12	4.2994	3	—	12	13	13		13	13	[59]
Laisk '89	+	12	—	14	—	12	14	14		14	14	[61]
Fridlyand	+	12	—	14	—	12	14	14		14	14	[57]

✓ stable   ✗ unstable   + no steady state at all

Appendix B.

Bottom-up metabolic reconstruction of *Arabidopsis thaliana* and its application to determining the metabolic costs of enzyme production[11]

B.1. Model characteristics

The assembled model is a large-scale metabolic network of Arabidopsis following the bottom-up reconstruction process. In contrast to genome-scale networks, the focus lies on the functionality and completeness of the metabolic objective, *e.g.*, amino acid synthesis, and the underlying pathways (Figure 5.2, Supplemental File 2 – Reaction list). The included reactions are assigned according to AraCyc 11.5 [32]. Exceptions are the light-dependent reactions, the importer and exporter, and the internal transport reactions. Moreover, the consideration of different environmental conditions and their effects on the cellular scenario are a main aspect of this study. The latter issue is modeled by different biomass compositions, one for each cellular scenario.

B.1.1. Light-dependent reactions

Metabolic databases, such as AraCyc [32] and KEGG [223], do not incorporate the proton pump into the thylakoid lumen in the context of the electron transport chain as well as the one of the ATP synthase back into the stroma. Four electrons enter the noncyclic electron transport chain per dissociated water molecule, and are utilized to form two NADPH molecules. Moreover, 12 protons in total are pumped into the lumen [315]. On the other hand, the corresponding ATP synthase needs 14 protons for a full rotation forming three molecules of ATP [227]. In contrast, the mitochondrial ATP synthase has a 12-fold symmetry of proton-powered turbine forming turbine.

B.1.2. Import and export reactions

The corresponding reactions are initially assigned as reversible and are then restricted to achieve a minimal in- and efflux of the system. For the first time, the uptake of the

---

required ions nitrate, sulfate and phosphate is coupled with energy demand. It is known that the uptake of these ions is a proton coupled co-transport. To maintain the intercellular pH value those protons have to be exported, again. Assuming that the pH value is only regulated by an ATP-driven proton pump, the the costs of exporting each proton can be approximated by one ATP. The stoichiometry of the co-transport differs for the different ions: 2 H/NO$_3$ [316, 317], 3 H/SO$_4$ [318], and 2–4 H/H$_2$PO$_4$ [319]. In the case of Pi, the determined physiological charge formula is HPO$_4$ which entails an adaption of the co-transport stoichiometry. Moreover, it is experimentally confirmed that the lower the phosphate concentration, the greater the ratio of protons and phosphate ions [320]. For this purpose, we assigned to the import of phosphate a cost of three ATP.

B.1.3. Internal transporter

For the incorporation of transport reactions the aim is twofold: include a minimal set of reactions, and avoid biologically unverified transporters. Therefore, we started with confirmed transport reactions, extended the list as far as required, and inspected it subsequently to remove redundant transporter if possible. The respective references are given in the reaction list (Supplemental File 2 – Reaction list).

B.1.4. Biomass reactions

To reinforce the biochemical reliability of a metabolic model, the careful assembly of the metabolic function to examine is of great importance. Therefore, we assembled three biomass compositions simulating growing cells exposed under carbon-limiting, nitrogen-limiting and optimal growth conditions. We would like to highlight again that, in contrast to the biomass compositions utilized in the other Arabidopsis reconstructions, the here presented compositions are experimentally obtained predominantly from plant fresh weight and, subsequently, converted into dry weight (DW) in a condition-specific manner (Table B.1). We accounted for all major cell components, namely cell wall, proteins, lipids,

Table B.1.: Condition-specific dry weight content (DWC; $\left[\frac{\text{g DW}}{\text{g FW}}\right]$) of Arabidopsis.

Condition	DWC	Reference
Optimal growth	0.088	[229, high N]
Carbon limitation	0.088	[321, data set S2: 24°C/16°C]
Nitrogen limitation	0.083	[229, low N]

soluble metabolites, starch, DNA and RNA. All of these components are incorporated via representative metabolites or precursors (Tables B.2–B.5). For instance, cell wall and starch are represented by cellulose and amylose, respectively, while a Glc dimer precursor of each is incorporated. Similarly, lipids, proteins, DNA and RNA are included. Regarding the lipids, we considered palmitic acid as a fatty acid representative and incorporated malonyl-acyl carrier protein (ACP) as the corresponding precursor, whereas for cell protein, and DNA and RNA only components, namely amino acids and nucleotides, respectively, are considered as representatives.

The experimental basis is the condition-specific metabolomics data set of [232] providing data for total protein, soluble metabolites and starch (Tables B.2–B.4). The soluble metabolite data are measured by GC-MS which, consequently, first have to be converted into absolute levels by means of calibration curves for each metabolite (Table B.2, Supplemental Data 1 – GC-MS data). For the protein composition, we used a proteomics data set of

Table B.2.: Soluble metabolite representatives measured in FW and converted into DW for model integration. The complete data set is provided as Supplemental Data 1 – GC-MS data

Metabolite		Carbon			Nitrogen			Optimal		
		$\left[\frac{\mu g}{g\,DW}\right]$	$\left[\frac{nmol}{g\,DW}\right]$	$\left[\frac{nmol}{g\,FW}\right]$	$\left[\frac{\mu g}{g\,DW}\right]$	$\left[\frac{nmol}{g\,DW}\right]$	$\left[\frac{nmol}{g\,FW}\right]$	$\left[\frac{\mu g}{g\,DW}\right]$	$\left[\frac{nmol}{g\,DW}\right]$	$\left[\frac{nmol}{g\,FW}\right]$
free Amino acids	Ala	59.20	664.43	58.47	10.26	115.16	9.56	11.18	125.47	11.04
	Arg	606.75	3483.04	306.51	6882.91	39511.56	3279.46	1218.57	6995.22	615.58
	Asn	196.53	1487.49	130.90	1119.68	8474.71	703.40	433.54	3281.42	288.76
	Asp	95.62	718.39	63.22	243.81	1831.76	152.04	98.56	740.46	65.16
	Cys									
	Gln	154.58	1057.76	93.08	1216.50	8324.18	690.91	214.16	1465.46	128.96
	Glu	182.32	1239.17	109.05	527.34	3584.19	297.49	371.84	2527.29	222.40
	Gly	88.36	1177.02	103.58	9.19	122.35	10.16	60.08	800.32	70.43
	His									
	Ile	15.28	116.52	10.25	42.24	322.04	26.73	17.31	131.97	11.61
	Leu	12.19	92.97	8.18	30.70	234.06	19.43	12.73	97.06	8.54
	Lys	35.02	239.58	21.08	275.01	1881.16	156.14	43.04	294.44	25.91
	Met	69.14	463.36	40.78	157.36	1054.64	87.54	39.21	262.82	23.13
	Phe	89.17	539.83	47.50	223.73	1354.40	112.41	125.92	762.28	67.08
	Pro	51.64	448.53	39.47	42.46	368.78	30.61	60.72	527.37	46.41
	Ser	128.72	1224.90	107.79	83.13	791.07	65.66	87.52	832.79	73.29
	Thr	1035.03	8688.94	764.63	777.43	6526.40	541.69	778.63	6536.54	575.22
	Trp	52.10	255.11	22.45	302.80	1482.62	123.06	68.84	337.09	29.66
	Tyr	76.86	424.19	37.33	112.54	621.13	51.55	55.39	305.70	26.90
	Val	20.32	173.43	15.26	31.03	264.85	21.98	21.18	180.79	15.91
Sugars	Fru	2661.71	14774.15	1300.13	1922.14	10669.09	885.53	2168.21	12034.90	1059.07
	Glc	6098.31	33849.42	2978.75	4240.82	23539.17	1953.75	3329.12	18478.68	1626.12
	Mas	779.56	2277.42	200.41	1841.45	5379.65	446.51	893.26	2609.58	229.64
	Suc	15539.82	45398.23	3995.04	29785.27	87015.11	7222.25	10497.78	30668.37	2698.82
TCA	Fum	9005.16	77583.91	6827.38	1344.95	11587.40	961.75	7800.75	67207.27	5914.24
	Mal	5110.44	38111.99	3353.86	1461.39	10898.56	904.58	6838.11	50996.41	4487.68
	SCA	153.75	1301.99	114.57	192.48	1629.92	135.28	116.56	987.09	86.86
Others	GABA	15.78	153.05	13.47	60.14	583.23	48.41	9.69	94.01	8.27
	Orn	1930.50	14607.27	1285.44	1274.92	9646.82	800.69	209.68	1586.54	139.62
	SA	54.32	311.89	27.45	33.19	190.59	15.82	66.05	379.24	33.37
	Tre	438.18	1158.18	101.92	1068.54	2824.36	234.42	218.43	577.36	50.81
	urea	332.23	5531.64	486.78	622.54	10365.35	860.32	170.60	2840.43	249.96

[259]. To this end, we determined the amino acid composition of the identified proteins in Arabidopsis leaves and divided the total amount of proteins into amino acid fractions (see Supplemental Data 2 – 2D-Gel data). Combined with the total protein data of [232],

we assigned the allocation to the amino acids (Table B.3). As there is no condition-

Table B.3.: Fractional amount of protein-bound amino acids of Arabidopsis leaf based on measured total protein [232, last row].

Amino acid	Fraction [%]	Carbon $\left[\frac{mg}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,FW}\right]$	Nitrogen $\left[\frac{mg}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,FW}\right]$	Optimal $\left[\frac{mg}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,FW}\right]$
Ala	8.4253	18.42	206.80	18.20	31.26	350.90	29.12	20.54	230.55	20.29
Arg	4.8366	10.58	60.71	5.34	17.95	103.02	8.55	11.79	67.69	5.96
Asn	3.6452	7.97	60.33	5.31	13.53	102.38	8.50	8.89	67.26	5.92
Asp	5.476	11.97	89.97	7.92	20.32	152.66	12.67	13.35	100.30	8.83
Cys	1.5465	3.38	27.91	2.46	5.74	47.36	3.93	3.77	31.12	2.74
Gln	2.9852	6.53	44.67	3.93	11.08	75.80	6.29	7.28	49.80	4.38
Glu	6.872	15.03	102.14	8.99	25.50	173.31	14.38	16.75	113.87	10.02
Gly	8.3967	18.36	244.59	21.52	31.16	415.04	34.45	20.47	272.68	24.00
His	2.1157	4.63	29.82	2.62	7.85	50.60	4.20	5.16	33.24	2.93
Ile	4.826	10.55	80.46	7.08	17.91	136.52	11.33	11.77	89.69	7.89
Leu	8.5374	18.67	142.33	12.52	31.68	241.51	20.05	20.81	158.67	13.96
Lys	6.1944	13.55	92.66	8.15	22.98	157.23	13.05	15.10	103.30	9.09
Met	2.0413	4.46	29.92	2.63	7.57	50.76	4.21	4.98	33.35	2.93
Phe	4.4653	9.76	59.11	5.20	16.57	100.30	8.33	10.89	65.90	5.80
Pro	5.0772	11.10	96.43	8.49	18.84	163.63	13.58	12.38	107.51	9.46
Ser	6.3223	13.83	131.56	11.58	23.46	223.23	18.53	15.41	146.67	12.91
Thr	6.3592	13.91	116.74	10.27	23.60	198.09	16.44	15.50	130.15	11.45
Trp	1.3851	3.03	14.83	1.31	5.14	25.17	2.09	3.38	16.53	1.45
Tyr	3.4726	7.59	41.91	3.69	12.89	71.12	5.90	8.47	46.72	4.11
Val	7.0199	15.35	131.04	11.53	26.05	222.35	18.46	17.11	146.09	12.86
Total	100	218.68	1803.92	158.75	371.06	3060.97	254.06	243.79	2011.09	176.98

specific proteomics data set available, the resulting partition into amino acids was used for all conditions. The ultimate amino acid fraction in the biomass composition is the combination of free and protein-bound amino acids (Tables B.2 and B.3, Supplemental File 2 – Reaction list). In a similar condition-unspecific manner also the remaining components were incorporated, namely cell wall, lipids, and DNA and RNA. The cell wall fraction was approximated by 118 µg cellulose mg^{-1} DW [258, Fig. 3A] which was converted into 363.88 µmol Glc units of cellulose g^{-1} DW by considering a molar mass of anhydroglucose, 162.14 g mol^{-1} [322] (Table B.4). For the lipid content, the amount of fatty acids was used amounting to 3.3 mg g^{-1} FW in leaves of wild type Arabidopsis [260]. Based on the specific dry weight content (Table B.1), the molar mass of palmitic acid as the major fatty acid, 256.42 g mol^{-1}, and its number of carbon atoms, 16, we calculated the corresponding amount of the three-carbon fatty acid precursor malonyl-ACP, 146.24 and 155.05 µmol g^{-1} DW, respectively, which ultimately serves as the lipid representative (Table B.4). Finally, 0.092 mg DNA g^{-1} FW is measured in Arabidopsis leaves [261] which we also assumed to be an appropriate level for RNA. Based on the DNA and cDNA sequence of Arabidopsis, we determined the nucleotide fraction which was converted into the respective nucleotide levels by means of the molar mass of each nucleotide (Table B.5).

Table B.4.: Starch, cell wall and lipid precursors.

Biomass precursor	Carbon			Nitrogen			Optimal		
	$\left[\frac{\mu g}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,FW}\right]$	$\left[\frac{\mu g}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,FW}\right]$	$\left[\frac{\mu g}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,DW}\right]$	$\left[\frac{\mu mol}{g\,FW}\right]$
Starch (Glc dimer)	73.72	227.35	20.01	108.81	335.54	27.85	95.62	294.86	25.95
Cellulose (Glc dimer)	118.00	363.88	32.02	118.00	363.88	30.20	118.00	363.88	32.02
Fatty acid (Malonyl-ACP)	37.50	146.24	12.87	39.76	155.05	12.87	37.50	146.24	12.87

Table B.5.: Fractional amount of nucleotides of DNA and RNA.

Nucleotide	Total	w/o N[b]	Carbon			Nitrogen			Optimal		
	[%]	[%]	$\left[\frac{mg}{g\,DW}\right]$	$\left[\frac{nmol}{g\,DW}\right]$	$\left[\frac{nmol}{g\,FW}\right]$	$\left[\frac{mg}{g\,DW}\right]$	$\left[\frac{nmol}{g\,DW}\right]$	$\left[\frac{nmol}{g\,FW}\right]$	$\left[\frac{mg}{g\,DW}\right]$	$\left[\frac{nmol}{g\,DW}\right]$	$\left[\frac{nmol}{g\,FW}\right]$
DNA	99.84	100	1.05			1.11			1.05		
dATP	31.94	31.99	0.33	680.91	59.92	0.35	721.93	59.92	0.33	680.91	59.92
dCTP	18.01	18.04	0.19	403.66	35.52	0.20	427.98	35.52	0.19	403.66	35.52
dGTP	17.99	18.02	0.19	371.41	32.68	0.20	393.78	32.68	0.19	371.41	32.68
dTTP	31.90	31.95	0.33	692.81	60.97	0.35	734.55	60.97	0.33	692.81	60.97
RNA[c]	99.99	100	1.05			1.11			1.05		
ATP	28.41	28.41	0.30	585.59	51.53	0.31	620.86	51.53	0.30	585.59	51.53
CTP	22.94	22.94	0.24	496.38	43.68	0.25	526.28	43.68	0.24	496.38	43.68
GTP	20.26	20.26	0.21	404.78	35.62	0.22	429.16	35.62	0.21	404.78	35.62
UTP	28.39	28.39	0.30	613.07	53.95	0.31	650.00	53.95	0.30	613.07	53.95

[b] – IUPAC nucleotide code: any base

[c] – based on cDNA

Table B.6.: Elemental composition of the three biomass functions.

Cellular scenario	Elemental composition $\left[\frac{mmol}{g\,DW}\right]$					
	C	H	N	O	P	S
Optimal growth	20.621	35.505	2.595	15.697	0.013	0.065
Carbon limitation	19.192	32.767	2.350	14.721	0.013	0.058
Nitrogen limitation	26.737	48.002	4.104	19.261	0.014	0.099

Overall, our findings indicate that the three biomass compositions differ in their coverage in a range of 495.08 to 695.78 µg mg^{-1} DW (Table B.7). This coverage specifies the fraction of measured biomass components to the reference fresh weight and results from converting the concentrations of all constituting metabolites into mass fraction and, subsequently, summing them up.

B.1.5. Extended GPR associations

Beyond the common features of a metabolic reconstruction, the Arabidopsis core model provides the stoichiometric subunit compositions of the included enzymes. In the existing genome-scale models, the GPR associations comprise the different genes assigned for each reaction and their linkage, whether they encode alternative or necessary proteins. Accordingly, the relation between isoforms are denoted by logical OR-operators and subunit

Table B.7.: Biomass coverage and free nitrate content for all three conditions

	Carbon limiting		Nitrogen limiting		Optimal growth	
	$\left[\frac{mg}{g\,DW}\right]$	$\left[\frac{mmol}{g\,DW}\right]$	$\left[\frac{mg}{g\,DW}\right]$	$\left[\frac{mmol}{g\,DW}\right]$	$\left[\frac{mg}{g\,DW}\right]$	$\left[\frac{mmol}{g\,DW}\right]$
Total biomass	495.08	3.44	695.78	4.84	533.03	3.67
Free NO_3	38.05	0.61	4.45	0.07	70.30	1.13

connections by logical AND-operators.

For the calculation of enzyme costs, the exact complex structure of an enzyme is required so that we additionally provide the stoichiometry of the subunit assembly. This is easily realized by an additional factor for each multimer, *e.g.*, the large and small subunit term of RuBisCO is complemented by factor eight (Equation B.1).

$$8 * (\texttt{ATCG00490})\ \texttt{AND}\ 8 * (\texttt{AT1G67090 OR AT5G38430 OR AT5G38420 OR AT5G38410})\quad \text{(B.1)}$$

The complex structure information is only rarely available and, moreover, often contradicting across the different databases, AraCyc [32], UniProt [224] and BRENDA [239]. As far as possible, we tried to confirm the utilized information by an additional reference (Supplemental File 2 – Reaction list).

B.2. Analyses and supplementary findings

The analyses within this study are based on constraint-based modeling techniques, in particular flux balance analysis (FBA; [28]). By means of FBA, the optimal synthesis and/or consumption rate of specific metabolites or combinations thereof can be determined (see section 1.4.3, paragraph flux balance analysis).

B.2.1. Constraints and boundaries of the flux balance analyses

All optimizations are based on the photoautotrophic scenario, namely only inorganic carbon, nitrogen, sulfur and phosphorus sources are available, and light is the only energy source. The exact import and export model boundaries are listed in Table B.8.

To compare the compartmentalized Arabidopsis models, we determined the maximum biomass production based on 1000 μmol CO_2 (Tables 5.1 and B.9, no. 1). Beforehand, we deactivated the unneeded import and export reactions to restrict the model boundaries to reactions listed in table B.8. Unfortunately, a comparison based on restricted photon import is not possible for the model of [188] since this restriction has no effect on the production of biomass in this model. Surprisingly, even without any photon import (and no other energy source) this model is able to produce biomass.

The determination of the minimum ATP consumption for cell performance, amino acid and RuBisCO costs was realized by a three-step optimization (Algorithm B.1). First, the minimum amount of required energy to produce the metabolite(s) of interest, in terms of photons, was computed (Algorithm B.1 LP1; Table B.9, no. 7, 10, 18, 21) and utilized as additional constraint in the subsequent optimization steps. Thereby, the amount of required light energy was translated into consumed metabolic energy equivalents. To this

Table B.8.: Model boundary constraints for photoautotrophic scenario. Columns marked with v^{min} and v^{max} represent the default lower and upper boundary, respectively.

	Reaction	v^{min}	v^{max}
98	G6PDH_h	0	0
136	PPIF6PK_c	0	0
446	Im_hnu	0	Inf
447	Im_CO2	-Inf	Inf
448	Im_H2O	-Inf	Inf
449	Im_Pi	0	Inf
450	Im_NO3	0	Inf
451	Im_NH4	0	Inf
452	Im_SO4	0	Inf
453	Im_H2S	0	Inf
454	Ex_O2	-Inf	Inf
455–534	Ex_Ala_c – Ex_Val_p	0	0
535–542	Ex_starch – Ex_Tre	0	0
546–549	Bio_AA – Bio_opt	0	0

end, we maximized the sum of the nongrowth-associated maintenance functions (NGAM) representing the unspecified ATP consuming reactions in the system. On the one hand, by disabling the metabolite's production of interest we determined the maximum conversion rate of photons into ATP (Algorithm B.1 LP2; Table B.9, no. 8, 11, 19, 22), and, on the other hand, by enabling the metabolite's production we were able to calculate the ATP surplus which can be utilized for other processes (Algorithm B.1 LP3; Table B.9, no. 9, 12, 20, 23). Accordingly, the number of consumed ATP results from the maximum amount of ATP that could be synthesized by the provided photons from which the ATP excess by producing the metabolite(s) of interest is subtracted. We note that such multiple-step optimizations may lead to solutions for the second optimization step which are suboptimal in comparison to the case when the solution of the second step would be obtained independently of the first.

B.2.2. Flux variability analysis

As a result of the bottom-up design, the model relies only on annotations from Arabidopsis and does not comprise any dead-ends and blocked reactions. On the other hand, it covers only a limited number of alternative pathways which means that the predicted flux distributions are not as flexible as genome-scale models [186, 188]. This is confirmed by a flux variability analysis (FVA; [323]; see section 1.4.3, paragraph flux variability analysis) of the photoautotrophic setting (Table B.8) maximizing the biomass function of the model of [186]. By means of FBA, the actual flux range of each reaction is determined and is used for predicting the flux variability of the reactions. A reaction is denoted as variable if the actual flux range is more than 1 % of the total flux range meaning maximal minus

---

Algorithm B.1: Three-step optimization of minimum ATP consumption for synthesizing the metabolite(s) of interest.

WLOG:

Import and export reactions are assigned as left-to-right operating such that the transport has positive flux

Input:

Metabolic network of the Arabidopsis core model, including

　　stoichiometric matrix, $\mathbf{S}$,

　　vector of lower and upper flux boundaries, $\mathbf{v}^{min}$ and $\mathbf{v}^{max}$

Initial constraints, *cond* (Table B.8)

Output: minimum ATP consumption for synthesizing the metabolite(s) of interest, $A_{\min}$

Begin

　　(De)Activate reactions according to *cond*,

　　Assign the flux of the export reaction of the respective metabolite(s) of interest, v_s, to one,

　　Minimize the flux through the photon import reaction, $v_e = \{\text{'Im\_hnu'}\}$:

$$z_1 = \min \mathbf{c}^T \mathbf{v}, \qquad c_i = \begin{cases} 1 & \text{if } i \in e \\ 0 & \text{otherwise.} \end{cases}$$

$$s.t. \quad \mathbf{S} \cdot \mathbf{v} = 0 \tag{LP1}$$
$$v_s^{min} = v_s^{max} = 1$$
$$\mathbf{v}^{min} \leq \mathbf{v} \leq \mathbf{v}^{max}$$

　　Define *minimum energy requirement* as $E_{\min} = z_1$

　　Assign the flux of the photon import reaction, v_e, to $E_{\min}$,

　　Maximize the flux through the maintenance reactions, v_m:

$$z_2 = \max \mathbf{c}^T \mathbf{v}, \qquad c_i = \begin{cases} 1 & \text{if } i = m \\ 0 & \text{otherwise.} \end{cases}$$

$$s.t. \quad \mathbf{S} \cdot \mathbf{v} = 0 \tag{LP2}$$
$$v_e^{min} = v_e^{max} = E_{\min}$$
$$\mathbf{v}^{min} \leq \mathbf{v} \leq \mathbf{v}^{max}$$

　　Define *maximum conversion rate of photons int ATP* as $A_{\text{tot}} = z_2$,

　　Assign the flux of the export reaction of the respective metabolite(s) of interest, v_s, to one,

　　Assign the flux of the photon import reaction, v_e, to $E_{\min}$,

　　Maximize the flux through the maintenance reactions, v_m:

$$z_3 = \max \mathbf{c}^T \mathbf{v}, \qquad c_i = \begin{cases} 1 & \text{if } i \in m \\ 0 & \text{otherwise.} \end{cases}$$

$$s.t. \quad \mathbf{S} \cdot \mathbf{v} = 0 \tag{LP3}$$
$$v_s^{min} = v_s^{max} = 1$$
$$v_e^{min} = v_e^{max} = E_{\min}$$
$$\mathbf{v}^{min} \leq \mathbf{v} \leq \mathbf{v}^{max}$$

　　Define *ATP surplus* as $A_{\text{sur}} = z_3$,

　　Set $A_{\min} = A_{\text{tot}} - A_{\text{sur}}$

End

---

minimal reaction boundary (Equation B.2):

$$\text{variability frequency} = \frac{\text{Actual flux range}}{\text{Total flux range}} = \frac{\text{FVA}_{\max} - \text{FVA}_{\min}}{v_{\max} - v_{\min}}. \tag{B.2}$$

Importantly, the maximal and minimal reaction boundary are arbitrarily set to 1000 and -1000, respectively. The variability threshold was chosen to be 1 % as the frequency of variable reactions for all three models can be meaningfully compared within this range (Figure B.1).
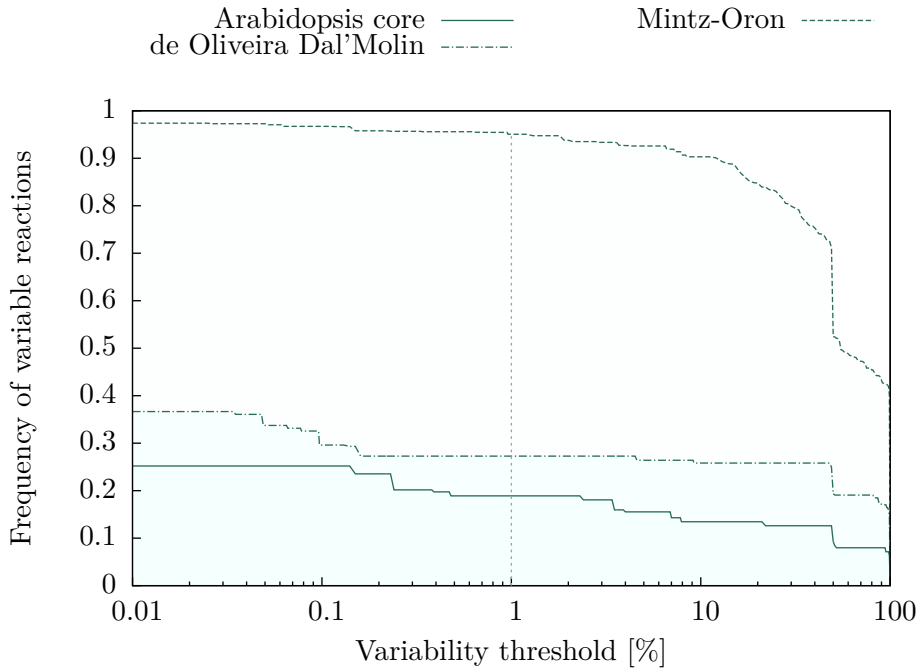
Figure B.1.: Fraction of variable reactions depending on the variability threshold. The gray line denotes the chosen threshold.

B.2.3. Flux coupling analysis

Another approach to make statements regarding the flexibility of metabolic networks is flux coupling analysis (FCA; 43). It describes the functional relation between two reactions, more precisely of their fluxes. By means of FBA, the effects of changing the flux of one coupling partner on the second reaction can be investigated [44]. Thereby, one distinguishes three types of couplings, namely fully-coupled, c_{ful}, partially-coupled, c_{par}, and directionally-coupled, c_{dir}, reactions (see section 1.4.3, paragraph flux coupling analysis). To determine the frequency of coupled reactions, we divided the sum of all coupled pairs of reactions, by the total number of reaction pairs (Equation B.3)

$$\text{coupling frequency} = \frac{\text{number of coupled pairs}}{\text{total number of pairs}} = \frac{\#c_{\text{ful}} + \#c_{\text{par}} + \#c_{\text{dir}}}{\frac{n(n-1)}{2}}, \tag{B.3}$$

where n denotes the number of reactions of the functional network.

B.2.4. Conversion of RuBisCO costs

For the cost estimation of RuBisCO all possible amino acid sequences were considered. As the eight small subunits are encoded by four genes with overall seven splice variants the total number of combinations is $3003 = \left(\binom{7}{8}\right) = \binom{7+8-1}{8} = \binom{7+8-1}{7-1} = \frac{(7+8-1)!}{8!(7-1)!}$.

To get a better feel for the magnitude of RuBisCO costs, we provided several alternative cost measures. By means of the standard Gibbs free energy of ATP synthesis [247] we converted the amount of ATP into kilojoule (kJ; Equations B.4 and B.5)

$$x \text{ nmol ATP} \cdot 36 \, \tfrac{\text{kJ}}{\text{mol ATP}}, \tag{B.4}$$

$$x \text{ molecules of ATP} \cdot 36 \, \tfrac{\text{kJ}}{\text{mol ATP}} \div 6.02214129e23 \, \tfrac{1}{\text{mol}}. \tag{B.5}$$

The costs in terms of sucrose were determined via FBA by computing the ATP costs for producing one molecule of sucrose (Table B.9, no. 21–23). As the optimization is a linear problem, the number of molecules sucrose can be achieved easily by cross multiplication with the 97 molecules of ATP per molecule sucrose.

Similarly, we determined the amount of biomass which can be produced for the energy 1 nmol *de novo* synthesized RuBisCO requires. Within the scope of cell performance, we also determined the amount of ATP required to synthesize one unit of biomass. Here, we relied on the biomass composition for optimal growth conditions. Together with the fact that one unit biomass represent 1 g FW, the amount of alternatively synthesized biomass was calculated by cross multiplication with the RuBisCO costs in terms of ATP.

Finally, we estimated the total number of reactions of a complex RuBisCO and, accordingly, the ATP costs per reaction. Therefore, we calculate the turnover number for RuBisCO by considering that each fourth reaction proceeds with O_2 instead of CO_2 [245]. Thereby we obtain a turnover number, $k_{\text{cat}}^{\text{RBC}}$, of 2.66 to 2.88 s^{-1} (Equation B.7). Assuming a half life of RuBisCO of about seven days [251], we estimated a catalysis capability of 1,605,744 to 1,741,824 reactions for one complex RuBisCO (Equation B.9). Eventually, by dividing the ATP costs per RuBisCO by the total number of reactions one complex of RuBisCO is able to catalyze, we achieved the estimated costs of one RuBisCO reaction (Equation B.11)

$$k\text{cat}^{\text{RBC}} = \tfrac{3}{4} \cdot k\text{cat}^{\text{RBC} - \text{CO}_2} + \tfrac{1}{4} \cdot k\text{cat}^{\text{RBC} - \text{O}_2}, \tag{B.6}$$

$$= \tfrac{3}{4} \cdot 3.4 \text{ to } 3.7 + \tfrac{1}{4} \cdot 0.42 = 2.655 \text{ to } 2.88 \left[\tfrac{1}{s}\right], \tag{B.7}$$

$$r_{\text{tot}} = t_{1/2} \cdot k\text{cat}^{\text{RBC}}, \tag{B.8}$$

$$= 7 \, [d] \cdot 2.655 \text{ to } 2.88 \left[\tfrac{1}{s}\right] = 1,605,744 \text{ to } 1,741,824, \tag{B.9}$$

$$C_{\text{RBC}}^{\text{rxn}} = \frac{C_{\text{RBC}}}{r_{\text{tot}}}, \tag{B.10}$$

$$= \frac{243,287.9 \text{ to } 269,133.9}{1,605,744 \text{ to } 1,741,824} = 0.140 \text{ to } 0.168. \tag{B.11}$$

Table B.9.: Compendium of performed flux balance analyses. Columns marked with c, v^{min} and v^{max} represent objective coefficients, lower and upper boundary for a minimization with the glpk solver, respectively. The initial model boundaries are given in Table B.8.

Description		Objective		Constraints			Comment	No.
		rxn	c	rxn	v^{min}	v^{max}		
Maximize biomass production		AraBio[a]	-1	Im_CO2	1000	1000		1
Cell performance	Minimize photon consumption	Im_hnu	1	Bio[b]	1	1		2
	Minimize CO_2 consumption	Im_CO2	1	Bio[b]	1	1		3
	Minimize NO_3 consumption	Im_NO3	1	Bio[b]	1	1		4
				Im_NH4	0	0	only NO_3	
	Minimize Pi consumption	Im_PO4	1	Bio[b]	1	1		5
	Minimize SO_4 consumption	Im_SO4	1	Bio[b]	1	1		6
				Im_H2S	0	0	only SO_4	
	Minimize ATP consumption	Im_hnu	1	Bio[b]	1	1	optimal value x	7
		NGAM[c]	-1	Im_hnu	x	x		8
				Bio[b]	0	0		
		NGAM[c]	-1	Im_hnu	x	x		9
				Bio[b]	1	1		
AA costs	Minimize ATP consumption	Im_hnu	1	AA[d]	1	1	optimal value x	10
		NGAM[c]	-1	Im_hnu	x	x		11
				AA[d]	0	0		
		NGAM[c]	-1	Im_hnu	x	x		12
				AA[d]	1	1		
RuBisCO costs	Minimize photon consumption	Im_hnu	1	Bio_AA[e]	1	1		13
	Minimize CO_2 consumption	Im_CO2	1	Bio_AA[e]	1	1		14
	Minimize NO_3 consumption	Im_NO3	1	Bio_AA[e]	1	1		15
				Im_NH4	0	0	only NO_3	
	Minimize Pi consumption	Im_PO4	1	Bio_AA[e]	1	1		16
	Minimize SO_4 consumption	Im_SO4	1	Bio_AA[e]	1	1		17
				Im_H2S	0	0	only SO_4	
	Minimize ATP consumption	Im_hnu	1	Bio_AA[e]	1	1	optimal value x	18
		NGAM[c]	-1	Im_hnu	x	x		19
				Bio_AA[e]	0	0		
		NGAM[c]	-1	Im_hnu	x	x		20
				Bio_AA[e]	1	1		
	Minimize ATP consumption	Im_hnu	1	Ex_Suc	1	1	optimal value x	21
		NGAM[c]	-1	Im_hnu	x	x		22
				Ex_Suc	0	0		
		NGAM[c]	-1	Im_hnu	x	x		23
				Ex_Suc	1	1		

[a] AraBio: `Bio_opt` with stoichiometric biomass coefficient of the model of [186]; [b] Bio: an optimization for each biomass function (`Bio_opt`, `Bio_CLim`, `Bio_NLim`); [c] NGAM: for each maintenance function (`NGAM_c`, `NGAM_h`, `NGAM_m`) objective coefficient -1 is assigned; [d] AA: an optimization for each of the 20 amino acids; [e] `Bio_AA`: an optimization for each of the 3003 amino acid compositions as stoichiometric coefficients

B.2.5. Comparison of amino acid cost measures

The complete comparison of the amino acid cost measures are provided in Table B.10.

B.2.6. Extended RuBisCO costs

The minimal and maximal costs and precursor requirements for all 3003 possible amino acid compositions of RuBisCO are provided in Table B.11. Interestingly, except for SO_4 the amino acid sequence denoting the minimal and maximal consumption of precursors and ATP is the same, respectively. The minimal costs correspond to the combination

$$8 * (\texttt{ATCG00490.1}) \text{ AND } 8 * (\texttt{AT1G67090.2}),$$

and the maximal costs to

$$8 * (\texttt{ATCG00490.1}) \text{ AND } 8 * (\texttt{AT5G38410.3}).$$

Table B.10.: Comparison of amino acid cost measures and the corresponding rankings of the Arabidopsis core model (Our costs; for optimal growth conditions), Sajitz-Hermstein and Nikoloski [101, SN], Craig and Weber [99, CW], Akashi and Gojobori [241, AG], Wagner [243, Wag], Seligmann [242, Sel], and Barton et al. [102, Bar]. The 16th column denotes their averaged rank. The columns 17 to 25 capture the amino acid composition with respect to the size of the carbon-skeleton, and the number of incorporated amine groups and sulfur as well as the essential pathways for the amino acid precursors. The columns labeled with CBC, Glyc, PRPP, Pyr, Shik and TCA represent the pathways Calvin-Benson cycle, glycolysis, phosphoribosyl pyrophosphate synthesis, pyruvate decarboxylation, Shikimate pathway and citric acid cycle, respectively.

Amino acid	Our costs [ATP]	SN rank	SN [ATP]	CW rank	CW [ATP]	AG rank	AG [ATP]	Wag rank	Wag [ATP]	Sel rank	Sel [$\frac{g}{mol}$]	Bar rank	Bar [$\frac{mmol}{g\,DW}$]	Average rank	Carbon skeleton	Amine groups	Sulfur presents	CBC	Glyc	PRPP	Pyr	Shik	TCA
Gly	25.16	2	13.33	1	14.5	8	11.7	1	14.5	3	75.07	1	0.31	1	2	1	0	✓	✓				✓
Ala	22.48	1	19.33	2	12.5	6	11.7	1	14.5	3	89.09	3	0.5	2	3	1	0	✓	✓				
Ser	26.83	5	19.33	2	15	9	11.7	1	14.5	3	105.09	2	0.49	3	3	1	0	✓	✓				
Asp	25.16	2	25.33	5	1	1	12.7	4	15.5	7	133.10	4	0.61	4	4	1	0						✓
Asn	26.34	4	26.66	7	4	2	14.7	5	18.5	8	132.12	7	0.79	5	4	2	0		✓				✓
Thr	31.93	7	25.33	5	6	3	18.7	8	21.5	10	119.12	5	0.69	6	4	1	0						✓
Glu	46.01	8	31.33	4	8.5	4	15.3	6	9.5	1	147.13	8	0.86	7	5	1	0						✓
Pro	47.90	11	31.33	8	12.5	6	20.3	9	14.5	3	115.13	11	0.99	8	5	1	0				✓		✓
Gln	46.32	9	32.66	12	9.5	5	16.3	7	10.5	2	146.15	9	0.92	9	5	2	0				✓		✓
Cys	30.13	6	19.33	2	24.5	14	24.7	11	26.5	11	121.16	6	0.75	10	3	1	1		✓		✓		
Val	46.34	10	31.33	8	25	15	23.3	10	29	12	117.15	10	0.96	11	5	1	0		✓				
Met	55.75	13	31.33	8	18.5	10	34.3	16	36.5	15	149.21	14	1.25	12	5	1	1		✓		✓		✓
Ile	56.50	14	37.33	13	20	13	32.3	15	38	17	131.17	12	1.21	13	6	1	0		✓				✓
Leu	65.51	15	37.33	13	33	16	27.3	12	37	16	131.17	12	1.21	13	6	1	0		✓		✓		
Lys	53.00	12	38.66	15	18.5	10	30.3	14	36	13	146.19	15	1.31	15	6	2	0		✓				✓
Arg	65.67	16	41.32	17	18.5	10	27.3	12	20.5	18	174.20	16	1.39	16	6	4	0				✓		✓
His	69.46	17	39.99	16	33	16	38.3	17	29	16	155.16	17	1.46	17	6	3	0			✓			
Tyr	86.74	18	55.33	18	56.5	18	50	18	59	19	181.19	18	1.77	18	9	1	0				✓	✓	
Phe	87.57	19	55.33	18	63	19	52	19	61	17	165.19	19	1.84	19	9	1	0		✓			✓	
Trp	117.36	20	68.66	20	78.5	20	74.3	20	75.5	20	204.23	20	2.39	20	11	2	0		✓	✓		✓	✓

Table B.11.: Minimal and maximal requirements of precursors and energy to synthesize one complex RuBisCO. The number of required Pi are the same for all 3003 possible combinations to assemble RuBisCO.

Encoding genes		Costs [ATP]		$h\nu$		CO_2		H_2O		Pi		NO_3		SO_4	
		min	max	min	max	min	max	min	max	min	max	min	max	min	max
RbcL	ATCG00490.1	8	8	8	8	8	8	8	8	8	8	8	8	8	8
RbcS1A	AT1G67090.1	–	–	–	–	–	–	–	–			–	–	–	–
	AT1G67090.2	8	–	8	–	8	–	8	–			8	–	–	8
RbcS1B	AT5G38430.1	–	–	–	–	–	–	–	–			–	–		–
RbcS2B	AT5G38420.1	–	–	–	–	–	–	–	–	8	8	–	–		–
	AT5G38410.1	–	–	–	–	–	–	–	–			–	–	8	–
RbcS3B	AT5G38410.2	–	–	–	–	–	–	–	–			–	–		–
	AT5G38410.3	–	8	–	8	–	8	–	8			–	8		–
Minimal costs		243,287.9		231,066.4		24,024		13,536		0		6,520		216	
Maximal costs		269,133.9		256,262.3		26,520		29,128		0		7,112		240	

References

[1] Harris, PM (1992) *The Potato Crop.* Chapman & Hall. doi: 10.1007/978-94-011-2340-2.

[2] Hoffmann, AA and Sgrò, CM (2011) Climate change and evolutionary adaptation. *Nature*, 470(7335):479–485. doi: 10.1038/nature09670.

[3] Bouton, JH, *et al.* (1986) Photosynthesis, Leaf Anatomy, and Morphology of Progeny from Hybrids between C_3 and C_3/C_4 *Panicum* Species. *Plant Physiol*, 80(2):487–492. doi: 10.1104/pp.80.2.487.

[4] Taiz, L and Zeiger, E (2010) *Plant Physiology*, chapter 7, pages 163–198. Sinauer Associates, Inc., 5 edition.

[5] Truper, HG and Fischer, U (1982) Anaerobic oxidation of sulphur compounds as electron donors for bacterial photosynthesis. *Phil Trans R Soc B*, 298(1093):529–542. doi: 10.1098/rstb.1982.0095.

[6] Griffin, BM, *et al.* (2007) Nitrite, an Electron Donor for Anoxygenic Photosynthesis. *Science*, 316(5833):1870. doi: 10.1126/science.1139478.

[7] Kennedy, DO and Wightman, EL (2011) Herbal Extracts and Phytochemicals: Plant Secondary Metabolites and the Enhancement of Human Brain Function. *Adv Nutr*, 2(1):32–50. doi: 10.3945/an.110.000117.

[8] Sauer, U (2006) Metabolic networks in motion: ^{13}C-based flux analysis. *Mol Syst Biol*, 2(1):62. doi: 10.1038/msb4100109.

[9] Taiz, L and Zeiger, E (2010) *Plant Physiology*, chapter 2, pages 35–64. Sinauer Associates, Inc., 5 edition.

[10] Yin, X, *et al.* (2004) Role of crop physiology in predicting gene-to-phenotype relationships. *Trends Plant Sci*, 9(9):426–432. doi: 10.1016/j.tplants.2004.07.007.

[11] Hammer, G, *et al.* (2006) Models for navigating biological complexity in breeding improved crop plants. *Trends Plant Sci*, 11(12):587–593. doi: 10.1016/j.tplants.2006.10.006.

[12] Weckwerth, W, *et al.* (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc Natl Acad Sci U S A*, 101(20):7809–7814. doi: 10.1073/pnas.0303415101.

[13] Gutierrez, RA, *et al.* (2008) Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene cca1. *Proc Natl Acad Sci U S A*, 105(12):4939–4944. doi: 10.1073/pnas.0800211105.

References

[14] Schwall, GP, *et al.* (2000) Production of very-high-amylose potato starch by inhibition of SBE A and B. *Nat Biotechnol*, 18(5):551–554. doi: 10.1038/75427.

[15] Slattery, CJ, *et al.* (2000) Engineering starch for increased quantity and quality. *Trends Plant Sci*, 5(7):291–298. doi: 10.1016/S1360-1385(00)01657-5.

[16] Brigelius-Flohé, R and Traber, MG (1999) Vitamin E: function and metabolism. *FASEB J*, 13(10):1145–1155.

[17] Soll, J and Schultz, G (1979) Comparison of geranylgeranyl and phytyl substituted methylquinols in the tocopherol synthesis of spinach chloroplasts. *Biochem Biophys Res Commun*, 91(3):715–720. doi: 10.1016/0006-291X(79)91939-9.

[18] Li, Y, *et al.* (2008) Current Opinions on the Functions of Tocopherol Based on the Genetic Manipulation of Tocopherol Biosynthesis in Plants. *J Integr Plant Biol*, 50 (9):1057–1069. doi: 10.1111/j.1744-7909.2008.00689.x.

[19] Almeida, J, *et al.* (2011) Genetic dissection of vitamin E biosynthesis in tomato. *J Exp Bot*, 62(11):3781–3798. doi: 10.1093/jxb/err055.

[20] Kohl, P and Noble, D (2009) Systems biology and the virtual physiological human. *Mol Syst Biol*, 5:292. doi: 10.1038/msb.2009.51.

[21] Ayala, FJ (2008) Reduction, emergence, naturalism, dualism, teleology: A précis. In Cobb, JB, ed., *Back to Darwin: A Richer Account of Evolution*, pages 76–87. William B. Eerdmans Publishing Company, 7 edition.

[22] Mishler, BD and Brandon, RN (1987) Individuality, pluralism, and the phylogenetic species concept. *Biol Philos*, 2(4):397–414. doi: 10.1007/BF00127698.

[23] Donnet, S and Samson, A (2013) A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. *Adv Drug Deliv Rev*, 65 (7):929–939. doi: 10.1016/j.addr.2013.03.005.

[24] Ditlevsen, S and Samson, A (2013) Introduction to stochastic models in biology. In Mostafa Bachar, Jerry Batzel, SD, ed., *Stochastic Biomathematical Models*, pages 3–35. Springer, Berlin Heidelberg. doi: 10.1007/978-3-642-32157-3_1.

[25] Dudukovic, M (2005) Chemical kinetics: Rate and mechanistic models. In *Reading Material for the Reaction Engineering Short Course*, chapter 4N ChE 505. Chemical Reaction Engineering Laboratory (CREL), Washington University, St. Louis, MO.

[26] Schuster, S, *et al.* (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol*, 18(3):326–332. doi: 10.1038/73786.

[27] Schilling, C, *et al.* (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway oriented perspective. *J Theor Biol*, 203:229–248. doi: 10.1006/jtbi.2000.1073.

[28] Orth, JD, *et al.* (2010) What is flux balance analysis? *Nat Biotechnol*, 28(3):245–248. doi: 10.1038/nbt.1614.

[29] Llaneras, F and Picó, J (2008) Stoichiometric modelling of cell metabolism. *J Biosci Bioeng*, 105(1):1–11. doi: 10.1263/jbb.105.1.

[30] Gowik, U and Westhoff, P (2011) The path from c_3 to c_4 photosynthesis. *Plant Physiol*, 155(1):56–63. doi: 10.1104/pp.110.165308.

[31] Taiz, L and Zeiger, E (2010) *Plant Physiology*, chapter 7 and 8, pages 163–242. Sinauer Associates, Inc., 5 edition.

[32] Mueller, LA, *et al.* (2003) AraCyc: A Biochemical Pathway Database for Arabidopsis. *Plant Physiol*, 132(2):453–460. doi: 10.1104/pp.102.017236.

[33] Rapoport, TA, *et al.* (1976) The regulatory principles of glycolysis in erythrocytes in vivo and in vitro. a minimal comprehensive model describing steady states, quasi-steady states and time-dependent processes. *Biochem J*, 154(2):449–469.

[34] Segel, LA (1988) On the validity of the steady state assumption of enzyme kinetics. *Bull Math Biol*, 50(6):579–593. doi: 10.1007/BF02460092.

[35] Schauer, M and Heinrich, R (1983) Quasi-steady-state approximation in the mathematical modeling of biochemical reaction networks. *Math Biosci*, 65(2):155–170. doi: 10.1016/0025-5564(83)90058-5.

[36] Cannon, W (1932) *The Wisdom of the Body*. WW Norton & Company Inc., New York.

[37] Kitano, H (2007) Towards a theory of biological robustness. *Mol Syst Biol*, 3:137. doi: 10.1038/msb4100179.

[38] Arnold, A and Nikoloski, Z (2014) Bottom-up metabolic reconstruction of *Arabidopsis thaliana* and its application to determining the metabolic costs of enzyme production. *Plant Physiol*, under review.

[39] Mahadevan, R, *et al.* (2002) Dynamic flux balance analysis of diauxic growth in escherichia coli. *Biophys J*, 83(3):1331–1340. doi: 10.1016/S0006-3495(02)73903-9.

[40] Holzhütter, HG (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur J Biochem*, 271(14):2905–2922. doi: 10.1111/j.1432-1033.2004.04213.x.

[41] Smallbone, K and Simeonidis, E (2009) Flux balance analysis: A geometric perspective. *J Theor Biol*, 258(2):311–315. doi: 10.1016/j.jtbi.2009.01.027.

[42] Haggart, CR, *et al.* (2011) Whole-genome metabolic network reconstruction and constraint-based modeling. *Methods Enzymol*, 500:411–433. doi: 10.1016/B978-0-12-385118-5.00021-9.

References

[43] Marashi, SA and Bockmayr, A (2011) Flux coupling analysis of metabolic networks is sensitive to missing reactions. *BioSystems*, 103(1):57–66. doi: 10.1016/j.biosystems. 2010.09.011.

[44] Larhlimi, A, *et al.* (2012) F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. *BMC Bioinformatics*, 13:57. doi: 10.1186/ 1471-2105-13-57.

[45] Burgard, AP (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res*, 14(2):301–312. doi: 10.1101/gr.1926504.

[46] Arnold, A and Nikoloski, Z (2011) A quantitative comparison of Calvin-Benson cycle models. *Trends Plant Sci*, 16(12):676–683. doi: 10.1016/j.tplants.2011.09.004.

[47] Arnold, A and Nikoloski, Z (2014) In search for an accurate model of the photosynthetic carbon metabolism. *Math Comput Simulat*, 96:171–194. doi: 10.1016/j.matcom.2012.03.011.

[48] Arnold, A and Nikoloski, Z (2013) Comprehensive classification and perspective for modelling photorespiratory metabolism. *Plant Biol*, 15(4):667–675. doi: 10.1111/j. 1438-8677.2012.00708.x.

[49] Arnold, A, *et al.* (2014) Effects of varying nitrogen sources on amino acid costs in *Arabidopsis thaliana* under different light and carbon-source conditions. *J Theor Biol*, under review.

[50] Nedbal, L, *et al.* (2009) Scaling and Integration of Kinetic Models of Photosynthesis: Towards Comprehensive E-Photosynthesis. In *Photosynthesis in silico*, volume 29 of *Advances in Photosynthesis and Respiration*, pages 17–29. Springer Netherlands. doi: 10.1007/978-1-4020-9237-4_2.

[51] Zhu, XG, *et al.* (2007) Optimizing the Distribution of Resources between Enzymes of Carbon Metabolism Can Dramatically Increase Photosynthetic Rate: A Numerical Simulation Using an Evolutionary Algorithm. *Plant Physiol*, 145:513–526. doi: 10.104/pp.107.103713.

[52] Bar-Even, A, *et al.* (2010) Design and analysis of synthetic carbon fixation pathways. *Proc Natl Acad Sci U S A*, 107(19):8889–8894. doi: 10.1073/pnas.0907176107.

[53] Farazdaghi, H (2011) The single-process biochemical reaction of Rubisco: A unified theory and model with the effects of irradiance, CO_2 and rate-limiting step on the kinetics of C_3 and C_4 photosynthesis from gas exchange. *BioSystems*, 103(2):265–284. doi: 10.1016/j.biosystems.2010.11.004.

[54] Calvin, M (1954) Chemical and photochemical reactions of thioctic acid and related disulfides. *Fed Proc*, 13(3):697–711.

[55] Taylor, TC and Andersson, I (1996) Structural transitions during activation and ligand binding in hexadecameric Rubisco inferred from the crystal structure of the activated unliganded spinach enzyme. *Nat Struct Mol Biol*, 3(1):95–101. doi: 10.1038/nsb0196-95.

[56] Laisk, A, *et al.* (2006) C_3 photosynthesis in silico. *Photosynth Res*, 90(1):45–66. doi: 10.1007/s11120-006-9109-1.

[57] Fridlyand, LE and Scheibe, R (1999) Regulation of the Calvin cycle for CO_2 fixation as an example for general control mechanisms in metabolic cycles. *BioSystems*, 51 (2):79–93. doi: 10.1016/S0303-2647(99)00017-9.

[58] Pettersson, G and Ryde-Pettersson, U (1988) A mathematical model of the Calvin photosynthesis cycle. *Eur J Biochem*, 175(3):661–672. doi: 10.1111/j.1432-1033.1988. tb14242.x.

[59] Woodrow, I and Mott, K (1993) Modeling C_3 photosynthesis: a sensitivity analysis of the photosynthetic carbon-reduction cycle. *Planta*, 191(4):421–432. doi: 10.1007/BF00195743.

[60] Poolman, MG, *et al.* (2000) Modelling Photosynthesis and its Control. *J Exp Bot*, 51:319–328. doi: 10.1093/jexbot/51.suppl_1.319.

[61] Laisk, A, *et al.* (1989) A Mathematical Model of Carbon Metabolism in Photosynthesis: Difficulties in Explaining Oscillations by Fructose 2,6-bisphosphate Regulation. *Proc R Soc Lond B Biol Sci*, 237:389–415. doi: 10.1098/rspb.1989.0057.

[62] Giersch, C (1977) A kinetic model for translocators in the chloroplast envelope as an element of computer simulation of the dark reaction of photosynthesis. *Z Naturforsch*, 32c:263–270.

[63] Heldt, HW, *et al.* (1978) Phosphate requirement for the light activation of ribulose 1,5-bisphosphate carboxylase in intact spinach chloroplasts. *FEBS Letters*, 92:234–240. doi: 10.1016/0014-5793(78)80761-3.

[64] Wirtz, W, *et al.* (1982) Light activation of Calvin cycle enzymes as measured in pea leaves. *FEBS Letters*, 142:223–226. doi: 10.1016/0014-5793(82)80139-7.

[65] Stitt, M, *et al.* (1983) Regulation of Sucrose Synthesis by Cytoplasmic Fructose-bisphosphatase and Sucrose Phosphate Synthase during Photosynthesis in Varying Light and Carbon Dioxide. *Plant Physiol*, 72(3):767–774. doi: 10.1104/pp.72.3.767.

[66] Badger, MR and Andrews, TJ (1974) Effects of CO_2, O_2 and temperature on a high-affinity form of ribulose diphosphate carboxylase-oxygenase from spinach. *Biochem Biophys Res Commun*, 60(1):204–210. doi: 10.1016/0006-291X(74)90192-2.

[67] Preiss, I, *et al.* (1967) Regulation of the biosynthesis of starch in spinach leaf chloroplasts. In Goodwin, TW, ed., *Biochemistry of Chloroplasts*, volume 2, chapter 13, pages 1–53. New York: Academic.

[68] Preiss, J and Kosuge, T (1970) Regulation of enzyme activity in photosynthetic system. *Annu Rev Plant Physiol*, 21:433–466. doi: 10.1146/annurev.pp.21.060170.002245.

[69] Heldt, HW, *et al.* (1977) Role of orthophosphate and other factors in the regulation of starch formation in leaves and isolated chloroplasts. *Plant Physiol*, 59:1146–1155. doi: 10.1104/pp.59.6.1146.

References

[70] Preiss, J (1982) Regulation of the biosynthesis and degradation of starch in higher plants. *Annu Rev Plant Physiol*, 33:431–454. doi: 10.1146/annurev.pp.33.060182. 002243.

[71] Gardemann, A, *et al.* (1986) Control of CO_2 fixation regulation of stromal fructose-1,6-bisphosphatase in spinach by pH and Mg^{2+} concentration. *Planta*, 168:536–545. doi: 10.1007/BF00392274. 10.1007/BF00392274.

[72] Naula, C, *et al.* (2008) A new erythrose 4-phosphate dehydrogenase coupled assay for transketolase. *J Biochem Biophys Methods*, 70(6):1185–1187. doi: 10.1016/j.jprot. 2007.11.002.

[73] Farquhar, GD, *et al.* (1980) A Biochemical Model of photosynthetic CO_2 Assimilation in Leaves of C_3 Species. *Planta*, 149(1):78–90. doi: 10.1007/BF00386231.

[74] Medlyn, BE, *et al.* (2002) Temperature response of parameters of a biochemically based model of photosynthesis. II. a review of experimental data. *Plant Cell Environ*, 25(9):1167–1179. doi: 10.1046/j.1365-3040.2002.00891.x.

[75] Schultz, HR (2003) Extension of a Farquhar model for limitations of leaf photosynthesis induced by light environment, phenology and leaf age in grapevines (*Vitis vinifera* l. cvv. White Riesling and Zinfandel). *Funct Plant Biol*, 30(6):673–687. doi: 10.1071/FP02146.

[76] Sharkey, TD, *et al.* (2007) Fitting photosynthetic carbon dioxide response curves for C_3 leaves. *Plant Cell Environ*, 30(9):1035–1040. doi: 10.1111/j.1365-3040.2007.01710. x.

[77] Damour, G and Urban, L (2007) Models of photosynthesis need to and can be upgraded to include the effects of drought, phenological changes, sink activity and carbohydrate accumulation on the light exposure/photosynthetic capacity relationship. In *Proceedings of the 5th International Workshop on Functional-Structural Plant Models*, pages 15–1–15–5.

[78] Zhu, XG, *et al.* (2009) A simple model of the Calvin cycle has only one physiologically feasible steady state under the same external conditions. *Nonlinear Anal Real World Appl*, 10(3):1490–1499. doi: 10.1016/j.nonrwa.2008.01.021.

[79] Giersch, C, *et al.* (1990) Control Analysis of Photosynthetic CO_2 Fixation. *Photosynth Res*, 24:151–165. doi: 10.1007/BF00032595.

[80] Hahn, BD (1986) A Mathematical Model of the Calvin Cycle: Analysis of the Steady State. *Ann Bot*, 57:639–653.

[81] Grimbs, S, *et al.* (2011) Spatiotemporal dynamics of the Calvin cycle: Multistationarity and symmetry breaking instabilities. *BioSystems*, 103:212–223. doi: 10.1016/j.biosystems.2010.10.015.

[82] Woodrow, IE and Berry, JA (1988) Enzymatic Regulation of Photosynthetic co^2 Fixation in C_3 Plants. *Annu Rev Plant Physiol Plant Mol Biol*, 39:533–594. doi: 10.1146/annurev.pp.39.060188.002533.

[83] Poolman, M (1999) *Computer Modelling Applied to the Calvin Cycle.* PhD thesis, Oxford Brookes University.

[84] Laisk, A and Edwards, GE (2000) A mathematical model of c_4 photosynthesis: The mechanism of concentrating co_2 in NADP-malic enzyme type species. *Photosynth Res*, 66(3):199–224. doi: 10.1023/A:1010695402963.

[85] Fridlyand, L, *et al.* (1999) Homeostatic regulation upon changes of enzyme activities in the Calvin cycle as an example for general mechanisms of flux control. What can we expect from transgenic plants? *Photosynth Res*, 61:227–239. doi: 10.1023/A: 1006342812049. 10.1023/A:1006342812049.

[86] Leegood, RC (1990) Enzymes of the Calvin cycle. In J.Lea, P, ed., *Methods in Plant Biochemistry*, volume 3, pages 15–38. Academic Press Ltd, London.

[87] Anderson, LE (1971) Chloroplast and cytoplasmic enzymes III. pea leaf ribose 5-phosphate isomerases. *Biochim Biophys Acta - Enzymology*, 235(1):245–249. doi: 10.1016/0005-2744(71)90052-0.

[88] Arrivault, S, *et al.* (2009) Use of reverse-phase liquid chromatography, linked to tandem mass spectrometry, to profile the Calvin cycle and other metabolic intermediates in Arabidopsis rosettes at different carbon dioxide concentrations. *Plant J*, 59(5): 826–839. doi: 10.1111/j.1365-313X.2009.03902.x.

[89] Gerhardt, R, *et al.* (1987) Subcellular Metabolite Levels in Spinach Leaves: Regulation of Sucrose Synthesis during Diurnal Alterations in Photosynthetic Partitioning. *Plant Physiol*, 83(2):399–407. doi: 10.1104/pp.83.2.399.

[90] Stitt, M, *et al.* (1982) Adenine Nucleotide Levels in the Cytosol, Chloroplasts, and Mitochondria of Wheat Leaf Protoplasts. *Plant Physiol*, 70(4):971–977. doi: 10.1104/pp.70.4.971.

[91] Winter, H, *et al.* (1994) Subcellular volumes and metabolite concentrations in spinach leaves. *Planta*, 193(4):530–535. doi: 10.1007/BF02411558.

[92] Bassham, J and Krause, G (1969) Free energy changes and metabolic regulation in steady-state photosynthetic carbon reduction. *Biochim Biophys Acta*, 189(2):207–221. doi: 10.1016/0005-2728(69)90048-6.

[93] Heinrich, R and Schuster, S (1996) *The Regulation of Cellular Systems*, chapter 5, pages 138–291. Chapman & Hall, London, England. doi: 10.1007/978-1-4613-1161-4_ 5.

[94] Abdi, H (2007) Kendall rank correlation. In Salkind, N, ed., *Encyclopedia of Measurement and Statistics.* Thousand Oaks (CA): Sage Publications.

References

[95] Vemuri, GN and Aristidou, AA (2005) Metabolic Engineering in the -omics Era: Elucidating and Modulating Regulatory Networks. *Microbiol Mol Biol Rev*, 69(2): 197–216. doi: 10.1128/MMBR.69.2.197-216.2005.

[96] Penny, W, *et al.* (2006) Variational bayes. In K., F, *et al.*, eds., *Statistical Parametric Mapping: The analysis of functional brain images*, pages 303–312. Elsevier.

[97] Gibon, Y, *et al.* (2009) Adjustment of growth, starch turnover, protein content and central metabolism to a decrease of the carbon supply when Arabidopsis is grown in very short photoperiods. *Plant Cell Environ*, 32(7):859–874. doi: 10.1111/j.1365-3040. 2009.01965.x.

[98] Liu, B, *et al.* (2009) Probabilistic approximations of signaling pathway dynamics. In *Computational Methods in Systems Biology, 7th International Conference, CMSB 2009, Bologna, Italy, August 31-September 1, 2009. Proceedings*, pages 251–265. Springer. doi: 10.1007/978-3-642-03845-7_17.

[99] Craig, CL and Weber, RS (1998) Selection costs of amino acid substitutions in ColE1 and ColIa gene clusters harbored by *Escherichia coli*. *Mol Biol Evol*, 15(6):774–776. doi: 10.1093/oxfordjournals.molbev.a025981.

[100] Penning de Vries, FW, *et al.* (1974) Products, requirements and efficiency of biosynthesis: a quantitative approach. *J Theor Biol*, 45(2):339–377. doi: 10.1016/ 0022-5193(74)90119-2.

[101] Sajitz-Hermstein, M and Nikoloski, Z (2010) A novel approach for determining environment-specific protein costs: the case of *Arabidopsis thaliana*. *Bioinformatics*, 26(18):i582–i588. doi: 10.1093/bioinformatics/btq390.

[102] Barton, MD, *et al.* (2010) Evolutionary systems biology of amino acid biosynthetic cost in yeast. *PLoS One*, 5(8):e11935. doi: 10.1371/journal.pone.0011935.

[103] Kaleta, C, *et al.* (2013) Metabolic costs of amino acid and protein production in *Escherichia coli*. *Biotechnol J*, 8(9):1105–1114. doi: 10.1002/biot.201200267.

[104] Poolman, MG, *et al.* (2009) A Genome-Scale Metabolic Model of Arabidopsis and Some of Its Properties. *Plant Physiol*, 151:1570–1581. doi: 10.1104/pp.109.141267.

[105] Jackson, RJ, *et al.* (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*, 11(2):113–127. doi: 10.1038/ nrm2838.

[106] Zerihun, A, *et al.* (1998) Photosynthate costs associated with the utilization of different nitrogen-forms: influence on the carbon balance of plants and shoot-root biomass partitioning. *New Phytologist*, 138(1):1–11. doi: 10.1046/j.1469-8137.1998.00893.x.

[107] de Visser, R, *et al.* (1992) Energy costs of protein turnover: Theoretical calculation and experimental estimation from regression of respiration on protein concentration of fullgrown leaves. In Lambers H, VdPL, ed., *Molecular, Biochemical & Physiological Aspects of Plant Respiration.*, pages 493–508. SPB Academic Publishing, Den Haag.

[108] Noguchi, K, *et al.* (2001) Costs of protein turnover and carbohydrate export in leaves of sun and shade species. *Funct Plant Biol*, 28(1):37–47. doi: 10.1071/PP00057.

[109] Mastrobuoni, G, *et al.* (2012) Proteome dynamics and early salt stress response of the photosynthetic organism *Chlamydomonas reinhardtii*. *BMC Genomics*, 13:215. doi: 10.1186/1471-2164-13-215.

[110] Tcherkez, GGB, *et al.* (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci U S A*, 103(19):7246–7251. doi: 10.1073/pnas.0600605103.

[111] Kumar, A and Malhotra, OP (1990) Steady-state kinetic properties of phosphoglycerate kinase of mung beans. *Indian J Biochem Biophys*, 27(5):311–315.

[112] Sparla, F, *et al.* (2004) Coenzyme Site-directed Mutants of Photosynthetic A4-GAPDH Show Selectively Reduced NADPH-dependent Catalysis, Similar to Regulatory AB-GAPDH Inhibited by Oxidized Thioredoxin. *J Mol Biol*, 340(5):1025–1037. doi: 10.1016/j.jmb.2004.06.005.

[113] Graciet, E, *et al.* (2003) Characterization of native and recombinant A4 glyceraldehyde 3-phosphate dehydrogenase. kinetic evidence for confromation changes upon association with the small protein CP12. *Eur J Biochem*, 270(1):129–136. doi: 10.1046/j.1432-1033.2003.03372.x.

[114] Graciet, E, *et al.* (2004) Involvement of two positively charged residues of *Chlamydomonas reinhardtii* glyceraldehyde-3-phosphate dehydrogenase in the assembly process of a bi-enzyme complex involved in CO_2 assimilation. *Eur J Biochem*, 271(23-24): 4737–4744. doi: 10.1111/j.1432-1033.2004.04437.x.

[115] González-Mondragón, E, *et al.* (2004) Conserved cysteine 126 in triosephosphate isomerase is required not for enzymatic activity but for proper folding and stability. *Biochemistry*, 43(11):3255–3263. doi: 10.1021/bi036077s.

[116] Knobeloch, D, *et al.* (2010) A coleopteran triosephosphate isomerase: X-ray structure and phylogenetic impact of insect sequences. *Insect Mol Biol*, 19(1):35–48. doi: 10.1111/j.1365-2583.2009.00928.x.

[117] Singh, SK, *et al.* (2001) Synthetic peptides as inactivators of multimeric enzymes: inhibition of *Plasmodium falciparum* triosephosphate isomerase by interface peptides. *FEBS Lett*, 501(1):19–23. doi: 10.1016/S0014-5793(01)02606-0.

[118] Banerjee, M, *et al.* (2011) Engineered dimer interface mutants of triosephosphate isomerase: the role of inter-subunit interactions in enzyme function and stability. *Protein Eng Des Sel*, 24(5):463–472. doi: 10.1093/protein/gzr005.

[119] Lal, A, *et al.* (2005) Purification and characterization of an allosteric fructose-1,6-bisphosphate aldolase from germinating mung beans (*Vigna radiata*). *Phytochemistry*, 66(9):968–974. doi: 10.1016/j.phytochem.2005.03.009.

References

[120] Moorhead, G and Plaxton, W (1990) Purification and characterisation of cytosolic aldolase from carrot storage root. *Biochem J*, 269(1):133–139.

[121] Cadet, F, *et al.* (1987) Effects of pH and fructose 2,6-bisphosphate on oxidized and reduced spinach chloroplastic fructose-1,6-bisphosphatase. *Eur J Biochem*, 162(2): 393–398. doi: 10.1111/j.1432-1033.1987.tb10614.x.

[122] Tang, GL, *et al.* (2000) Overexpression in *Escherichia coli* and characterization of the chloroplast fructose-1,6-bisphosphatase from wheat. *Protein Expr Purif*, 19(3): 411–418. doi: 10.1006/prep.2000.1267.

[123] Teige, M, *et al.* (1998) Purification, properties and in situ localization of the amphibolic enzymes D-ribulose 5-phosphate 3-epimerase and transketolase from spinach chloroplasts. *Eur J Biochem*, 252:237–244. doi: 10.1046/j.1432-1327.1998.2520237.x.

[124] Cadet, F and Meunier, JC (1988) pH and kinetic studies of chloroplast sedoheptulose-1,7-bisphosphatase from spinach (*Spinacia oleracea*). *Biochem J*, 253(1):249–254.

[125] Chen, YR, *et al.* (1999) Identification of a catalytic aspartyl residue of D-ribulose 5-phosphate 3-epimerase by site-directed mutagenesis. *J Biol Chem*, 274(4):2132–2136. doi: 10.1074/jbc.274.4.2132.

[126] Jung, C, *et al.* (2000) D-ribose-5-phosphate isomerase from spinach: Heterologous overexpression, purification, characterization, and site-directed mutagenesis of the recombinant enzyme. *Arch Biochem Biophys*, 373(2):409–417. doi: 10.1006/abbi. 1999.1554.

[127] Avilan, L, *et al.* (1997) Information transfer in multienzyme complexes–2. the role of Arg64 of *Chlamydomonas reinhardtii* phosphoribulokinase in the information transfer between glyceraldehyde-3-phosphate dehydrogenase and phosphoribulokinase. *Eur J Biochem*, 250(2):296–302. doi: 10.1111/j.1432-1033.1997.0296a.x.

[128] Lin, HY, *et al.* (2009) Effects of inherited mutations on catalytic activity and structural stability of human glucose-6-phosphate isomerase expressed in *Escherichia coli*. *Biochim Biophys Acta - Proteins and Proteomics*, 1794(2):315–323. doi: 10. 1016/j.bbapap.2008.11.004.

[129] Sun, LC, *et al.* (2009) Glucose-6-phosphate isomerase is an endogenous inhibitor to myofibril-bound serine proteinase of crucian carp (*Carassius auratus*). *J Agric Food Chem*, 57(12):5549–5555. doi: 10.1021/jf9004669.

[130] Ye, RW, *et al.* (1994) Purification and characterization of phosphomannomutase/phosphoglucomutase from pseudomonas aeruginosa involved in biosynthesis of both alginate and lipopolysaccharide. *J Bacteriol*, 176(16):4851–4857.

[131] Hwang, SK, *et al.* (2006) ATP binding site in the plant ADP-glucose pyrophosphorylase large subunit. *FEBS Lett*, 580(28-29):6741–6748. doi: 10.1016/j.febslet.2006.11.029.

[132] Georgelis, N and Hannah, LC (2008) Isolation of a heat-stable maize endosperm ADP-glucose pyrophosphorylase variant. *Plant Sci*, 175(3):247–254. doi: 10.1016/j.plantsci.2008.04.005.

[133] Senoura, T, *et al.* (2007) Enzymatic characterization of starch synthase III from kidney bean (*Phaseolus vulgaris* l.). *FEBS J*, 274(17):4550–4560. doi: 10.1111/j.1742-4658.2007.05984.x.

[134] Yanase, M, *et al.* (2005) Cumulative effect of amino acid replacements results in enhanced thermostability of potato type L alpha-glucan phosphorylase. *Appl Environ Microbiol*, 71(9):5433–5439. doi: 10.1128/AEM.71.9.5433-5439.2005.

[135] Hondo, T, *et al.* (1983) The Purification and Some Properties of the UDP-Glucose Pyrophosphorylase from Pollen of *Typha latifolia* Linné. *Plant Cell Physiol*, 24(1):61–69.

[136] Yoon, M, *et al.* (2009) Cloning, Expression, and Characterization of UDP-glucose Pyrophosphorylase from *Sphingomonas chungbukensis* DJ77. *Bull Korean Chem Soc*, 30(6):1360–1364.

[137] Iino, R, *et al.* (2009) Mechanism of inhibition by C-terminal alpha-helices of the epsilon subunit of *Escherichia coli* FoF1-ATP synthase. *J Biol Chem*, 284(26):17457–17464. doi: 10.1074/jbc.M109.003798.

[138] Ceccarelli, EA, *et al.* (2004) Functional plasticity and catalytic efficiency in plant and bacterial ferredoxin-NADP(H) reductases. *Biochim Biophys Acta*, 1698(2):155–165. doi: 10.1016/j.bbapap.2003.12.005.

[139] Nogués, I, *et al.* (2004) Role of the C-terminal tyrosine of ferredoxin-nicotinamide adenine dinucleotide phosphate reductase in the electron transfer processes with its protein partners ferredoxin and flavodoxin. *Biochemistry*, 43(20):6127–6137. doi: 10.1021/bi049858h.

[140] Johansson, M, *et al.* (2004) Structure and mutational analysis of a plant mitochondrial nucleoside diphosphate kinase. identification of residues involved in serine phosphorylation and oligomerization. *Plant Physiol*, 136(2):3034–3042. doi: 10.1104/pp.104.044040.

[141] Johansson, M, *et al.* (2008) The activities of nucleoside diphosphate kinase and adenylate kinase are influenced by their interaction. *Plant Sci*, 174(2):192–199. doi: http://dx.doi.org/10.1016/j.plantsci.2007.11.005.

[142] Spreitzer, RJ and Salvucci, ME (2002) Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu Rev Plant Biol*, 53:449–475. doi: 10.1146/annurev.arplant.53.100301.135233.

[143] Köpke-Secundo, E, *et al.* (1990) Isolation and Characterization of the Cytosolic and Chloroplastic 3-Phosphoglycerate Kinase from Spinach Leaves. *Plant Physiol*, 93(1):40–47. doi: 10.1104/pp.93.1.40.

References

[144] Zaffagnini, M, *et al.* (2013) High-Resolution Crystal Structure and Redox Properties of Chloroplastic Triosephosphate Isomerase from *Chlamydomonas reinhardtii. Mol Plant*, 7(1):101–120. doi: 10.1093/mp/sst139.

[145] Krüger, I and Schnarrenberger, C (1983) Purification, subunit structure and immunological comparison of fructose-bisphosphate aldolases from spinach and corn leaves. *Eur J Biochem*, 136(1):101–106. doi: 10.1111/j.1432-1033.1983.tb07711.x.

[146] Kelly, GJ, *et al.* (1982) Fructose-bisphosphatase from spinach leaf chloroplast and cytoplasm. *Methods Enzymol*, 90 Pt E:371–378.

[147] Cadet, F, *et al.* (1987) Isolation and purification of chloroplastic spinach (*Spinacia oleracea*) sedoheptulose-1,7-bisphosphatase. *Biochem J*, 241(1):71–74.

[148] Marri, L, *et al.* (2005) Reconstitution and properties of the recombinant glyceraldehyde-3-phosphate dehydrogenase/CP12/phosphoribulokinase supramolecular complex of Arabidopsis. *Plant Physiol*, 139(3):1433–1443. doi: 10.1104/pp.105. 068445.

[149] Weeden, NF and Gottlieb, LD (1982) Dissociation, Reassociation, and Purification of Plastid and Cytosolic Phosphoglucose Isomerase Isozymes. *Plant Physiol.*, 69(3): 717–723. doi: 10.1104/pp.69.3.717.

[150] Takamiya, S and Fukui, T (1978) Purification and multiple forms of phosphoglucomutase from potato tubers. *Plant Cell Physiol*, 19:759–768.

[151] Li, L and Preiss, J (1992) Characterization of ADPglucose pyrophosphorylase from a starch-deficient mutant of *Arabidopsis thaliana* (l.). *Carbohydr Res*, 227:227–239. doi: 10.1016/0008-6215(92)85074-A.

[152] Baba, T, *et al.* (1990) Properties of primer-dependent starch synthesis catalysed by starch synthase from potato tubers. *Phytochemistry*, 29(3):719–723. doi: 10.1016/ 0031-9422(90)80007-4.

[153] Steup, M, *et al.* (1980) Purification of a non-chloroplastic alpha-glucan phosphorylase from spinach leaves. *Planta*, 148:168–173.

[154] Blair, A, *et al.* (1996) Phylogenetic analyses of the homologous transmembrane channel-forming proteins of the F0F1-ATPases of bacteria, chloroplasts and mitochondria. *Microbiology*, 142 (Pt 1):17–32. doi: 10.1099/13500872-142-1-17.

[155] Lintala, M, *et al.* (2007) Structural and functional characterization of ferredoxin-NADP+-oxidoreductase using knock-out mutants of Arabidopsis. *Plant J*, 49(6): 1041–1052. doi: 10.1111/j.1365-313X.2006.03014.x.

[156] Meng, M, *et al.* (2008) Molecular and kinetic characterization of two UDP-glucose pyrophosphorylases, products of distinct genes, from Arabidopsis. *Biochim Biophys Acta*, 1784(6):967–972. doi: 10.1016/j.bbapap.2008.02.021.

[157] Sowokinos, JR, *et al.* (1993) Pyrophosphorylases in *Solanum tuberosum* (IV. Purification, Tissue Localization, and Physicochemical Properties of UDP-Glucose Pyrophosphorylase). *Plant Physiol*, 101(3):1073–1080. doi: 10.1104/pp.101.3.1073.

[158] Harbron, S, *et al.* (1981) The purification and properties of sucrose-phosphate synthetase from spinach leaves: the involvement of this enzyme and fructose bisphosphatase in the regulation of sucrose biosynthesis. *Arch Biochem Biophys*, 212(1):237–246. doi: 10.1016/0003-9861(81)90363-5.

[159] Whitaker, DP (1984) Purification and properties of sucrose-6-phosphatase from *Pisum sativum* shoots. *Phytochemistry*, 23(11):2429–2430. doi: 10.1016/S0031-9422(00) 84069-8.

[160] Nomura, T, *et al.* (1991) Purification and characterization of nucleoside diphosphate kinase from spinach leaves. *Biochim Biophys Acta*, 1077(1):47–55. doi: 10.1016/ 0167-4838(91)90524-4.

[161] Foyer, CH, *et al.* (2009) Photorespiratory metabolism: Genes, mutants, energetics, and redox signaling. *Annu Rev Plant Biol*, 60(1):455–484. doi: 10.1146/annurev. arplant.043008.091948.

[162] Bauwe, H, *et al.* (2012) Photorespiration has a dual origin and manifold links to central metabolism. *Curr Opin Plant Biol*, 15(3):269–275. doi: 10.1016/j.pbi.2012.01.008.

[163] Leegood, RC, *et al.* (1995) The regulation and control of photorespiration. *J Exp Bot*, 46(special):1397–1414. doi: 10.1093/jxb/46.special_issue.1397.

[164] Bauwe, H, *et al.* (2010) Photorespiration: players, partners and origin. *Trends Plant Sci*, 15(6):330–336. doi: 10.1016/j.tplants.2010.03.006.

[165] Timm, S, *et al.* (2008) A cytosolic pathway for the conversion of hydroxypyruvate to glycerate during photorespiration in arabidopsis. *Plant Cell*, 20(10):2848–2859. doi: 10.1105/tpc.108.062265.

[166] Ogren, WL and Bowes, G (1971) Ribulose diphosphate carboxylase regulates soybean photorespiration. *Nat New Biol*, 230(13):159–160. doi: 10.1038/newbio230159a0.

[167] Neill, SJ, *et al.* (2002) Hydrogen peroxide and nitric oxide as signalling molecules in plants. *J Exp Bot*, 53(372):1237–1247. doi: 10.1093/jexbot/53.372.1237.

[168] Zhang, P, *et al.* (2010) Creation of a genome-wide metabolic pathway database for *Populus trichocarpa* using a new approach for reconstruction and curation of metabolic pathways for plants. *Plant Physiol*, 153(4):1479–1491. doi: 10.1104/pp.110.157396.

[169] Spalding, MH (2009) The co2-concentrating mechanism and carbon assimilation. In Harris, EH, *et al.*, eds., *The Chlamydomonas Sourcebook (Second Edition)*, volume 2, chapter 8, pages 257–301. Academic Press, London. doi: 10.1016/B978-0-12-370873-1. 00016-2.

References

[170] Forrester, ML, *et al.* (1966) Effect of oxygen on photosynthesis, photorespiration and respiration in detached leaves. i. soybean. *Plant Physiol*, 41(3):422–427. doi: 10.1104/pp.41.3.422.

[171] Tregunna, EB, *et al.* (1966) Effect of oxygen on the rate of photorespiration in detached tobacco leaves. *Physiol Plant*, 19(3):723–733. doi: 10.1111/j.1399-3054.1966.tb07057.x.

[172] Laing, WA, *et al.* (1974) Regulation of soybean net photosynthetic co2 fixation by the interaction of co2, o2, and ribulose 1,5-diphosphate carboxylase. *Plant Physiol*, 54(5):678–685. doi: 10.1104/pp.54.5.678.

[173] Farquhar, GD (1979) Models describing the kinetics of ribulose biphosphate carboxylase-oxygenase. *Arch Biochem Biophys*, 193(2):456–468. doi: 10.1016/0003-9861(79)90052-3.

[174] Farazdaghi, H and Edwards, GE (1988) A mechanistic model for photosynthesis based on the multisubstrate ordered reaction of ribulose 1,5 bisphosphate carboxylase. *Plant Cell Environ*, 11(9):789–798. doi: 10.1111/j.1365-3040.1988.tb01904.x.

[175] Sharkey, TD (1985) Photosynthesis in intact leaves of c3 plants: Physics, physiology and rate limitations. *Botanical Review*, 51(1):53–105. doi: 10.1007/BF02861058.

[176] Urban, L, *et al.* (2003) A biochemical model of photosynthesis for mango leaves: evidence for the effect of fruit on photosynthetic capacity of nearby leaves. *Tree Physiol*, 23(5):289–300. doi: 10.1093/treephys/23.5.289.

[177] Gross, LJ, *et al.* (1991) A dynamic model of photosynthesis in varying light taking account of stomatal conductance, c3-cycle intermediates, photorespiration and rubisco activation. *Plant Cell Environ*, 14(9):881–893. doi: 10.1111/j.1365-3040.1991.tb00957.x.

[178] Kirschbaum, MUF, *et al.* (1988) Observed and modelled stomatal responses to dynamic light environments in the shade plant alocasia macrorrhiza. *Plant Cell Environ*, 11(2):111–121. doi: 10.1111/1365-3040.ep11604898.

[179] Lawlor, DW and Pearlman, JG (1981) Compartmental modelling of photorespiration and carbon metabolism of water stressed leaves. *Plant Cell Environ*, 4(1):37–52. doi: 10.1111/j.1365-3040.1981.tb00833.x.

[180] Hahn, BD (1987) A mathematical model of photorespiration and photosynthesis. *Ann Bot*, 60:157–169.

[181] Hahn, B (1991) Photosynthesis and photorespiration: modelling the essentials. *J Theor Biol*, 151(1):123–139. doi: 10.1016/S0022-5193(05)80147-X.

[182] Young, JD, *et al.* (2011) Mapping photoautotrophic metabolism with isotopically nonstationary ^{13}c flux analysis. *Metab Eng*, 13(6):656–665. doi: 10.1016/j.ymben.2011.08.002.

154

[183] Hahn, BD (1984) A mathematical model of leaf carbon metabolism. *Ann Bot*, 54: 325–339.

[184] Tolbert, NE (1980) Photorespiration. In Stumpf, P and Conn, E, eds., *The Biochemistry of Plants: Metabolism and respiration*, volume 2, pages 487–523. Academic Press.

[185] Jablonsky, J, *et al.* (2011) Modeling the Calvin-Benson cycle. *BMC Syst Biol*, 5:185. doi: 10.1186/1752-0509-5-185.

[186] de Oliveira Dal'Molin, CG, *et al.* (2010) AraGEM, a Genome-Scale Reconstruction of the Primary Metabolic Network in Arabidopsis. *Plant Physiol*, 152(2):579–589. doi: 10.1104/pp.109.148817.

[187] Saha, R, *et al.* (2011) Zea mays irs1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism. *PLoS One*, 6(7):e21784. doi: 10.1371/journal. pone.0021784.

[188] Mintz-Oron, S, *et al.* (2012) Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc Natl Acad Sci U S A*, 109(1):339–344. doi: 10.1073/pnas.1100358109.

[189] Boyle, N and Morgan, J (2009) Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC Syst Biol*, 3(1):4. doi: 10.1186/1752-0509-3-4.

[190] de Oliveira Dal'Molin, CG, *et al.* (2011) AlgaGEM–a genome-scale metabolic reconstruction of algae based on the *Chlamydomonas reinhardtii* genome. *BMC Genomics*, 12 Suppl 4:S5. doi: 10.1186/1471-2164-12-S4-S5.

[191] Chang, RL, *et al.* (2011) Metabolic network reconstruction of chlamydomonas offers insight into light-driven algal metabolism. *Mol Syst Biol*, 7:518. doi: 10.1038/msb. 2011.52.

[192] Nogales, J, *et al.* (2012) Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proc Natl Acad Sci U S A*, 109(7):2678–2683. doi: 10.1073/pnas.1117907109.

[193] Knoop, H, *et al.* (2010) The metabolic network of synechocystis sp. pcc 6803: systemic properties of autotrophic growth. *Plant Physiol*, 154(1):410–422. doi: 10.1104/pp.110.157198.

[194] Vu, TT, *et al.* (2012) Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium cyanothece sp. atcc 51142. *PLoS Comput Biol*, 8(4):e1002460. doi: 10.1371/journal.pcbi.1002460.

[195] de Oliveira Dal'Molin, CG, *et al.* (2010) C4GEM, a genome-scale metabolic model to study c_4 plant metabolism. *Plant Physiol*, 154(4):1871–1885. doi: 10.1104/pp.110. 166488.

References

[196] Noctor, G, *et al.* (2002) Drought and oxidative load in the leaves of c₃ plants: a predominant role for photorespiration? *Ann Bot*, 89:841–850. doi: 10.1093/aob/mcf096.

[197] Cheeseman, J (2007) Hydrogen peroxide and plant stress: a challenging relationship. *Plant Stress*, 1:4–15.

[198] Covert, MW, *et al.* (2001) Regulation of gene expression in flux balance models of metabolism. *J Theor Biol*, 213(1):73–88. doi: 10.1006/jtbi.2001.2405.

[199] Covert, MW, *et al.* (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in escherichia coli. *Bioinformatics*, 24(18):2044–2050. doi: 10.1093/bioinformatics/btn352.

[200] Feng, X, *et al.* (2012) Integrating flux balance analysis into kinetic models to decipher the dynamic metabolism of shewanella oneidensis mr-1. *PLoS Comput Biol*, 8(2): e1002376. doi: 10.1371/journal.pcbi.1002376.

[201] Hanson, AD, *et al.* (2000) Plant one-carbon metabolism and its engineering. *Trends Plant Sci*, 5(5):206–213. doi: 10.1016/S1360-1385(00)01599-5.

[202] Long, SP, *et al.* (2004) Rising atmospheric carbon dioxide: plants face the future. *Annu Rev Plant Biol*, 55:591–628. doi: 10.1146/annurev.arplant.55.031903.141610.

[203] Luo, Y, *et al.* (2004) Progressive Nitrogen Limitation of Ecosystem Responses to Rising Atmospheric Carbon Dioxide. *BioScience*, 54(8):731. doi: 10.1641/0006-3568(2004)054[0731:PNLOER]2.0.CO;2.

[204] Bloom, AJ (2009) As carbon dioxide rises, food quality will decline without careful nitrogen management. *Calif Agric*, 63(2):67–72. doi: 10.3733/ca.v063n02p67.

[205] Igamberdiev, AU, *et al.* (2001) The role of photorespiration in redox and energy balance of photosynthetic plant cells: A study with a barley mutant deficient in glycine decarboxylase. *Physiol Plant*, 111(4):427–438. doi: 10.1034/j.1399-3054.2001.1110402.x.

[206] Bloom, AJ, *et al.* (2002) Nitrogen assimilation and growth of wheat under elevated carbon dioxide. *Proc Natl Acad Sci U S A*, 99(3):1730–1735. doi: 10.1073/pnas.022627299.

[207] Backhausen, J. E.and Kitzmann, C, *et al.* (2000) Electron acceptors in isolated intact spinach chloroplasts act hierarchically to prevent over-reduction and competition for electrons. *Photosynth Res*, 64(1):1–13. doi: 10.1023/A:1026523809147.

[208] Baldwin, RL and Denham, SC (1979) Quantitative and dynamic aspects of nitrogen metabolism in the rumen: a modeling analysis. *J Anim Sci*, 49(6):1631–1639.

[209] van Riel, NA, *et al.* (1998) A Structured, Mminimal Parameter Model of the Central Nitrogen Metabolism in *Saccharomyces cerevisiae*: the Prediction of the Behavior of Mutants. *J Theor Biol*, 191(4):397–414. doi: 10.1006/jtbi.1997.0600.

[210] Rabouille, S, *et al.* (2006) Modeling the dynamic regulation of nitrogen fixation in the cyanobacterium *Trichodesmium* sp. *Appl Environ Microbiol*, 72(5):3217–3227. doi: 10.1128/AEM.72.5.3217-3227.2006.

[211] Resendis-Antonio, O, *et al.* (2007) Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. *PLoS Comput Biol*, 3(10):1887–1895. doi: 10.1371/journal.pcbi.0030192.

[212] Bartl, M, *et al.* (2010) Model-based Optimization to Explain Lliver Zonation in Nitrogen Metabolism. In *55th International Scientific Colloquium*, pages 235–240, Ilmenau, Germany.

[213] Sweetlove, LJ and Ratcliffe, RG (2011) Flux-balance modeling of plant metabolism. *Front Plant Sci*, 2:38. doi: 10.3389/fpls.2011.00038.

[214] Hannah, MA, *et al.* (2010) Combined transcript and metabolite profiling of Arabidopsis grown under widely variant growth conditions facilitates the identification of novel metabolite-mediated regulation of gene expression. *Plant Physiol*, 152(4): 2120–2129. doi: 10.1104/pp.109.147306.

[215] Caldana, C, *et al.* (2011) High-density kinetic analysis of the metabolomic and transcriptomic response of Arabidopsis to eight environmental conditions. *Plant J*, 67(5):869–884. doi: 10.1111/j.1365-313X.2011.04640.x.

[216] Lewis, NE, *et al.* (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nat Rev Microbiol*, 10(4):291–305. doi: 10.1038/nrmicro2737.

[217] McCloskey, D, *et al.* (2013) Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol Syst Biol*, 9:661. doi: 10.1038/msb.2013.18.

[218] Satish Kumar, V, *et al.* (2007) Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics*, 8:212. doi: 10.1186/1471-2105-8-212.

[219] Williams, TCR, *et al.* (2010) A genome-scale metabolic model accurately predicts fluxes in central carbon metabolism under stress conditions. *Plant Physiol*, 154(1): 311–323. doi: 10.1104/pp.110.158535.

[220] Cheung, CYM, *et al.* (2013) A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. *Plant J*, 75(6):1050–1061. doi: 10.1111/tpj.12252.

[221] Thiele, I and Palsson, BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, 5(1):93–121. doi: 10.1038/nprot.2009.203.

[222] Lamesch, P, *et al.* (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res*, 40(Database issue):D1202–D1210. doi: 10.1093/nar/gkr1090.

References

[223] Kanehisa, M and Goto, S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, 28(1):27–30. doi: 10.1093/nar/28.1.27.

[224] The UniProt Consortium (2012) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res*, 41(D1):D43–D47. doi: 10.1093/nar/gks1068.

[225] Thimm, O, *et al.* (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J*, 37(6): 914–939. doi: 10.1111/j.1365-313X.2004.02016.x.

[226] Sun, Q, *et al.* (2009) PPDB, the Plant Proteomics Database at Cornell. *Nucleic Acids Res*, 37(Database issue):D969–D974. doi: 10.1093/nar/gkn654.

[227] Seelert, H, *et al.* (2000) Structural biology. Proton-powered turbine of a plant motor. *Nature*, 405(6785):418–419. doi: 10.1038/35013148.

[228] Kayser, A, *et al.* (2005) Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state. *Microbiology*, 151(Pt 3):693–706. doi: 10.1099/mic.0.27481-0.

[229] Tschoep, H, *et al.* (2009) Adjustment of growth and central metabolism to a mild but sustained nitrogen-limitation in *Arabidopsis*. *Plant Cell Environ*, 32(3):300–318. doi: 10.1111/j.1365-3040.2008.01921.x.

[230] Gibon, Y, *et al.* (2004) Adjustment of diurnal starch turnover to short days: depletion of sugar during the night leads to a temporary inhibition of carbohydrate utilization, accumulation of sugars and post-translational activation of adp-glucose pyrophosphorylase in the following light period. *Plant J*, 39(6):847–862. doi: 10.1111/j.1365-313X.2004.02173.x.

[231] Stitt, M, *et al.* (2007) Multilevel genomics analysis of carbon signalling during low carbon availability: coordinating the supply and utilisation of carbon in a fluctuating environment. *Funct Plant Biol*, 34(6):526. doi: 10.1071/FP06249.

[232] Sulpice, R, *et al.* (2013) Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of Arabidopsis accessions. *Plant Physiol*, 162(1):347–363. doi: 10.1104/pp.112.210104.

[233] Wang, R (2000) Genomic Analysis of a Nutrient Response in Arabidopsis Reveals Diverse Expression Patterns and Novel Metabolic and Potential Regulatory Genes Induced by Nitrate. *Plant Cell*, 12(8):1491–1510. doi: 10.1105/tpc.12.8.1491.

[234] Varshavsky, A (1992) The N-end rule. *Cell*, 69(5):725–735. doi: 10.1101/SQB.1995.060.01.051.

[235] Bachmair, A, *et al.* (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science*, 234(4773):179–186. doi: 10.1126/science.3018930.

[236] Gonda, DK, *et al.* (1989) Universality and structure of the N-end rule. *J Biol Chem*, 264(28):16700–16712.

[237] Apel, W, *et al.* (2010) Identification of protein stability determinants in chloroplasts. *Plant J*, 63(4):636–650. doi: 10.1111/j.1365-313X.2010.04268.x.

[238] Graciet, E, *et al.* (2010) Structure and evolutionary conservation of the plant N-end rule pathway. *Plant J*, 61(5):741–751. doi: 10.1111/j.1365-313X.2009.04099.x.

[239] Schomburg, I, *et al.* (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res*, 41(Database issue):D764–D772. doi: 10.1093/nar/gks1049.

[240] Löw, R, *et al.* (1996) Early salt stress effects on the differential expression of vacuolar H(+)-ATPase genes in roots and leaves of *Mesembryanthemum crystallinum*. *Plant Physiol*, 110(1):259–265. doi: 10.1104/pp.110.1.259.

[241] Akashi, H and Gojobori, T (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci U S A*, 99(6):3695–3700. doi: 10.1073/pnas.062526999.

[242] Seligmann, H (2003) Cost-Minimization of Amino Acid Usage. *J Mol Evol*, 56(2): 151–161. doi: 10.1007/s00239-002-2388-z.

[243] Wagner, A (2005) Energy constraints on the evolution of gene expression. *Mol Biol Evol*, 22(6):1365–1374. doi: 10.1093/molbev/msi126.

[244] Ellis, R (1979) The most abundant protein in the world. *Trends Biochem Sci*, 4(11): 241–244. doi: 10.1016/0968-0004(79)90212-3.

[245] Peterhansel, C and Maurino, VG (2011) Photorespiration redesigned. *Plant Physiol*, 155(1):49–55. doi: 10.1104/pp.110.165019.

[246] Whitney, SM, *et al.* (2009) Rubisco oligomers composed of linked small and large subunits assemble in tobacco plastids and have higher affinities for CO2 and O2. *Plant Physiol*, 149(4):1887–1895. doi: 10.1104/pp.109.135210.

[247] Turina, P, *et al.* (2003) H+/ATP ratio of proton transport-coupled ATP synthesis and hydrolysis catalysed by CF0F1-liposomes. *EMBO J*, 22(3):418–426. doi: 10.1093/emboj/cdg073.

[248] Rigden, J (1996) Body, Physics of. In *Macmillan Encyclopedia of Physics: A-D*, volume 1. Simon & Schuster Macmillan, New York.

[249] Drubach, D (2000) Structure of the Brain. In *The Brain Explained*. Prentice Hall, Upper Saddle River.

[250] Williams, SG, *et al.* (2001) Peak exercise cardiac power output; a direct indicator of cardiac function strongly predictive of prognosis in chronic heart failure. *Eur Heart J*, 22(16):1496–1503. doi: 10.1053/euhj.2000.2547.

References

[251] Piques, M, *et al.* (2009) Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. *Mol Syst Biol*, 5:314. doi: 10.1038/msb.2009.68.

[252] Meyer, RC, *et al.* (2004) Heterosis of biomass production in arabidopsis. establishment during early development. *Plant Physiol*, 134(4):1813–1823. doi: 10.1104/pp.103. 033001.

[253] Pick, TR, *et al.* (2013) PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. *Proc Natl Acad Sci U S A*, 110(8):3185–3190. doi: 10.1073/pnas.1215142110.

[254] Williams, LE, *et al.* (2000) Sugar transporters in higher plants–a diversity of roles and complex regulation. *Trends Plant Sci*, 5(7):283–290. doi: 10.1016/S1360-1385(00) 01681-2.

[255] Geiger, D (2011) Plant sucrose transporters from a biophysical point of view. *Mol Plant*, 4(3):395–406. doi: 10.1093/mp/ssr029.

[256] Obata, T and Fernie, AR (2012) The use of metabolomics to dissect plant responses to abiotic stresses. *Cell Mol Life Sci*, 69(19):3225–3243. doi: 10.1007/s00018-012-1091-5.

[257] Williams, TCR, *et al.* (2008) Metabolic network fluxes in heterotrophic Arabidopsis cells: stability of the flux distribution under different oxygenation conditions. *Plant Physiol*, 148(2):704–718. doi: 10.1104/pp.108.125195.

[258] DeBolt, S, *et al.* (2009) Mutations in UDP-Glucose:sterol glucosyltransferase in Arabidopsis cause transparent testa phenotype and suberization defect in seeds. *Plant Physiol*, 151(1):78–87. doi: 10.1104/pp.109.140582.

[259] Mooney, BP, *et al.* (2006) Using quantitative proteomics of *Arabidopsis* roots and leaves to predict metabolic activity. *Physiol Plant*, 128(2):237–250. doi: 10.1111/j. 1399-3054.2006.00746.x.

[260] Doermann, P, *et al.* (1995) Isolation and characterization of an Arabidopsis mutant deficient in the thylakoid lipid digalactosyl diacylglycerol. *Plant Cell*, 7(11):1801–1810. doi: 10.1105/tpc.7.11.1801.

[261] Sharrock, RA and Clack, T (2002) Patterns of expression and normalized levels of the five Arabidopsis phytochromes. *Plant Physiol*, 130(1):442–456. doi: 10.1104/pp. 005389.

[262] Heinrich, R and Schuster, S (1996) *The Regulation of Cellular Systems*, chapter 5, pages 9–74. Chapman & Hall, London, England. doi: 10.1007/978-1-4613-1161-4_2.

[263] Geiger, DR and Servaites, JC (1994) Diurnal Regulation of Photosynthetic Carbon Metabolism in C_3 Plants. *Annu Rev Plant Physiol Plant Mol Biol*, 45:235–256. doi: 10.1146/annurev.pp.45.060194.001315.

[264] Heizer, Jr, EM, *et al.* (2006) Amino acid cost and codon-usage biases in 6 prokaryotic genomes: a whole-genome analysis. *Mol Biol Evol*, 23(9):1670–1680. doi: 10.1093/ molbev/msl029.

[265] Masakapalli, SK, *et al.* (2013) The metabolic flux phenotype of heterotrophic Arabidopsis cells reveals a complex response to changes in nitrogen supply. *Plant J*, 74(4): 569–582. doi: 10.1111/tpj.12142.

[266] Helali, SM, *et al.* (2010) Influence of nitrate-ammonium ratio on growth and nutrition of *Arabidopsis thaliana*. *Plant Soil*, 336(1-2):65–74. doi: 10.1007/s11104-010-0445-8.

[267] Miller, AJ and Cramer, MD (2005) Root nitrogen acquisition and assimilation. *Plant Soil*, 274(1-2):1–36. doi: 10.1007/s11104-004-0965-1.

[268] von Wirén, N, *et al.* (2000) The molecular physiology of ammonium uptake and retrieval. *Curr Opin Plant Biol*, 3(3):254–261. doi: 10.1016/S1369-5266(00)80074-6.

[269] Buchanan, BB (1980) Role of Light in the Regulation of Chloroplast Enzymes. *Annu Rev Plant Physiol*, 31:341–374. doi: 10.1146/annurev.pp.31.060180.002013.

[270] Geigenberger, P, *et al.* (2005) Redox regulation of carbon storage and partitioning in response to light and sugars. *J Exp Bot*, 56(416):1469–1479. doi: 10.1093/jxb/eri178.

[271] Kirchsteiger, K, *et al.* (2012) NADPH thioredoxin reductase C is localized in plastids of photosynthetic and nonphotosynthetic tissues and is involved in lateral root formation in Arabidopsis. *Plant Cell*, 24(4):1534–1548. doi: 10.1105/tpc.111.092304.

[272] Ruelland and Miginiac-Maslow (1999) Regulation of chloroplast enzyme activities by thioredoxins: activation or relief from inhibition? *Trends Plant Sci*, 4(4):136–141. doi: 10.1016/S1360-1385(99)01391-6.

[273] Eichelmann, H, *et al.* (2011) The rate of nitrite reduction in leaves as indicated by O_2 and CO_2 exchange during photosynthesis. *J Exp Bot*, 62(6):2205–2215. doi: 10.1093/jxb/erq428.

[274] Geiger, M, *et al.* (1998) Enhanced carbon dioxide leads to a modified diurnal rhythm of nitrate reductase activity in older plants, and a large stimulation of nitrate reductase activity and higher levels of amino acids in young tobacco plants. *Plant Cell Environ*, 21(3):253–268. doi: 10.1046/j.1365-3040.1998.00277.x.

[275] Chapin, FS, *et al.* (1987) Plant Responses to Multiple Environmental Factors: Physiological ecology provides tools for studying how interacting environmental resources control plant growth. *BioScience*, 37(1):49–57. doi: 10.2307/1310177.

[276] Nielsen, TH, *et al.* (2004) Fructose-2,6-bisphosphate: a traffic signal in plant metabolism. *Trends Plant Sci*, 9(11):556–563. doi: 10.1016/j.tplants.2004.09.004.

[277] Stitt, M (1990) Fructose-2,6-Bisphosphate as a Regulatory Molecule in Plants. *Annu Rev Plant Physiol Plant Mol Biol*, 41:153–185. doi: 10.1146/annurev.pp.41.060190.001101.

[278] Carrillo, N and Ceccarelli, EA (2003) Open questions in ferredoxin-NADP+ reductase catalytic mechanism. *Eur J Biochem*, 270(9):1900–1915. doi: 10.1046/j.1432-1033.2003.03566.x.

References

[279] Berg, JM, *et al.* (2002) *Biochemistry*. W H Freeman, New York, 5 edition.

[280] Sicher, RC, *et al.* (1987) A Comparative Analysis of Fructose 2,6-Bisphosphate Levels and Photosynthate Partitioning in the Leaves of Some Agronomically Important Crop Species. *Plant Physiol*, 83(4):768–771. doi: 10.1104/pp.83.4.768.

[281] Reed, JL (2012) Shrinking the Metabolic Solution Space Using Experimental Datasets. *PLoS Comput Biol*, 8(8):e1002662. doi: 10.1371/journal.pcbi.1002662.

[282] Smallbone, K, *et al.* (2010) Towards a genome-scale kinetic model of cellular metabolism. *BMC Syst Biol*, 4(1):6. doi: 10.1186/1752-0509-4-6.

[283] Li, P, *et al.* (2010) Systematic integration of experimental data and models in systems biology. *BMC Bioinformatics*, 11:582. doi: 10.1186/1471-2105-11-582.

[284] Stanford, NJ, *et al.* (2013) Systematic construction of kinetic models from genome-scale metabolic networks. *PLoS One*, 8(11):e79195. doi: 10.1371/journal.pone.0079195.

[285] Schallau, K and Junker, BH (2010) Simulating plant metabolic pathways with enzyme-kinetic models. *Plant Physiol*, 152(4):1763–1771. doi: 10.1104/pp.109.149237.

[286] Teusink, B, *et al.* (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem*, 267(17):5313–5329. doi: 10.1046/j.1432-1327.2000.01527.x.

[287] Bennett, BD, *et al.* (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol*, 5(8):593–599. doi: 10.1038/nchembio.186.

[288] Heijnen, JJ and Verheijen, PJT (2013) Parameter identification of in vivo kinetic models: limitations and challenges. *Biotechnol J*, 8(7):768–775. doi: 10.1002/biot.201300105.

[289] Emanuelsson, O, *et al.* (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*, 300(4):1005–1016. doi: 10.1006/jmbi.2000.3903.

[290] Small, I, *et al.* (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics*, 4(6):1581–1590. doi: 10.1002/pmic.200300776.

[291] Picault, N, *et al.* (2004) The growing family of mitochondrial carriers in Arabidopsis. *Trends Plant Sci*, 9(3):138–146. doi: 10.1016/j.tplants.2004.01.007.

[292] Palmieri, F, *et al.* (2011) Evolution, structure and function of mitochondrial carriers: a review with new insights. *Plant J*, 66(1):161–181. doi: 10.1111/j.1365-313X.2011.04516.x.

[293] Fischer, K (2011) The import and export business in plastids: transport processes across the inner envelope membrane. *Plant Physiol*, 155(4):1511–1519. doi: 10.1104/pp.110.170241.

[294] Weber, APM and Linka, N (2011) Connecting the plastid: transporters of the plastid envelope and their role in linking plastidial with cytosolic metabolism. *Annu Rev Plant Biol*, 62:53–77. doi: 10.1146/annurev-arplant-042110-103903.

[295] Lehmann, S, *et al.* (2010) Proline metabolism and transport in plant development. *Amino Acids*, 39(4):949–962. doi: 10.1007/s00726-010-0525-3.

[296] Reed, JL, *et al.* (2006) Systems approach to refining genome annotation. *Proc Natl Acad Sci U S A*, 103(46):17480–17484. doi: 10.1073/pnas.0603364103.

[297] Christian, N, *et al.* (2009) An integrative approach towards completing genome-scale metabolic networks. *Mol Biosyst*, 5(12):1889–1903. doi: 10.1039/B915913b.

[298] Henry, CS, *et al.* (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol*, 28(9):977–982. doi: 10.1038/nbt.1672.

[299] Jerby, L, *et al.* (2010) Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol Syst Biol*, 6:401. doi: 10.1038/msb.2010.56.

[300] Vitkin, E and Shlomi, T (2012) MIRAGE: a functional genomics-based approach for metabolic network model reconstruction and its application to cyanobacteria networks. *Genome Biol*, 13(11):R111. doi: 10.1186/gb-2012-13-11-r111.

[301] Hatzimanikatis, V, *et al.* (2005) Exploring the diversity of complex metabolic networks. *Bioinformatics*, 21(8):1603–1609. doi: 10.1093/bioinformatics/bti213.

[302] Palsson, B (2002) In silico biology through ömics". *Nat Biotechnol*, 20(7):649–650. doi: 10.1038/nbt0702-649.

[303] Petersen, J, *et al.* (2003) Origin, Evolution, and Metabolic Role of a Novel Glycolytic GAPDH Enzyme Recruited by Land Plant Plastids. *J Mol Evol*, 57(1):16–26. doi: 10.1007/s00239-002-2441-y.

[304] Endy, D (2005) Foundations for engineering biology. *Nature*, 438(7067):449–453. doi: 10.1038/nature04342.

[305] Winkel, BS (2004) Metabolic Channeling in Plants. *Annu Rev Plant Biol*, 55(1): 85–107. doi: 10.1146/annurev.arplant.55.031903.141714.

[306] Pérez-Bercoff, A, *et al.* (2011) Patterns of indirect protein interactions suggest a spatial organization to metabolism. *Mol Biosyst*, 7(11):3056–3064. doi: 10.1039/c1mb05168g.

[307] Marcotte, EM (1999) Detecting Protein Function and Protein-Protein Interactions from Genome Sequences. *Science*, 285(5428):751–753. doi: 10.1126/science.285.5428.751.

[308] Shlomi, T, *et al.* (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol*, 3:101. doi: 10.1038/msb4100141.

References

[309] Bernacchi, CJ, *et al.* (2001) Improved temperature response functions for models of Rubisco-limited photosynthesis. *Plant Cell Environ*, 24(2):253–259. doi: 10.1111/j.1365-3040.2001.00668.x.

[310] Urban, L and Alphonsout, L (2007) Girdling decreases photosynthetic electron fluxes and induces sustained photoprotection in mango leaves. *Tree Physiol*, 27(3):345–352. doi: 10.1093/treephys/27.3.345.

[311] Harley, PC, *et al.* (1992) Modelling photosynthesis of cotton grown in elevated CO_2. *Plant Cell Environ*, 15(3):271–282. doi: 10.1111/j.1365-3040.1992.tb00974.x.

[312] Lei, HB, *et al.* (2011) A parameter condition for ruling out multiple equilibria of the photosynthetic carbon metabolism. *Asian J Control*, 13(5):1–14. doi: 10.1002/asjc.359.

[313] Segel, IH (1993) *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*. John Wiley & Sons, 1 edition.

[314] Gardemann, A, *et al.* (1983) Control of CO_2 fixation. regulation of spinach ribulose-5-phosphate kinase by stromal metabolite levels. *Biochim Biophys Acta - Bioenergetics*, 722(1):51–60. doi: 10.1016/0005-2728(83)90156-1.

[315] Allen, J (2002) Photosynthesis of ATP-electrons, proton pumps, rotors, and poise. *Cell*, 110(3):273–276. doi: 10.1016/S0092-8674(02)00870-x.

[316] Pate, JS and Layzell, DB (1990) Energetics and biological costs of nitrogen assimilation. In Conn, PM, *et al.*, eds., *The Biochemistry of Plants: Intermediary Nitrogen Metabolism*, volume 16, chapter 1, pages 1–42. Academic Press, San Diego, 1 edition.

[317] Meharg, AA and Blatt, MR (1995) NO3- transport across the plasma membrane of *Arabidopsis thaliana* root hairs: kinetic control by pH and membrane voltage. *J Membr Biol*, 145(1):49–66. doi: 10.1007/BF00233306.

[318] Buchner, P, *et al.* (2004) Plant sulphate transporters: co-ordination of uptake, intracellular and long-distance transport. *J Exp Bot*, 55(404):1765–1773. doi: 10.1093/jxb/erh206.

[319] Raghothama, KG (1999) Phosphate acquisition. *Annu Rev Plant Physiol Plant Mol Biol*, 50:665–693. doi: 10.1146/annurev.arplant.50.1.665.

[320] Sakano, K (1990) Proton/Phosphate Stoichiometry in Uptake of Inorganic Phosphate by Cultured Cells of *Catharanthus roseus* (L.) G. Don. *Plant Physiol*, 93(2):479–483. doi: 10.1104/pp.93.2.479.

[321] Pyl, ET, *et al.* (2012) Metabolism and growth in Arabidopsis depend on the daytime temperature but are temperature-compensated against cool nights. *Plant Cell*, 24(6):2443–2469. doi: 10.1105/tpc.112.097188.

[322] Smith, AM and Zeeman, SC (2006) Quantification of starch in plant tissues. *Nat Protoc*, 1(3):1342–1345. doi: 10.1038/nprot.2006.232.

[323] Mahadevan, R and Schilling, CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng*, 5(4):264–276. doi: 10.1016/j.ymben.2003.09.002.

References

Publications and scientific contributions

Publications

Axmann IM, Dühring U, Seeliger L, **Arnold A**, Vanselow JT, Kramer A, Wilde A (2009) Biochemical evidence for a timing mechanism in prochlorococcus. *J Bacteriol* 191(17):5342–5347.

Grimbs S, **Arnold A**, Koseska A, Kurths J, Selbig J, Nikoloski Z (2011) Spatiotemporal dynamics of the Calvin cycle: multistationarity and symmetry breaking instabilities. *BioSystems* 103(2):212–223.

Arnold A, Nikoloski Z (2011) A quantitative comparison of Calvin-Benson cycle models. *Trends Plant Sci*, 16(12):676–683.

Arnold A, Nikoloski Z (2013) Comprehensive classification and perspective for modelling photorespiratory metabolism. *Plant Biol* 15(4):667–75.

Arnold A, Nikoloski Z (2014) In search for an accurate model of the photosynthetic carbon metabolism. *Math Comput Simulat*, 96:171–194.

Arnold A, Nikoloski Z (2014) Bottom-up metabolic reconstruction of *Arabidopsis thaliana* and its application to determining the metabolic costs of enzyme production. *Plant Physiol*, under review.

Arnold A[*], Sajtiz-Hermstein M[*],Nikoloski Z (2014) Effects of varying nitrogen sources on amino acid costs in *Arabidopsis thaliana* under different light and carbon-source conditions. *J Theor Biol*, under review. (* contributed equally)

Mettler T, Mühlhaus T, Hemme D, Schöttler M-A, Rupprecht J, Idoine A, Veyel D, Pal SK, Yaneva-Roder L, Winck FV, Sommer F, Vosloh D, Seiwert B, Erban A, Burgos A, Arvidsson S, Schlede S, **Arnold A**, Günther M, Krause U, Lohse M, Kopka J, Nikoloski Z, Mueller- Roeber B, Willmitzer L, Bock R, Schroda M, Stitt M (2014) Systems analysis of the response of photosynthesis, metabolism and growth to an increase in irradiance in the photosynthesis model organism *Chlamydomonas reinhardtii. Plant Cell*, submitted

Conferences

09/2010 Workshop on Differential and Integral Equations with Applications in Biology and Medicine, Karlovasi, Griechenland

   Talk Design and analysis of a reference model of sugar phosphate metabolism

02/2011 FORSYS/FORSYS Partner Statusseminar, Deutsches Krebsforschungszentrum Heidelberg, Deutschland

   Poster Comparative model-based analysis of carbon metabolism based on integration of data from *C. reinhardtii* and *A. thaliana*

06/2011 European Conference on Mathematical and Theoretical Biology, Krakau, Polen

   Talk Comparative model analysis of the Calvin-Benson cycle

08/2011 International Conference on Systems Biology, Heidelberg/Mannheim, Deutschland

   Poster A quantitative comparison of Calvin-Benson cycle models

08/2012 International Conference on Systems Biology, Toronto, Kanada

   Talk Reconstruction of central metabolism integrating enzyme turnover for *A. thaliana*