

# Metabolic modeling of energy supply balance in photosynthesis



Inaugural-Dissertation

zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

**Nima P. Saadat**

aus Düsseldorf

Düsseldorf, Dezember 2022

aus dem Institut für Quantitative und Theoretische Biologie  
der Heinrich-Heine-University, Düsseldorf

Gedruckt mit der Genehmigung der  
Mathematisch-Naturwissenschaftlichen Fakultät der  
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Oliver Ebenhöf
2. Prof. Dr. Martin Lercher

Tag der mündlichen Prüfung: 17.04.2023

## **Erklärung**

Ich versichere an Eides statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der "Grundsätze zur Sicherung guter wissenschaftlicher Praxis" an der Heinrich Heine Universität Düsseldorf erstellt worden ist. Die Dissertation habe ich in dieser oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen oder erfolgreichen Promotionsversuche unternommen.

Düsseldorf, den 20. Dezember 2022

Nima Philipp Saadat

# Acknowledgement

The time that I spent working in the Institute for Quantitative and Theoretical Biology has been full of challenging and demanding, but yet very interesting, fun and rewarding projects. Most of my experiences in scientific work are positive and fulfilling thanks to QTB and the people working there, for which I am extremely grateful.

In particular, I would like to thank my supervisor Prof. Dr. Oliver Ebenhöf for being an excellent supervisor, mentor and boss. I am deeply grateful for all the patience, scientific discussions, continued support, as well as for listening and helping me when I encountered obstacles and problems. Furthermore, I want to thank you for listening to my ideas, for discussing them and for encouraging me to pursue them. Thanks to your supervision and teachings, I genuinely enjoy research and science.

I want to thank Jun.-Prof. Anna Matuszyńska for all the help, shared projects and support over the years. We first met when I was still a Bachelor student, and who would have thought how much we worked together since then. I am very happy and grateful that most of our projects resulted in very rewarding and fun experiences. I also want to thank you for always lending an ear when I needed help and someone to talk to, and for being such a good example of a young scientist to me.

I would also like to thank Dr. Ovidiu Popa for being the best office roommate I could wish for. When I started at QTB, you have been one of the most welcoming people, and always supported me and my ideas with valuable input and discussions. I am very grateful for your openness and our great scientific exchanges. Thank you for always cheering me up and keeping me in a good mood.

Thank you to all members of the QTB Team, who all created a great working environment that kept me motivated and always allowed great exchanges and fun while working, and special thanks to all members of the photosynthesis taskforce.

Lastly, words can not express my gratitude to my parents, my sister Ayda and of course to Mona for their unrelenting support and love that gives me the strength and motivation to always give my best. I would not be where I am without you.

# Contents

<b>1</b>	<b>Thesis Abstract</b>	<b>2</b>
<b>2</b>	<b>Introduction to Photobiology</b>	<b>3</b>
2.1	The relevance of photosynthesis . . . . .	3
2.2	Light dependent reactions . . . . .	4
2.2.1	The photosynthetic electron transport chain . . . . .	4
2.2.2	Electron Flows (Linear vs. Cyclic) . . . . .	5
2.2.3	Non-photochemical quenching . . . . .	6
2.2.4	Alternative electron flows and photoprotection . . . . .	7
2.3	Light independent reactions . . . . .	9
2.3.1	The Calvin Benson Bassham Cycle . . . . .	9
2.3.2	Export and storage of carbohydrates . . . . .	12
2.3.3	Regulation of the Calvin Cycle . . . . .	14
2.3.4	Carbon fixation in C3, C4, CAM plants . . . . .	14
<b>3</b>	<b>Introduction to modeling</b>	<b>17</b>
3.1	General Introduction to metabolic modeling . . . . .	17
3.2	Kinetic modeling . . . . .	18
3.2.1	Balance Equations . . . . .	19
3.2.2	Rate equations . . . . .	21
3.2.3	Steady States and their stabilities . . . . .	25
3.2.4	Metabolic Control Analysis . . . . .	27
3.3	Constraint-based modeling . . . . .	29
3.3.1	Structural modeling . . . . .	29
3.3.2	Flux Balance Analysis . . . . .	31
<b>4</b>	<b>Aims of the thesis</b>	<b>33</b>
<b>5</b>	<b>Research topics in this thesis</b>	<b>34</b>
<b>6</b>	<b>Manuscripts</b>	<b>35</b>
6.1	Building mathematical models of biological systems with modelbase . . . . .	35
6.1.1	Authors and Details . . . . .	35
6.1.2	Contributions . . . . .	35

6.2	Network Reconstruction and modeling Made Reproducible with <code>moped</code> . . . . .	47
6.2.1	Authors and Details . . . . .	47
6.2.2	Contributions . . . . .	47
6.3	Balancing energy supply during photosynthesis - a theoretical perspective . . . . .	60
6.3.1	Authors and Details . . . . .	60
6.3.2	Contributions . . . . .	60
6.4	Computational Analysis of Alternative Photosynthetic Electron Flows Linked With Oxidative Stress . . . . .	72
6.4.1	Authors and Details . . . . .	72
6.4.2	Contributions . . . . .	72
6.5	Bioenergetics of the secondary metabolites production in photosynthetic glandular trichomes . . . . .	84
6.5.1	Authors and Details . . . . .	84
6.5.2	Contributions . . . . .	84
<b>7</b>	<b>Discussion and Outlook</b>	<b>100</b>

# 1 Thesis Abstract

Although the two main processes of photosynthesis, the light dependent and the light independent reactions, are often studied in isolation, both processes exhibit a high interdependence. It is true that the energy provided by the light reactions of photosynthesis are, amongst others, limiting for the light independent reactions which utilize it. However, the fact that the rate and efficiency of the light independent reactions are apparently having a strong effect on all components of the light dependent reactions like the linear flow of electrons, the alternative electron routes and even the NPQ mechanisms show that the links between them are far more complex than just the consumption of provided energy.

In this thesis, mathematical models are used to further understand the interdependency between the light dependent reactions and the light independent reactions of photosynthesis quantitatively. It is shown that the light dependent and light independent reactions of photosynthesis can be regarded as an molecular-economic supply-demand system. In this system, the light dependent reactions represent the supply side of the system, and the light independent reactions represent the demand side. It is shown that both sides of the supply-demand system maintain control over the efficiency and rate of photosynthesis. Our model explains how a tight regulation of supply and demand reactions leads to efficient carbon fixation, and how a standby mode is necessary for maintaining intermediates necessary for carbon fixation in longer periods of darkness.

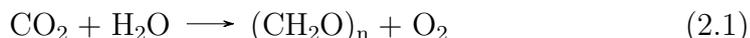
Furthermore, the effect of alternative electron flows in the electron transport chain on the rest of the photosynthetic apparatus has been investigated. It is shown that the rate of the cyclic electron flow, as well as the rate of the Calvin cycle, regulate the rate of linear electron flow and the rate of the Mehler reaction by providing  $\text{NADP}^+$  as an electron acceptor. Lastly, this thesis investigates the increased production of secondary metabolites in high light conditions in photosynthetic glandular trichomes. It is shown that higher light availability allows a shift in carbon partitioning from catabolic to anabolic pathways, as well as isoprenoid production shifting from the MEV to the MEP pathway. The results of this thesis highlight how the interdependence of light dependent and independent reactions change in different environments and reveal crucial interactions between reactions in different pathways of photosynthesis and plant metabolism.

## 2 Introduction to Photobiology

### 2.1 The relevance of photosynthesis

Photosynthetic organisms are established as the foundation of almost all ecosystems and life on earth as we know it. Photosynthetic organisms are also of utmost importance to the global human population. Especially with increasing global population, environmental challenges like climate change or shortage of fossil fuels, the interest for an increased efficiency in photosynthetic production of resources (like foods, biofuels and pharmaceuticals) has never been higher.

The word photosynthesis literally means 'building with light' and describes the process of converting light energy into chemical energy by a multitude of chemical reactions within living organisms. While this process is typically known to be found in plants, it can also be found in algae and some bacteria (e.g. Cyanobacteria). Photosynthetic organisms use light energy to drive the synthesis of sugars and other energy storing carbohydrates from carbon dioxide, usually producing oxygen as a byproduct in the process [5].



Photosynthesis in living organisms is so powerful that it changed the entire atmosphere and biosphere on earth in the past billions of years. It has been shown that there has been a time period on earth in which there was almost no oxygen in the atmosphere [35]. Approximately 2.5 billion years ago, when the first Cyanobacteria evolved, the atmospheric oxygen content slowly increased to 1-2% within the 'great oxygenation event'. The oxygen content further increased until about 850 million years ago. Today, the oxygen content in our atmosphere is approximately 21%. It is considered that most organisms at that time died out due to the sudden and massive changes in atmospheric composition, with the only survivors being species that found oxygen-free niches or adapted to the new environmental conditions [23].

Due to the immense efficiency of photosynthesis in living organisms, a non-photosynthetic organism enveloped a photosynthetic bacterium in an endosymbiotic event. This cell formed the eukaryotic ancestor to all algae and plants, with the photosynthetic bacterium developing into a cellular organelle called chloroplast [49]. The chloroplast is the main location in which the entirety of photosynthesis is taking place. Chloroplasts possess a complex organisation of endomembranes, called the thylakoids. Thylakoid

membranes are the main location for the absorption of photons and the space within thylakoid membranes is called the lumen. The space inside chloroplasts and outside of thylakoids is called the stroma. While the thylakoid membrane and lumen is the location of the light dependent part of photosynthesis, the chloroplast stroma is the location of the light independent part of photosynthesis [5, 20]. Both processes are explained in detail in the following sections. The general organisation and structure of chloroplasts and photosynthesis can be found amongst all eukaryotic photosynthetic organisms.

## 2.2 Light dependent reactions

### 2.2.1 The photosynthetic electron transport chain

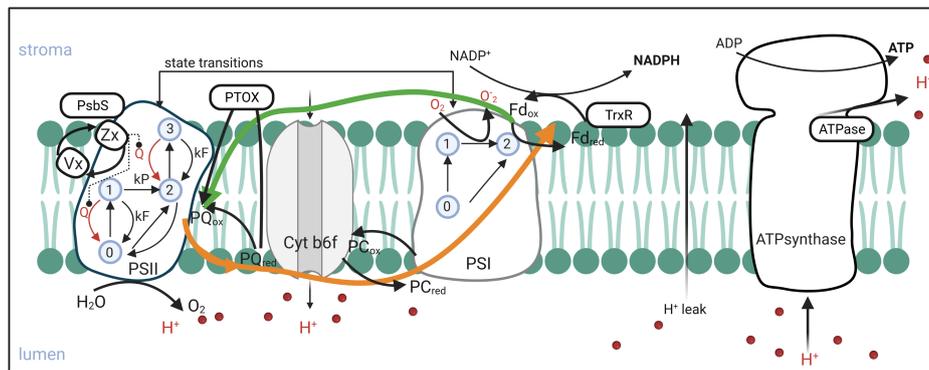


Figure 1: Schematic of the reactions of the electron transport chain (Figure taken from [55])

The light reactions, or light dependent reactions, of photosynthesis are driving the conversion of light energy to chemical energy. They take place in the thylakoid membrane and lumen of the chloroplast [2]. There, chlorophyll pigments funnel light energy to a chain of redox reactions which produces ATP and NADPH as chemical energy equivalents. This chain of redox reactions is named the photosynthetic electron chain [65]. Photosystems I and II are protein complexes which are responsible for the collection of light energy and initiate the energy conversion process in the photosynthetic electron transport chain. Photosystems consist of an antenna complex and a reaction center. The antenna complex, or light harvesting complex, is an aggregation of pigments like chlorophyll and carotenoids [32]. The role of the light harvesting complexes in the photosynthetic electron transport chain is to improve light absorption for the reaction centers of photosystems.

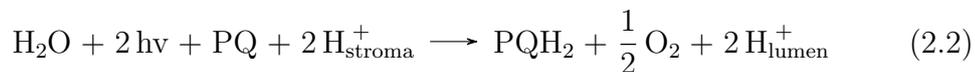
In the reaction centers, the channeled light energy is used to bring a dimeric chlorophyll molecule to a higher excitation state [46]. These dimeric chlorophyll molecules are one of the main differences between photosystems I and II. In photosystem I, the dimeric chlorophyll molecules are mainly absorbing far-red light at wavelengths of approximately 700nm (P700), while in photosystem II the main absorption is in red light at wavelengths of 680nm (P680) [54]. The dimeric chlorophyll molecules in reaction centers, which are considered primary electron donors, are rapidly transferring an electron to an adjacent electron acceptor.

### 2.2.2 Electron Flows (Linear vs. Cyclic)

The flow of electrons in the electron transport chain can take two routes: One is the linear electron flow, and the other is the cyclic electron flow. The linear electron flow is the primary route for the electron transport chain to produce both NADPH and ATP [47, 25].

Beginning at photosystem II, light energy is used to donate two electrons from P680 into the electron transport chain and transfer two electrons back to P680 from a water molecule, producing half a oxygen molecule and two protons in the lumen [3].

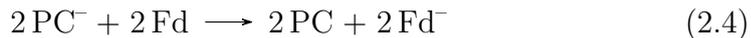
These two electrons are transferred to the electron carrier plastoquinone, which is then reduced to plastoquinol [63].



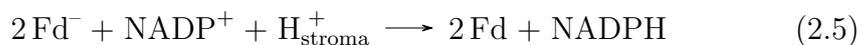
The two electrons from plastoquinol are then transferred to plastocyanin via the cytochrome B<sub>6</sub>f complex, which is pumping 2 protons from the stroma and 2 protons from plastoquinol into the lumen [61].



In photosystem I, the light energy is used to donate two electrons from P700 to ferredoxin, while the electrons from plastocyanin are transferred back to P700, which is why no electrons from water are required at photosystem I [7].



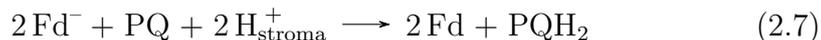
The two electrons are then transferred from ferredoxin to NADP<sup>+</sup>, reducing it to NADPH, which is one of the main reduction equivalents in plant cells [15].



The protons which have been transferred to the lumen in the process of the electron transport chain form a proton gradient, which is essential for the ATP production in photosynthesis. The proton gradient causes protons to be pumped from the lumen into the stroma through the ATP synthase [18]. This enzyme uses the proton motive force to produce ATP from ADP and orthophosphate.



An alternative flow of electrons, the cyclic electron flow, is taking place on the same components of the electron transport chain as the linear electron flow. However, in this alternative electron flow, electrons which are transferred from photosystem I to ferredoxin can be transferred back to the cytochrome B<sub>6</sub>f complex via the PGR5 protein [42].



The cycling of electrons through this route causes the pumping of protons in to the lumen without producing NADPH, and therefore increases the rate of ATP production through the ATP synthase.

### 2.2.3 Non-photochemical quenching

The absorption of light through pigments in antenna complexes in photosystems I and II is meant to funnel the respective energy to chlorophylls. This energy excites the chlorophyll molecules to a higher energy state, the first singlet excited state, in which the chlorophyll molecules are able to initiate electron transfers [44]. However, excess light conditions can cause the oversaturation of excited chlorophyll molecules, meaning that the capacity to turn the chlorophyll molecules back to their ground state via photosynthesis is full. Chlorophyll molecules that are in their singlet excited state for an extended period of time may enter its triplet state, which can provoke the evolution of reaction oxygen species (ROS) that are extremely harmful to cells [31]. A variety of mechanisms for quenching this higher energy level is present to prevent this oversaturation of excited chlorophyll molecules. Besides the necessary photosynthetic electron transfer, excited chlorophyll molecules can relax back to their ground state by emitting fluorescence and heat. Energy quenching mechanisms which are not driving photosynthesis are called non-photochemical quenching (NPQ) mechanisms [38]. Amongst such mechanisms is the energy dependent factor, called qE. Excess excitation in chlorophyll molecules leads to an increase in protons pumped into the lumen and therefore a higher proton gradient along the

thylakoid membrane. This initiates the activation of two proteins; PsbS (photosystem II subunit S) and VDE (violaxanthin de-epoxidase). VDE transforms Violaxanthin molecules to Zeaxanthin, while ZEP (zeaxanthin epoxidase) converts Zeaxanthin back to Violaxanthin. This process, called the Xanthophyll cycle, activates conformational modifications in light harvesting complexes. These conformational modifications are enhanced by PsbS activity which increases flexibility in protein complexes of photosystem II [71].

Another factor is the state transition factor, qT. This mechanism regulates excitations of antenna complex pigments in photosystem I and II by shifting light harvesting complexes from photosystem II to photosystem I via protein kinases and phosphorylases. In the first state, the light harvesting complexes of photosystem II are associated with its primary photosystem. The transition to the second state attaches them to photosystem I instead. This mechanism of non-photochemical quenching is primarily present in lower light intensities [9].

The final mechanism of non-photochemical quenching is photoinhibition, qI. In this process, photosystem II complexes and reaction centers are degraded, irreversibly inactivating the photosystems. This severe procedure to prevent photodamage is usually present in extended periods of high light intensities [71].

#### 2.2.4 Alternative electron flows and photoprotection

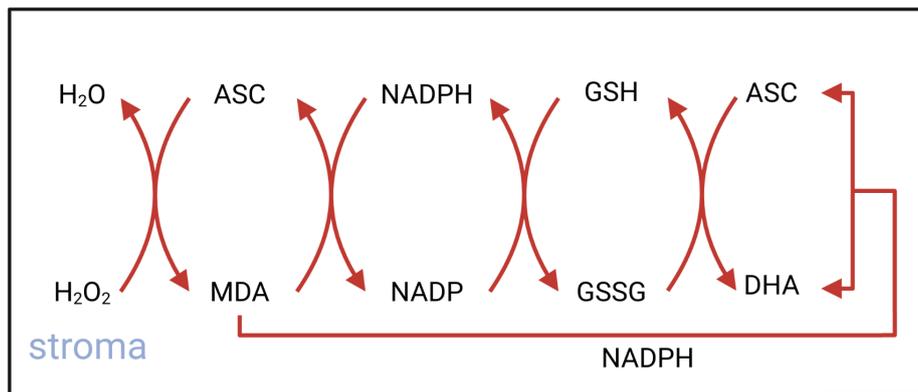
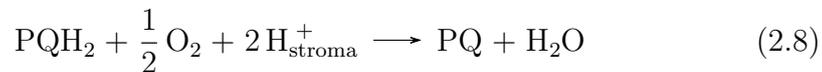


Figure 2: Schematic of the reactions of the ascorbate-glutathione cycle (Figure taken from [55])

In certain scenarios, the photosynthetic electron transport chain may be overflowed with electrons. In prolonged periods of excess light, or if

the consumption of NADPH is not fast enough and not enough  $\text{NADP}^+$  is available as a terminal electron acceptor, the overflow of electrons may lead to the production of reactive oxygen species [36]. Reactive oxygen species are extremely reactive chemicals that are, amongst others, the result of oxygen reduction. This scenario regularly happens at the photosynthetic electron transport chain, and mechanisms have evolved to mitigate the damage and production of reaction oxygen species. The photoreduction of oxygen is a consequence of an overflow of electrons on the transport chain. In order to counter this overflow, the enzyme called plastid terminal oxidase (PTOX) can act as a valve for the electron transport chain. This enzyme oxidizes plastoquinone and transfers its electrons and protons onto oxygen to produce water [43].



This reaction can ease the flow of electrons on the transport chain to prevent the formation of reactive oxygen species. Due to the fact that in the beginning of the linear electron flow, water is used as an electron donor and transformed to oxygen. These donated electrons are, if used by PTOX, spent to transform oxygen and protons to water. Therefore, this electron flow is one of two electron routes referred to as the "water-to-water cycle".

The other alternative electron flow, which can act as an rather unintended valve for electrons, is the Mehler reaction. During linear electron flow, oxygen may be photoreduced by photosystem I instead of ferredoxin. This leads to the formation of superoxide  $\text{O}_2^-$ , a reactive oxygen species. The produced superoxide is rapidly transformed to  $\text{H}_2\text{O}_2$  (Hydrogen peroxide) and oxygen by the enzyme superoxide dismutase.

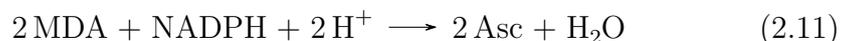


Hydrogen peroxide itself is a reactive oxygen species as well, however it is not as volatile as most others. In order to detoxify hydrogen peroxide, an entire biochemical cycle has evolved in higher plants, the ascorbate-glutathione cycle [36].

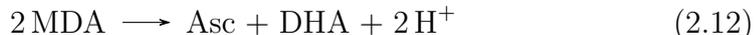
In this cycle, the ascorbate peroxidase (APX) reduces hydrogen peroxide to water by transferring electrons from ascorbate, producing monodehydroascorbate.



Monodehydroascorbate is a highly reactive radical, and can be reduced by the monodehydroascorbate reductase (MDAR), consuming an reduction equivalent in the form of NADPH in the process.



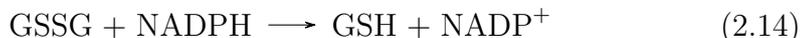
In the case that monodehydroascorbate is not reduced rapidly enough, it might split into ascorbate and dehydroascorbate.



Dehydroascorbate can then be reduced by the dehydroascorbate reductase (DHAR) consuming an glutathione, and producing ascorbate and glutathione disulfide.



Glutathione disulfide can be transformed back into glutathione by the glutathione reductase, which uses NADPH as an reduction equivalent in the process.



This cycle is an efficient way to quickly detoxify hydrogen peroxide which evolved during the Mehler reaction at photosystem I [62]. Overall, electrons that have been transferred from water, splitting it into oxygen, have now been used via the Mehler reaction, the superoxide dismutase and the ascorbate-glutathione cycle to end up as water again, therefore this alternative electron flow is the second route considered to be a "water-to-water cycle". Although this alternative electron flow is not only unintended, but includes the risk of damage via reactive oxygen species and the consumption of energy in the form of NADH or NADPH, it offers an additional valve for electron overflow on the photosynthetic electron transport chain. Furthermore, this electron flow increases the proton gradient on the thylakoid membrane, and therefore activates processes of non-photochemical quenching to protect the plant cell [36].

## 2.3 Light independent reactions

### 2.3.1 The Calvin Benson Bassham Cycle

The light dependent reactions of photosynthesis are converting light energy into chemical energy in the form of the energy equivalent ATP and the redox equivalent NADPH. However, as every other living organism, plants need to maintain an extensive carbohydrate metabolism. To access carbohydrates, plants use ATP and NADPH produced by light dependent reactions in order to fix carbon dioxide in a series of biochemical reactions. These reactions are found in the stroma of the chloroplasts and are commonly known as the light independent reactions of photosynthesis [6].

Amongst these light independent reactions, the reactions of the Calvin-Benson-Bassham cycle (CBB cycle) are the main reactions responsible for

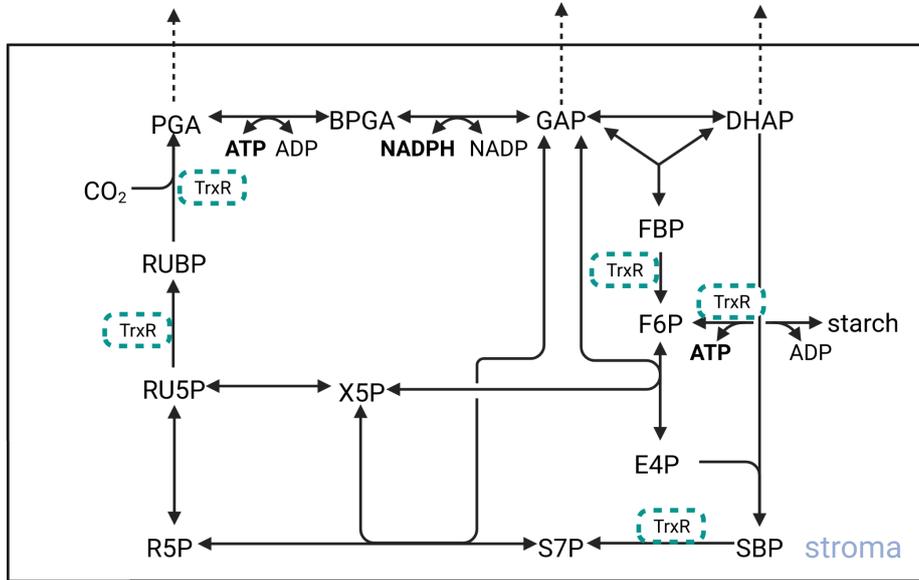


Figure 3: Schematic of the reactions of the Calvin-Benson-Bassham cycle reactions (Figure taken from [55])

photosynthetic carbon fixation [56]. This biochemical cycle can be split up into two phases. In the first phase, which can be considered the fixation phase, the enzyme Ribulose-1,5-bisphosphate-carboxylase-oxygenase (RuBisCO) converts a molecule of carbon dioxide and Ribulose-1,5-bisphosphate (RUBP), which is a molecule containing 5 carbon atoms, into two molecules of 3-phosphoglycerate (PGA).



This enzymatic step is crucial and universal for carbon fixation in plants, and therefore RuBisCO is the most abundant enzyme on planet earth [4]. The two molecules of PGA themselves are already 3C carbohydrates, but are converted to glycerate-1,3-bisphosphate (BPGA) by the enzyme phosphoglycerate kinase by consuming a molecule of ATP and returning a molecule of ADP for each reaction [26].



In the next step, the enzyme glyceraldehyde 3-phosphate dehydrogenase converts the two BPGA molecules into glyceraldehyde 3-phosphate (GAP) by consuming a molecule of NADPH and returning a  $\text{NADP}^+$  molecule for each reaction [66].



While PGA is a carbohydrate as well and could be in principle used for other remaining metabolic pathways, GAP is a crucial carbohydrate because it can be used as a metabolite in the plant metabolism, and it is also a critical precursor in the next phase of the CBB cycle, which is the regenerative phase [52]. This phase is responsible for the regeneration of RUBP, which is the initial substrate of RuBisCO and therefore essential for further carbon fixation. GAP is a 3C molecule, while RUBP is a 5C molecule. Therefore, five GAP molecules are required to regenerate three RUBP molecules. Hence, three carbon fixation phases (and therefore three carbon dioxide molecules) are required to overall fix one molecule of GAP and replenish the essential RUBP in the process.

The regenerative phase starts with the catalyzation of GAP into dihydroxyacetone phosphate (DHAP) by the enzyme triose phosphate isomerase [16]. This isomerase is not changing the atomic composition of GAP, but its conformation.



DHAP and GAP are transformed into fructose-1,5-bisphosphate (FBP) by the enzyme aldolase [24].



In the next step, FBP is dephosphorylated into fructose-6-phosphate (F6P) by the enzyme fructose-1,6-bisphosphatase [60].



F6P and GAP are then converted into erythrose-4-phosphate (E4P), a 4C molecule, and xylulose-5-phosphate (X5P), a 5C molecule, by the transketolase [22].



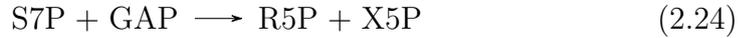
The aldolase then transforms E4P and DHAP into sedoheptulose-1,7-bisphosphate (SBP), a 7C molecule.



SBP is then dephosphorylated into sedoheptulose-7-phosphate (S7P) by the sedoheptulose-1,7-bisphosphatase [24, 60].



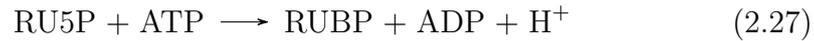
S7P and GAP are then converted into ribose-5-phosphate (R5P) and X5P, both 5C molecules, by the transketolase [22].



Both R5P and X5P are converted into ribulose-5-phosphate (RU5P) by the enzymes ribose-5-phosphate isomerase and ribulose-5-phosphate epimerase, respectively [70, 29].



In the last step of the regeneration phase, the RU5P molecules are converted into the RuBisCO substrate RUBP by the enzyme phosphoribulokinase by consuming ATP and returning ADP [40].



### 2.3.2 Export and storage of carbohydrates

Within the previously described CBB cycle, carbon dioxide is fixed and the main form in which it is used is as triose phosphates. Any usage of energy and mass in the form of fixed carbohydrates starts with these phosphorylated 3C molecules [14]. Triose-phosphates in the chloroplast, which are participating in the regenerative phase of the CBB cycle, are mainly used for two things: Either the synthesis of starch, which is a polysaccharidal energy storage in all plants, or the synthesis of sucrose, which is a saccharide which is the main form of energy and carbon transportation between cells and organs in plants [39]. The synthesis of starch is taking place within the chloroplast, where the triose phosphates themselves are being formed. Within the CBB cycle, the triose phosphates DHAP and GAP are converted into FBP by the aldolase, which is then dephosphorylated into F6P by the fructose-1,6-bisphosphatase. F6P is then converted into glucose-6-phosphate (G6P) by the glucose-6-phosphate isomerase. G6P is then transformed into glucose-1-phosphate (G1P) by the enzyme phosphoglucomutase.



G1P is the main substrate for starch synthesis. The enzyme ADPglucose pyrophosphorylase is using G1P and ATP to produce pyrophosphate and ADP-glucose.



The substrate ADP-glucose can be used by multiple enzymes of the family of starch synthases, to release ADP and add the glucosyl structure of ADP-glucose to a specific position on an existing starch granule. This position is specific to the enzyme utilizing the ADP-glucose [50]. If not converted into starch, triose phosphates may be used for sucrose synthesis, which takes place in the cytosol. Therefore, this process begins with the relocation of triose phosphates from the chloroplast into the cytosol. This relocation is facilitated by triosephosphate/phosphate translocators (TPTs) which are present in the chloroplast membrane and translocate triose phosphates from the chloroplasts to the cytosol, and orthophosphate from the cytosol into the chloroplast [39]. Therefore, the chloroplast stroma is overall releasing carbon without depleting phosphate. This means that the phosphate concentration in the stroma is more or less constant. In the cytosol, GAP and DHAP are converted into FBP and converted into Sucrose via the UDP-glucose-pyrophosphorylase, the sucrose-phosphate synthase and the sucrose-phosphate phosphatase [33].



The translocation of triose phosphates and orthophosphate has many effects on other upstream reactions in photosynthesis. If the rate of triosephosphate translocation is high relative to the rate of carbon fixation, problems might occur. For example, a depletion of the CBB cycle intermediates may cause problems to the regenerative phase of the cycle, even risking its collapse. On the other hand, increased levels of orthophosphate within the chloroplast may lead to inhibition of starch synthesis [39]. However, if the rate of triosephosphate translocation is low relative to the rate of carbon fixation, other problems might occur. Besides the decreased availability of triose phosphates and sucrose for the rest of the cells, a decrease in orthophosphate levels in the chloroplast may cause problems for the ATP-synthesis, which needs orthophosphate as a substrate [59]. Therefore, the rate of triosephosphate translocation is a crucial part of the regulation and limitation of the process of photosynthesis.

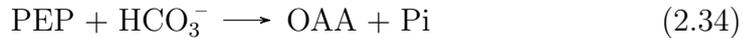
### 2.3.3 Regulation of the Calvin Cycle

The CBB cycle is dependent on the energy and redox equivalents ATP and NADPH, which are provided by the light-reactions of photosynthesis, to fix carbon dioxide into triose phosphates which can be translocated out of the chloroplast. However, as previously described, imbalances in rates of carbon fixation and triosephosphate translocation may cause problems for the CBB cycle. One of these problems is the risk to deplete the cycle of necessary intermediates, needed to enable the regeneration of RUBP, through the triosephosphate translocators. In order to prevent this scenario from happening, the intermediates and the rate of the CBB cycle are regulated by biochemical mechanisms. One regulatory mechanism is the oxidative pentose phosphate pathway, which uses starch to synthesize additional 5C and 6C sugars to replenish the intermediate pool of the CBB cycle [58]. This regulatory mechanism may consume fixed carbon, but prevents the CBB cycle from collapsing during low CBB cycle fluxes, e.g. in extended periods of low light. Another major regulatory mechanism is the light-regulation of the rate of specific CBB cycle enzymes. The rate of these enzymatic reactions is lowered in low-light conditions in order to prevent the CBB cycle from working at constant rates in low light (and therefore low energetic) conditions, in which the CBB cycle intermediates may deplete. This regulatory mechanism is facilitated by the thioredoxin reductase system, which regulates the activity of CBB cycle enzymes dependent on the amount of reduced ferredoxin in the photosynthetic electron transport chain [27]. The enzymes fructose-1,6-bisphosphatase, sedoheptulose-1,6-bisphosphatase, phosphoribulokinase and ADPglucose pyrophosphorylase [27, 57] are light-regulated via the thioredoxin system, while RuBisCO is regulated by light activation via the RuBisCO activase [69]. The light-regulation of these enzymes is not only beneficial for maintaining the stability of the CBB cycle, but is also a component of the delicate balancing between the input and output of the light dependent and light independent reactions of photosynthesis for maximal photosynthetic efficiency.

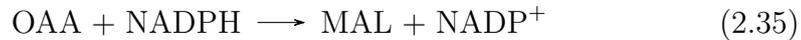
### 2.3.4 Carbon fixation in C3, C4, CAM plants

In all plants, RuBisCO is the key enzyme for carbon fixation within the CBB cycle. However, aside from using carbon dioxide as a substrate, the RuBisCO also may use oxygen instead of carbon dioxide as substrate [1]. In this reaction, which is the oxygenation reaction of RuBisCO, two molecules of 2-phosphoglycolate (2PG) are produced instead of two molecules of PGA. 2PG is toxic and is therefore detoxified in a long and energy costly metabolic

pathway named photorespiration [13]. In most plants photorespiration is a well known process, while some other plant types developed cell structures to minimize the rate of photorespiration. Generally, plants can be distinguished into three types of carbon fixation pathways; C3, C4 and CAM. The most abundant type is the C3 carbon fixation pathway, in which carbon dioxide enters the plant via the stomata, from which it diffuses into the chloroplasts of cells and is used as a substrate for RuBisCO [67]. Notably, this same path may be used by oxygen and therefore result in an oxygenation reaction of RuBisCO. However, C4 plants evolved a different pathway, in which the carbon dioxide is initially diffusing into the mesophyll cells of the leaves [8]. There, the enzyme phosphoenolpyruvate carboxylase (PEPC) is using carbon dioxide (in the form of bicarbonate) and phosphoenolpyruvate (PEP) to produce oxalacetate (OAA).



OAA is then reduced to malate by the malate dehydrogenase, which is then transported to the bundle sheath cells of the leaves. In the chloroplasts of the bundle sheath cells, malate is decarboxylated by the NADP-malic enzyme (NADP-ME) to carbon dioxide and Pyruvate (PYR), and carbon dioxide can be used by RuBisCO as a substrate.



The spatial separation of the location of carbon dioxide (and oxygen) diffusion, and the location of carbon fixation by RuBisCO is the key difference between C3 and C4 plants, and enables a much higher carbon dioxide concentration near RuBisCO. The name C4 refers to the 4 carbon molecules (OAA and malate) which are shuttling the carbon dioxide to the bundle sheath chloroplasts [67].

Plants that use the CAM (Crassulacean acid metabolism) also accomplish a higher carbon dioxide concentration at the RuBisCO location, but this separation is not spatial but temporal [34]. CAM plants take up carbon dioxide at night via the stomata, where it diffuses into the cells and is converted into OAA via the phosphoenolpyruvate carboxylase, and then reduced to malate via an NAD-malic enzyme. The synthesized malate is accumulated within the vacuole. During the day, the plant stomata remain closed and malate is released from the vacuole, decarboxylated, and the resulting carbon dioxide can be used by RuBisCO. The closed stomata

prevent any unwanted oxygen molecules from diffusing into the cells during the day through the stomata [67]. The name crassulacean acid metabolism is derived from the discovery of the pathway in plants of the Crassulaceae family.

## 3 Introduction to modeling

### 3.1 General Introduction to metabolic modeling

Understanding biological phenomena and behaviors begins with observations through experimentation. However, even with extensive knowledge about the subject and its details, it is very difficult to understand the underlying mechanisms responsible for the observation from studying individual observations. Furthermore, understanding the role and effects of such observations in the context of a bigger, complex biological system is almost impossible. In order to get a deep understanding about these biological phenomena, it is important to study them within the context of their surrounding system. The collection and integration of biological knowledge and data from separate measurements and experiments can culminate in a model of the investigated biological system, which can be used to understand such behaviors within complex systems.

A model is used as a simplified representation of systems and offers a more global perspective on processes. In contrast to models in experimental biology, which are usually species or individuals which can be useful for particular investigations, models can be purely theoretical in order to qualitatively or quantitatively represent the investigated system. Theoretical models can be as simple as a visual schematic representation of e.g. interactions between species within an ecosystem. However, models can also integrate far more information and be used for simulations like constrained based networks, systems of ordinary differential equations, or even systems of partial differential equations. Such models require in-depth information about the biological systems like stoichiometric coefficients, appropriate rate equations, measured kinetic parameters and more. Such models can be used for simulations of biological processes over time and are extremely useful to understand the role of individual processes within the context of a more complex system. More recently, the usage of mathematical models to investigate biological systems has been known as systems biology. Mathematical models are usually constructed not only to encapsulate existing information about the studied biological system, but also to answer specific research questions. Therefore, the level of complexity of the model is also highly dependent on the research question. Generally, mathematical models should be constructed as simplistic as possible, and should only grow more complex if needed to answer the underlying research questions. Even slightly more complex mathematical models in systems biology are

difficult to simulate and understand by hand. In order to access the true capabilities of mathematical models, the implementation and simulation of models using computer programs allows a broad range of calculations and analyses. While there are existing programs which are intended for mathematical modeling of biological systems (e.g. CellNetAnalyzer, Copasi, Matlab), scientists are also able to implement models within programming languages (e.g. Python, Julia, C, R) for access to modeling libraries and generally less restrictions.

In the following section, the representation of biochemical systems as models using mathematical methods is addressed. It is supposed to gain an overview of the concepts of kinetic and constraint based modeling necessary to understand the techniques applied in the presented manuscripts. This section draws heavy inspiration from the books "The regulation of cellular systems" from 1996 by Heinrich and Schuster [21] and "Systems biology: a textbook" from 2016 by Klipp et al [28].

### 3.2 Kinetic modeling

In kinetic modeling of biochemical systems, the focus lies on the description of the dependence of systems on time and space. One mathematical method is the description of changes over time of quantities using differential equations. For changes over time, ordinary differential equations are adequate. However, if temporal and spatial changes are considered, partial differential equations are needed. In this thesis, all models are not describing changes in space and are therefore not using partial differential equations. Instead, in this section, the concept of modeling dynamic biochemical systems with ordinary differential equations is described.

The behavior of biochemical systems over time can be described by a set of differential equations

$$\frac{dS_i}{dt} = f_i(S_1, \dots, S_n, p_1, \dots, p_j, t) \quad i, j = 1, \dots, n \quad (3.1)$$

in which  $S_i$  denotes quantities (e.g. the metabolite concentrations) and  $p_j$  denotes parameters, or rate constants, and  $t$  represents time. Equation (3.1) can be written in vector notation as

$$\frac{d\mathbf{S}}{dt} = f(\mathbf{S}, \mathbf{p}, t) \quad (3.2)$$

in which  $\mathbf{S} = (S_1, \dots, S_n)^T$ ,  $\mathbf{f} = (f_1, \dots, f_n)^T$  and  $\mathbf{p} = (p_1, \dots, p_j)^T$ .

In Eq. (3.1),  $f_i$  and  $(p_1, \dots, p_j)$  depict a simplified description of the modeled biochemical system. This kinetic description is represented as rate equations, which are described in more detail in a later part of this section.

In order to analyze the modelled biochemical system of interest, one may be interested in finding solutions of ordinary differential equations. Implicit ordinary differential equations

$$F(t, x, x', \dots, x^{(n)}) = 0 \quad (3.3)$$

contain the time variable  $t$ , the unknown function  $x$ , as well as all derivatives of  $x$  up to the  $n$ th order. Explicit ordinary differential equations of the  $n$ th order are represented as

$$x^{(n)} = f(t, x, x', \dots, x^{(n-1)}) \quad (3.4)$$

in which, again, the order of the ordinary differential equation is determined by  $n$  (the highest derivative). The solution of an  $n$ th order ordinary differential equation depends on  $n$  integration parameters. Such a solution is considered a general solution. However, when certain integration constants are specified (like the initial concentration of the system's metabolites), one may obtain a particular solution for the ordinary differential equation. While it is extremely difficult to solve multi-dimensional systems of ordinary differential equations analytically, it is certainly possible to use numerical integration for approximations to particular solutions. In the following sections, the specifics of balance equations, stoichiometric coefficients and reaction rates are described in more detail.

### 3.2.1 Balance Equations

The foundation of biochemical reaction kinetics is that reaction rate  $v$  at a point  $\mathbf{r}$  in space at a time  $t$  can be expressed as a nonlinear function of the concentrations of all included metabolites  $S_i$  (and frequently of time) at point  $r$  and at time point  $t$ .

$$v(\mathbf{r}, t) = v[\mathbf{S}(\mathbf{r}, t), t] \quad (3.5)$$

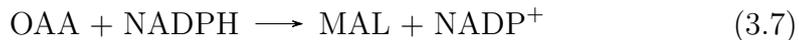
In Eq. (3.5), the vector of metabolite concentrations is represented by  $S$ , and the reaction rate  $v$  is allowed to be dependent on the time  $t$ . Systems which are explicitly dependent on time, such as systems with oscillating clocking inputs, are considered non-autonomous systems. However, in this thesis such systems are not described and therefore the focus is on autonomous systems which are not explicitly dependent on time. The

states of autonomous biochemical systems are determined by all metabolite concentrations at a certain point in space. However, this state is not only given by the metabolite concentrations, but also by rate constants, or parameters, which remain constant at any time point. Temperature and pressure may be such parameters when systems are considered isothermic and isobaric, like it is the case in this thesis. In some systems, certain metabolite concentrations can be considered constant and can therefore be treated as parameters as well. In the models presented in this thesis, spatial homogeneity is assumed, meaning that all metabolite concentrations are uniform in the considered volumes.

In biochemical systems, the stoichiometry of reactions is essential to characterize the proportions of molecularities of substrates and products of reactions. If we use the reaction



catalyzed by the aldolase, the stoichiometric coefficients of GAP and DHAP are -1, while the stoichiometric coefficient of FBP is 1. The signs of the stoichiometric coefficients depend on the role as substrate or products of the metabolites within the reaction. In reversible reactions, the signs of stoichiometric coefficients may change, depending which reaction direction is considered. Sets of stoichiometric coefficients of reactions can be described as vectors, which is especially useful in systems including a multitude of different reactions, as these vectors can be applied in a matrix. In these stoichiometric matrices, the rows indicate the metabolites and the columns indicate the reactions of the system. For example, if we consider a system including the malate dehydrogenase (MDH) and the NADP-malic enzyme (NME)



the vectors of stoichiometric coefficients can construct a stoichiometric matrix  $\mathbf{N}$

$$\mathbf{N} = \begin{array}{cc} & \begin{array}{cc} \text{MDH} & \text{NME} \end{array} \\ \left[ \begin{array}{cc} -1 & 0 \\ 1 & -1 \\ 0 & 1 \\ 0 & 1 \\ -1 & 1 \\ 1 & -1 \end{array} \right] & \begin{array}{l} \text{OAA} \\ \text{MAL} \\ \text{PYR} \\ \text{CO}_2 \\ \text{NADPH} \\ \text{NADP}^+ \end{array} \end{array} \quad (3.9)$$

in which the stoichiometric coefficients of all metabolites are included that are not considered as constant. Furthermore, chemical species which are practically parameter-independent due to overabundance like water are considered external substrates and are usually not included in stoichiometric matrices.

Biochemical systems almost always consist of a multitude of reactions. While the reaction rates are represented by  $v_j (j = 1, \dots, r)$  and stoichiometric coefficients are represented by  $n_{ij}$  ( $i$  denotes the metabolite and  $j$  denotes the reaction). When biochemical conversions are considered the only source of concentration changes in the system, the time-dependent behavior of the system can be described as

$$\frac{dS_i}{dt} = \sum_{j=1}^r n_{ij} v_j(\mathbf{S}) \quad (3.10)$$

which may be represented in matrix notation as

$$\frac{d\mathbf{S}}{dt} = \mathbf{N}\mathbf{v} \quad (3.11)$$

in which  $\mathbf{S}$  is the vector of metabolite concentrations and  $\mathbf{v}$  is the vector of reaction rates.

Reaction rates, which represent mathematical descriptions and representations of biological and biochemical behaviors of processes (e.g. catalyzed by enzymes). The following section focuses on the description of specifics about rate equations, as well as giving examples for rate laws.

### 3.2.2 Rate equations

The functions  $v_j(\mathbf{S})$  in Eq. (3.10) are rate equations (also known as rate laws or kinetic functions) which, to be precise, are to be formulated as  $v_j(\mathbf{S}, \mathbf{p})$  in which  $\mathbf{p}$  represents a vector of parameters  $p_j$  which are also included in Eqs. (3.1) and (3.2). In this section, two of the most fundamental kinetic

functions used in biochemical modeling (mass-action kinetics and Michaelis-Menten kinetics) are explained. The quantities of kinetic functions are the metabolite concentration  $S$  (the number of molecules per volume) and the reaction rate  $v$  (the change of metabolites  $S$  per time  $t$ ). As mentioned before, for simplicity we assume a spatially homogeneous system which is autonomous. Therefore, rate equations may be written as

$$v(t) = v(S(t), p) \quad (3.12)$$

### 3.2.2.1 The law of mass-action

The mass-action law is the fundament of biochemical kinetics, stating that the rate of a reaction is proportional to the probability of reactants colliding. Therefore, the higher the concentrations of metabolites in reactions, the higher the probability of collision. Generally, mass-action kinetics for reactions in which  $z_i$  substrates with concentrations of  $S_i$  are converted into  $z_j$  products with concentrations of  $P_j$  can be formulated as

$$v = v_+ - v_- = k_+ \prod_{i=1}^{z_i} S_i^{n_i} - k_- \prod_{j=1}^{z_j} P_j^{n_j} \quad (3.13)$$

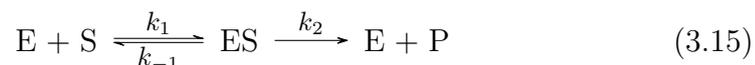
in which  $n_i$  and  $n_j$  represent the stoichiometric coefficients of substrates and products in the reaction. From this, the equilibrium constant  $q$  of reactions can be derived as

$$q = \frac{k_+}{k_-} = \frac{\prod_{j=1}^{z_j} P_{j,eq}^{n_j}}{\prod_{i=1}^{z_i} S_{i,eq}^{n_i}} \quad (3.14)$$

in which  $S_{eq}$  and  $P_{eq}$  represent the respective concentrations in equilibrium.

### 3.2.2.2 Michaelis-Menten kinetics

A general enzymatic mechanism for all one-substrate reactions can be formulated as



in which the formation of the enzyme-substrate complex ES from free enzyme and substrate is reversible, while the catalysis of ES to free enzyme

and product is irreversible. This set of reactions can be written as ordinary differential equations:

$$\begin{aligned}
 \frac{dS}{dt} &= -k_1 \cdot E \cdot S + k_{-1} \cdot ES \\
 \frac{dES}{dt} &= k_1 \cdot E \cdot S - (k_{-1} + k_2) \cdot ES \\
 \frac{dE}{dt} &= -k_1 E \cdot S + (k_{-1} + k_2) \cdot ES \\
 \frac{dP}{dt} &= k_2 \cdot ES
 \end{aligned}
 \tag{3.16}$$

Due to the multidimensionality of this system of ordinary differential equations, no analytical solutions are present. However, with the right assumptions, it is possible to derive mathematical expressions of this system's behavior. One assumption, made by Michaelis and Menten, considers that the rate of the association and dissociation reactions  $k_1$  and  $k_{-1}$  are much faster than the irreversible catalytic reaction  $k_2$ . That suggests that there is a state of quasi-equilibrium between free enzyme E and bound enzyme ES. In line with this, another assumption by Briggs and Haldane considers that if the initial substrate concentration is much higher than the enzyme concentration, the concentration of bound enzyme ES stays relatively constant within a quasi-steady-state. For Eq. 3.16, this means

$$\frac{dES}{dt} = 0
 \tag{3.17}$$

which allows the derivation of rate kinetics from the ordinary differential equations by describing that the total enzyme  $E_{total}$  is constant:

$$\frac{dES}{dt} + \frac{dE}{dt} = 0
 \tag{3.18}$$

meaning that

$$E_{total} = E + ES = constant.
 \tag{3.19}$$

which means that in steady state, the concentration of the enzyme-substrate complex can be expressed as

$$ES = \frac{k_1 \cdot E_{total} \cdot S}{k_1 \cdot S + k_{-1} + k_2} = \frac{E_{total} \cdot S}{S + \frac{(k_{-1} + k_2)}{k_1}} \quad (3.20)$$

and resulting in the reaction rate

$$v = \frac{k_2 \cdot E_{total} \cdot S}{S + \frac{(k_{-1} + k_2)}{k_1}} \quad (3.21)$$

In order to obtain a simpler form of this equation, the catalytic rate parameter and enzyme concentration can be expressed together as the maximum reaction rate is

$$E_{total} \cdot k_2 = V_{max} \quad (3.22)$$

and the substrate concentration at which the reaction is at half its maximum rate is calculated as

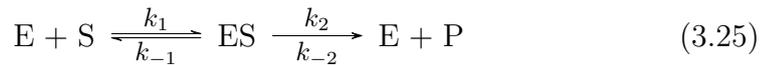
$$\frac{k_{-1} + k_2}{k_1} = K_m \quad (3.23)$$

in which  $K_m$  is known as the Michaelis constant. Eqts. (3.21), (3.22) and (3.23) can be used to obtain the practical expression

$$v = \frac{V_{max} \cdot S}{K_m + S} \quad (3.24)$$

for the reaction rate of enzymatic one-substrate Michaelis-Menten kinetics.

In the case of enzymatic one-substrate reactions which are entirely reversible, the enzymatic mechanism can be described as



which establishes maximum reaction rates ( $V_{max}^{for}$  and  $V_{max}^{back}$ ) and Michaelis constants ( $K_{mS}$  and  $K_{mP}$ ) for both forward and backward reactions, respectively. This establishes

$$v = \frac{\frac{V_{max}^{for}}{K_{mS}} \cdot S - \frac{V_{max}^{back}}{K_{mP}} \cdot P}{1 + \frac{S}{K_{mS}} + \frac{P}{K_{mP}}} \quad (3.26)$$

with the equilibrium constant  $q$  defined as

$$q = \frac{V_{max}^{for} \cdot K_{mS}}{V_{max}^{back} \cdot K_{mP}} \quad (3.27)$$

While generally, an increase of substrate concentration  $S$  increases reaction rates up to a maximum rate  $V_{max}$ , there are enzymatic reactions in which a decrease in reaction rate is observed after a certain threshold of substrate concentration. This phenomenon is regarded as substrate inhibition. One suggested explanation for this scenario is the binding of additional substrates to the enzyme, forming an reversible enzyme-substrate-substrate complex which occupies enzymes without product catalysis. This inhibition can be competitive (occupation of enzymes from metabolites similar to the substrate) or noncompetitive (occupation by more than one molecule of the substrate). Both cases can be formulated in an extended version of Michaelis-Menten kinetics which is expressed as

$$v = k_2 \cdot ES = \frac{V_{max} \cdot S}{K_m + S \cdot (1 + \frac{I}{K_I})} \quad (3.28)$$

in which the inhibition constant  $K_I$  is determining the extent of inhibition while the inhibitor  $I$  can be substrate  $S$  in the case of noncompetitive inhibition, or a different metabolite in the case of competitive inhibition.

### 3.2.2.3 Convenience kinetics

In an attempt to establish an expression of reaction rates in which all parameters are independent of each other and to simplify estimations of mechanisms and parameters, convenience kinetics have been proposed. Convenience kinetics apply a formulation of Michaelis-Menten kinetics which can be applied to all reaction stoichiometries. Convenience kinetics are expressed as

$$v = E_{total} \cdot \frac{k_{cat}^{for} \prod_i (\frac{S_i}{K_{m,S_i}})^{n_i} - k_{cat}^{back} \prod_j (\frac{P_j}{K_{m,P_j}})^{n_j}}{\prod_i (1 + (\frac{S_i}{K_{m,S_i}}) + \dots + (\frac{S_i}{K_{m,S_i}})^{n_i}) + \prod_j (1 + (\frac{P_j}{K_{m,P_j}}) + \dots + (\frac{P_j}{K_{m,P_j}})^{n_j}) - 1} \quad (3.29)$$

in which  $E_{total}$  represents enzyme concentration,  $K_m$  represents metabolite concentrations in which the reaction rate (in respective reaction direction) is at half its maximum rate and  $k_{cat}^{for}$  and  $k_{cat}^{back}$  represent turnover rates.

### 3.2.3 Steady States and their stabilities

The investigation of systems in steady state is an essential component in the analysis of dynamic models. The concept of steady states is especially important for biochemical systems that do not change in metabolite

concentrations for a certain time. However, it is important to note that the analysis of systems in steady state is mathematically favorable, but remains an approximation of realistic behavior, as life is always dynamic and fluctuating. Nevertheless, the calculation of steady states, as well as the analyses that are offered by it, are very helpful to develop a deeper understanding of the investigated system.

One important aspect of systems is analyses of the stability of steady states. Linearizing systems in steady state is necessary to start such analyses. If a biochemical system with the steady state  $\bar{\mathbf{S}}$  is perturbed by  $\hat{\mathbf{S}}(t)$  with

$$\mathbf{S}(t) = \bar{\mathbf{S}} + \hat{\mathbf{S}}(t) \quad (3.30)$$

the change of the deviation over time can be written as

$$\dot{\mathbf{S}} = \frac{d}{dt}(\bar{\mathbf{S}} + \hat{\mathbf{S}}(t)) = \frac{d}{dt}\hat{\mathbf{S}}(t) \quad (3.31)$$

The system equations can be used to perform a Taylor expansion for further stability analyses. It follows

$$\frac{d}{dt}\hat{S}_i = f_i(\bar{S}_1, \dots, \bar{S}_n) + \sum_{j=1}^n \frac{\partial f_i}{\partial S_j} \hat{S}_j + \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \frac{\partial^2 f_i}{\partial S_j \partial S_k} \hat{S}_j \hat{S}_k + \dots \quad (3.32)$$

In steady state, in which  $f_i(\bar{S}_1, \dots, \bar{S}_n) = 0$ , we can disregard terms of higher order and it holds that

$$\frac{d}{dt}\hat{S}_i = \sum_{j=1}^n \frac{\partial f_i}{\partial S_j} \hat{S}_j = \sum_{j=1}^n a_{ij} \hat{S}_j \quad (3.33)$$

in which  $a_{ij} = \frac{\partial f_i}{\partial S_j}$  are constant coefficients at steady state. From  $a_{ij}$ , the Jacobian Matrix can be constructed as

$$\mathbf{J} = \{a_{ij}\} = \begin{pmatrix} \frac{\partial f_1}{\partial S_1} & \frac{\partial f_1}{\partial S_2} & \dots & \frac{\partial f_1}{\partial S_n} \\ \frac{\partial f_2}{\partial S_1} & \frac{\partial f_2}{\partial S_2} & \dots & \frac{\partial f_2}{\partial S_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial S_1} & \frac{\partial f_n}{\partial S_2} & \dots & \frac{\partial f_n}{\partial S_n} \end{pmatrix} \quad (3.34)$$

which is the system matrix  $\mathbf{A}$  in linear systems, meaning  $\mathbf{J} = \mathbf{A}$ .

In order to find the steady state  $\bar{\mathbf{S}}$  of the linear ODE system,  $\dot{\mathbf{S}} = 0$  indicates that

$$\mathbf{A} \cdot \bar{\mathbf{S}} + \mathbf{z} = 0 \quad (3.35)$$

which means that the steady state can be found by matrix inversion of  $\mathbf{A}$

$$\bar{\mathbf{S}} = -\mathbf{A}^{-1} \cdot \mathbf{z}. \quad (3.36)$$

In such linear homogeneous systems of ordinary differential equations, it holds

$$\dot{\mathbf{S}} = \mathbf{A} \cdot \mathbf{S} \quad (3.37)$$

which enables the approach

$$\dot{\mathbf{S}} = \mathbf{b} \cdot \lambda \cdot e^{\lambda t} = \mathbf{A} \cdot \mathbf{b} \cdot e^{\lambda t} \quad (3.38)$$

which can be adjusted to read

$$(\mathbf{A} - \lambda \mathbf{I}_n) \mathbf{b} = 0 \quad (3.39)$$

where  $\mathbf{I}$  denotes the identity matrix, and which can be practically rewritten as

$$\mathbf{S}(t) = \sum_{i=1}^n c_i \mathbf{b}_i e^{\lambda_i t} \quad (3.40)$$

where  $\mathbf{b}_i$  are eigenvectors and  $\lambda_i$  are eigenvalues of  $\mathbf{A}$ , while  $c_i$  are unknown coefficients which are determined by initial conditions for particular solutions.

The eigenvalues of the system can be calculated with

$$\text{Det}(\mathbf{A} - \lambda \mathbf{I}) = a_n \lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_1 \lambda + a_0 = 0 \quad (3.41)$$

which is a characteristic polynomial equation of order  $n$ . The eigenvalues that can be calculated from Eq. 3.41 give the desired information about the system's stability. Generally, if all eigenvalues (real or complex) are negative, the steady state is stable. However, if any eigenvalue is positive, the steady state is unstable.

### 3.2.4 Metabolic Control Analysis

The characterization of the effect of reactions on the system as a whole is very important for a deeper biological understanding of investigated systems. However, this can be particularly hard in dynamic systems containing many variables and many interdependent regulations. Metabolic Control analysis is a theoretical framework which allows the analysis of the steady-state relationships between single reactions and the biochemical

network in a quantitative, and qualitative fashion. The aim of metabolic control analysis is to understand the sensitivities of reactions and metabolite concentrations of networks in steady state on small parameter perturbations. The quantification of these sensitivities is performed by the calculation of coefficients, specifically elasticity-coefficients and control-coefficients.

To introduce the general form for such coefficients, we take the quantity  $y(x)$  as an example which is dependent on  $x$ . We can define the sensitivity coefficient  $c_x^y$  as

$$c_x^y = \left( \frac{x \Delta y}{y \Delta x} \right)_{\Delta x \rightarrow 0} \quad (3.42)$$

describing the effect of  $\Delta x$  on  $y$ .  $\frac{x}{y}$  is being used as a normalization factor to express the sensitivity coefficient independent of units. For limiting cases  $\Delta x \rightarrow 0$ , Eq. 3.42 can be expressed as

$$c_x^y = \frac{x}{y} \frac{\partial y}{\partial x} = \frac{\partial \ln y}{\partial \ln x} \quad (3.43)$$

This general form can be applied to the coefficients calculated for metabolic control analysis. These coefficients can either be local or global. Elasticity coefficients are local coefficients, which means that the coefficient quantifies the effect of perturbations on the reaction velocity itself, not on the entire system. These perturbations are either changes in a parameter or the concentration of a substrate. The elasticity  $\varepsilon$  is written as

$$\varepsilon_i^k = \frac{S_i}{v_k} \frac{\partial v_k}{\partial S_i} \quad (3.44)$$

which calculates the sensitivity of reaction rate  $v_k$  to changes in metabolite concentration  $S_i$ . The calculation of control coefficients relates to the steady state fluxes  $\mathbf{J}$  of the system, are given by the reaction rates depending on respective parameters and steady state metabolite concentrations. Control coefficients calculate the effect of small parameter perturbations on all fluxes (or concentrations) of the system in steady state, and therefore are considered global coefficients. When calculating the control of rate  $v_k$  on flux  $J_j$ , the flux control coefficient  $C_{v_k}^{J_j}$  can be written as

$$C_{v_k}^{J_j} = \frac{v_k}{J_j} \frac{\partial J_j}{\partial v_k} \quad (3.45)$$

where the change in reaction rate  $v_k$  is caused by perturbations in parameter  $p_k$ .

This calculation can be made for the steady state concentrations of the system as well. When calculating the control of rate  $v_k$  on the steady state

metabolite concentration  $S_i$ , the concentration control coefficient  $C_{v_k}^{S_i}$  can be expressed as

$$C_{v_k}^{S_i} = \frac{v_k}{S_i} \frac{\partial S_i}{\partial v_k}. \quad (3.46)$$

The control coefficients are used in the Summation Theorem of metabolic control analysis to explain the total control over fluxes and concentrations in a stable steady state. It states that

$$\sum_{k=1}^r C_{v_k}^{J_j} = 1 \quad \text{and} \quad \sum_{k=1}^r C_{v_k}^{S_i} = 0 \quad (3.47)$$

where  $r$  is the number of reactions in the system. The summation of flux control coefficients to one and the summation of concentration control coefficients to zero displays that the fluxes and concentrations in steady state in a biochemical system are systemic and the control over them is shared by all reactions (whether in a positive or in a negative fashion). Furthermore, it implies that changes in the control of one reaction are compensated by shifts in control of the other reactions in the system.

### 3.3 Constraint-based modeling

#### 3.3.1 Structural modeling

With increasing sequenced genomes accessible to science allows another approach in the mathematical representation of biochemical systems. The combination of available sequenced genomes and databases with knowledge about biochemical processes in living organisms allows for a systematic reconstruction of metabolic networks. Such networks are commonly known as genome scale metabolic models (GSMs). Such models can be used to simulate responses of metabolic fluxes to changes in gene activities and nutrients, as well as designing desired strains of organisms of interest *in silico*. GSMs describe all processes and biochemical conversions from the uptake of nutrients to the production of cellular biomass mainly using the stoichiometries of reactions. This is also why the construction and analyses of GSMs is also considered "stoichiometric" or "structural" modeling. The reason for the focus on structural model properties is, amongst others, the relatively low computing time and the relatively low number of parameters needed for model simulations.

To understand the concept of structural modeling, consider the reaction network described in Eqts. (3.7) and (3.8). We may introduce two additional

reactions to the model:



Eq. (3.48) represents a continuous influx of OAA into the system, while Eq. (3.49) removes PYR and CO<sub>2</sub> from the system. This network can be expressed as a set of ordinary differential equations by

$$\begin{aligned} \frac{d\text{OAA}}{dt} &= v_{\text{inf}} - v_{\text{MDH}} \\ \frac{d\text{MAL}}{dt} &= v_{\text{MDH}} - v_{\text{NME}} \\ \frac{d\text{PYR}}{dt} &= v_{\text{NME}} - v_{\text{out}} \\ \frac{d\text{CO}_2}{dt} &= v_{\text{NME}} - v_{\text{out}} \\ \frac{d\text{NADPH}}{dt} &= v_{\text{NME}} - v_{\text{MDH}} \\ \frac{d\text{NADP}^+}{dt} &= v_{\text{MDH}} - v_{\text{NME}} \end{aligned} \quad (3.50)$$

which can be expressed as a stoichiometric matrix  $\mathbf{N}$  in matrix notation as shown in Eq. (3.9). The set of ordinary differential equations from Eq.(3.50) may now be expressed in matrix notation as well by

$$\begin{bmatrix} -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_{\text{MDH}} \\ v_{\text{NME}} \\ v_{\text{inf}} \\ v_{\text{out}} \end{bmatrix} = \begin{bmatrix} d\text{OAA}/dt \\ d\text{MAL}/dt \\ d\text{PYR}/dt \\ d\text{CO}_2/dt \\ d\text{NADPH}/dt \\ d\text{NADP}^+/dt \end{bmatrix}$$

or in simpler terms as

$$\mathbf{N} \cdot \mathbf{v} = \frac{d\mathbf{S}}{dt} \quad (3.51)$$

in which  $\mathbf{S}$  is the vector of substrate concentration and  $\mathbf{v}$  is the vector of reaction rates.

### 3.3.2 Flux Balance Analysis

One approach to analyse fluxes in structural networks is Flux Balance Analysis (FBA) [45]. It calculates a solution which maximizes or minimizes fluxes of interest, while maintaining mass-conservation and being restricted by sets of constraints. These constraints play a very big role in FBA, which is why the combination of structural models with FBA is generally called constraint-based modeling.

Structural modeling is assuming a steady state of the biochemical system, in which the system is regarded to be in a state of metabolic exchange at which no substrate concentrations are changing over time. This assumption is expressed as

$$\mathbf{N} \cdot \mathbf{v} = \frac{d\mathbf{S}}{dt} = 0. \quad (3.52)$$

In steady state, the stoichiometric matrix imposes that the flux of mass is balanced and that the overall production of compounds has to be equal to the consumption. With this mass balance, it is given that no species can be produced without a valid source, and the consumption prevents any overaccumulation of compounds.

The stoichiometries and the steady state assumption are not the only constraints in FBA. The upper and lower bounds of reactions are another constraint which limits the solution space of constraint based modeling. These bounds limit the minimal and maximal flux reactions are allowed to carry.

With the feasible solution space constrained by steady state, mass-conservation and upper and lower bounds, fluxes are assigned to reactions via a linear programming based approach which is attempting to maximize or minimize the objective reaction(s). In many models, the objective is to maximize the flux of an organism-specific biomass reaction. Generally, the optimality problem investigated in FBA can be expressed as

$$\begin{array}{ll} \text{max./min.} & \mathbf{c}^T \mathbf{v} \\ \text{subject to} & \mathbf{S}\mathbf{v} = 0 \\ \text{and} & \text{lb} \leq v_i \leq \text{ub}, \quad i = 1, \dots, n \end{array} \quad (3.53)$$

in which  $\mathbf{v}$  is the vector of fluxes,  $\mathbf{c}^T \mathbf{v}$  is the vector representing the objective reaction,  $\mathbf{S} \mathbf{v} = 0$  is the steady state constraint and lb and ub are the lower and upper bounds of reaction  $v_i$ , respectively. There is the possibility to introduce further problem-specific constraints like thermodynamic constraints, maximum overall flux etc. can be introduced as well.

The uptake of substrates (e.g. from a medium) and excretion of byproducts (e.g. from respiration) is described by exchange reactions that can allow the in and efflux of matter through the metabolic model. All reactions introducing matter into the system and removing it out of the system (including the often used biomass reaction) have to be balanced as well to prevent matter from unrealistic disappearance or appearance.

It is important to note that multiple flux vectors may lead to the same optimization result under the same constraints. This means that FBA solutions from linear programming are usually non-unique solutions. In order to still efficiently investigate large metabolic systems, alternative but heavily related methods are used.

To estimate the range of fluxes in any solution that results in the same objective value, flux variability analysis (FVA) can be used [17]. FVA tries to find the range of fluxes for every reaction that can be found in solutions resulting in the same optimized objective value. This analysis is useful for understanding which fluxes are unbound and which fluxes are relatively restricted. This information can offer an overview of the solution space that is present in the current optimization, and how the model can behave in alternative flux solutions.

Another method is parsimonious FBA (pFBA) which is initially finding an optimal solution to the FBA problem [30]. In the next step, a flux vector is determined which represents the same objective value but carries the lowest overall flux. Due to the fact that FBA solutions are non-unique, pFBA solutions are especially useful when comparing simulation results after e.g. perturbations of the system.

## 4 Aims of the thesis

The light dependent and independent processes of photosynthesis are often studied apart from each other, ignoring the fundamental interdependence of both processes. One approach to study the detailed and complex interdependence is the development of mathematical models of photosynthesis that can help answering specific research questions which are difficult to answer experimentally. In the past decades, many kinetic mathematical models have been developed to investigate the complexity of photosynthesis. Still, most of these models have a clear focus on either the light dependent or light independent reactions of photosynthesis. Models which describe the light independent reactions, mainly the CBB cycle are not including the photosynthetic electron transport chain, but simplify the light dependent reaction by either constant ATP and NADPH concentrations, or by providing lumped reactions of ATP and NADPH synthesis [19, 48, 51, 12, 73, 72]. The research focus in these studies has been the kinetics of the reactions of photosynthetic carbon fixation, therefore it is understandable that no detailed descriptions for light dependent reactions have been included for the sake of simplification. Similarly, many kinetic models describing the light dependent processes, mainly the photosynthetic electron transport chain, are not including the light independent reactions [10, 11, 68, 38, 41]. Instead, lumped reactions are used as energy sinks. With the research focus on the details of the electron transport in the thylakoid membrane, it is understandable that a simplified description of the light independent reactions is sufficient. Models like the ‘e-photosynthesis’ model by Zhu et al. (2013) include as many known photosynthetic reactions as possible and therefore result in a very large model. While approach is providing a useful knowledge basis, it is difficult to use it for specific research questions due to its complexity and highly detailed description, making it hard to gain knowledge about precise interactions.

In this thesis, the aim is to use and develop mathematical models of photosynthesis to explore research questions that focus on the behavior of the CBB cycle in different light protocols, as well as the need for regulation of CBB cycle activity in changing light. In additional analyses, we tried to reveal the effect of different CBB cycle rates in changing light intensities on crucial system properties. Furthermore, the view of photosynthesis as a economic supply-demand model is used to find out in which conditions the supply or the demand reactions maintain highest overall control in the

system. Another aim is to examine the behavior of alternative electron flows in changing light conditions and their impact on the production of reactive oxygen species. Lastly, the advantages of photosynthetic activity for the production of secondary metabolites in glandular trichomes are investigated.

## 5 Research topics in this thesis

The results of the thesis are presented in submitted and/or published manuscripts. The results are presented in the following organisation:

- Section 5.1 contains a manuscript which presents the software `modelbase` which has been developed to implement ODE models in the programming language Python. All ODE models presented in this thesis have been implemented using this software package.
- Section 5.2 contains a manuscript which presents the software `moped` which has been developed to construct, modify and ensure reproducibility of constraint based genome scale metabolic models in the programming language Python. All constraint-based models presented in this thesis have been constructed using this software package.
- Section 5.3 contains a manuscript which investigates a merged model of photosynthesis as a supply-demand model. The results highlight the need of standby mechanisms in carbon fixation for long phases of darkness, as well as shifting control from supply to demand reactions in increasing light conditions.
- Section 5.4 contains a manuscript which researches the role of alternative electron flows in the production of reactive oxygen species and in the rate of linear electron flow and carbon fixation. It is shown that cyclic electron flow and carbon fixation improve linear electron flow and decrease reactive oxygen evolution by providing additional terminal electron acceptors and prevention of an overreduction of the photosynthetic electron transport chain.
- Section 5.5 contains a manuscript which highlights the light-induced increase in secondary metabolite production in photosynthetic glandular trichomes. The results show shifts in carbon partitioning from catabolic to anabolic reactions, as well as a switch in isoprenoid synthesis from the MEV to the MEP pathway.

## 6 Manuscripts

### 6.1 Building mathematical models of biological systems with `modelbase`

#### 6.1.1 Authors and Details

Authors: Oliver Ebenhöf, Marvin v. Aalst, Nima P. Saadat,  
Tim Nies, Anna Matuszyńska

Authorship: 2. Author

Journal: Journal of Open Research Software

Status: Published (<https://doi.org/10.5334/jors.236>)

#### 6.1.2 Contributions

Oliver Ebenhöf initiated the project, implementing code at an early stage and providing teaching materials. Marvin van Aalst further improved the code, as well as the documentation and implemented a kinetic CBB cycle model to show the utility of `modelbase`. Nima P. Saadat developed code to provide export options for `modelbase` models as SBML files for reproducibility. Tim Nies implemented a Pentose-Phosphate-Pathway model to show the utility of `modelbase`. Anna Matuszyńska further improved code and prepared all example notebooks, as well as the first draft of the manuscript.

## SOFTWARE METAPAPER

# Building Mathematical Models of Biological Systems with `modelbase`

Oliver Ebenhöh<sup>1,2</sup>, Marvin van Aalst<sup>1</sup>, Nima P. Saadat<sup>1</sup>, Tim Nies<sup>1</sup> and Anna Matuszyńska<sup>1,2</sup>

<sup>1</sup> Institute for Quantitative and Theoretical Biology, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, DE

<sup>2</sup> Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, DE

Corresponding author: Anna Matuszyńska ([anna.matuszynska@hhu.de](mailto:anna.matuszynska@hhu.de))

The `modelbase` package is a free expandable Python package for building and analysing dynamic mathematical models of biological systems. Originally it was designed for the simulation of metabolic systems, but it can be used for virtually any deterministic chemical processes. `modelbase` provides easy construction methods to define reactions and their rates. Based on the rates and stoichiometries, the system of differential equations is assembled automatically. `modelbase` minimises the constraints imposed on the user, allowing for easy and dynamic access to all variables, including derived ones, in a convenient manner. A simple incorporation of algebraic equations is, for example, convenient to study systems with rapid equilibrium or quasi steady-state approximations. Moreover, `modelbase` provides construction methods that automatically build all isotope-specific versions of a particular reaction, making it a convenient tool to analyse non-steady state isotope-labelling experiments.

**Keywords:** isotope labelling; mathematical modelling; metabolic networks; Python; ODEs; open science; systems biology; theoretical biology; QSSA

**Funding statement:** This work was financially supported by the Deutsche Forschungsgemeinschaft "Cluster of Excellence on Plant Sciences" CEPLAS (EXC 1028).

## (1) Overview

### Introduction

Well designed mathematical models are excellent theoretical frameworks to analyse and understand the dynamics of a biological system. Here, the design process itself is the first important scientific exercise, in which biological knowledge is collected, organised and represented in a new, systematic way, that allows defining the model assumptions and formulating them in the language of mathematics. A working model then enables testing new hypotheses and allows for novel predictions of the system's behaviour. Kinetic models allow simulating the dynamics of the complex biochemistry of cells. Therefore, they have the power to explain which processes are responsible for observed emergent properties and they facilitate predictions on how the system behaves under various scenarios, such as changed environmental conditions or modification of molecular components. Unfortunately, the construction of mathematical models is often already a challenging task, hampered by the limited availability of measured physiological and kinetic parameters, or even incomplete information regarding the network structure. It

is therefore highly desirable to make the overall process of model construction as easy, transparent and reproducible as possible. Providing a toolbox with a wide range of methods that flexibly adapt to the scientific needs of the user, `modelbase` greatly simplifies the model-building process, by facilitating a systematic construction of kinetic models fully embedded in the Python programming language, and by providing a set of functionalities that help to conveniently access and analyse the results.

Despite the fact that mathematical models vary significantly in their complexity, from very simple and abstract models to extremely detailed ones, they share a set of universal properties. The process of building a kinetic model can be divided into a number of mandatory steps such as i) establishing the biological network structure (the stoichiometry), ii) defining the kinetic rate expressions, iii) formulation of the differential equations, iv) parametrisation, v) validation and, finally, vi) application [1]. `modelbase` supports researchers in every step of model development and application with its simple design aimed at being minimally restrictive. It has been written in Python, an open source, general-purpose, interpreted,

interactive, object-oriented, and high-level programming language. Due to a long list of its general features, such as clear syntax, useful built-in objects, a wealth of general-purpose libraries, especially NumPy and SciPy, Python has become a widely used scientific tool [2]. Needless to say, the usage of Python over other, proprietary software, such as MATLAB or Wolfram Mathematica, decreases the risk of limited reproducibility and transparency, two critical factors while conducting research. Unfortunately, several powerful models of central biochemical pathways [3, 4] have been published before this need became apparent. As a consequence, some of these models are extremely difficult to implement to even attempt to reproduce their results. Therefore, `modelbase` provides an environment for relatively easy implementation of models that were published without source code, in a general-purpose and reusable format. Moreover, `modelbase` supports the export of a structural (stoichiometric) model into Systems Biology Markup Language (SBML) for further structural analysis with the appropriate software.

In recent years, several other Python-based modelling tools have been developed, such as ScrumPy [5] or PySCeS [6]. They allow performing various analyses of biochemical reaction networks, ranging from structural analyses (null-space analysis, elementary flux modes) to kinetic analyses and calculation of control coefficients. To the best of our knowledge they do not provide dedicated methods for model construction inside Python, and the standard usage relies on loading previously assembled model definition files.

The `modelbase` package presented here provides an alternative toolbox, complementing the functionalities of existing programs for computer modelling. Its power lies mainly in integrating the model construction process into the Python programming language. It is envisaged that `modelbase` will greatly facilitate the model construction and analysis process as an integral part of a fully developed programming environment.

### Motivation

In the course of our photosynthetic research, we identified several shortcomings that are not adequately met by available free and open source research software. When constructing a series of similar models, which share the same basic structure but differ in details, it is, in most modelling environments, necessary to copy the model definition file (or even pieces of code) and perform the desired modifications. This makes even simple tasks, such as changing a particular kinetic rate law, hideous and unnecessarily complicated, affecting the overall code readability. To facilitate a systematic and structured model definition, exploiting natural inheritance properties of Python objects, our intention was to fully integrate the model construction process into the Python programming language, allowing for an automated construction of model variants. The necessity for this fully Python-embedded approach became further evident for isotope label-specific models [7], where an automatic construction of isotope-specific reactions from a common rate law and an atom transition map is desired. Such models are, for

example, required to explain complex phenomena, such as the asymmetric label distribution during photosynthesis, first observed by Gibbs and Kandler in the 1950s [8, 9].

### Implementation and architecture

`modelbase` is a console-based application written in Python. It supplies methods to construct various dynamic mathematical models, using a bottom-up approach, to simulate the dynamic equations, and analyse the results. We deliberately separated construction methods from simulation and analysis, in order to reflect the experimental approach. In particular, a model object constructed using the `Model` class can be understood as a representation of a model organism or any subsystem, on which experiments are performed. Instances of the `Simulator` class in turn correspond to particular experiments. The software components of `modelbase` are summarised in the Unified Modeling Language (UML) diagram in **Figure 1**.

### Model construction

The user has the possibility to build two types of models, using one of the classes defined in the module `model`: `Model`, for differential-equation based systems, or `LabelModel`, for isotope-labelled models.

### Class `Model`

Every model object is defined by:

1. model parameters,
2. model variables,
3. rate equations,
4. stoichiometries.

Model parameters can be simply defined in a dictionary, `d`. To build a mathematical model the user needs first to import the `modelbase` package and instantiate a model object (called `m`):

```
import modelbase
m = modelbase.Model(d)
```

After instantiation, the keys of the parameter dictionary `d` become accessible as attributes of an object of the internal class `modelbase.parameters.ParameterSet`, which is stored as the model's attribute `m.par`.

To add reacting entities of the described system (referred to as species in SBML), e.g., metabolites, we pass a list of compounds names to the `set_cpds` method:

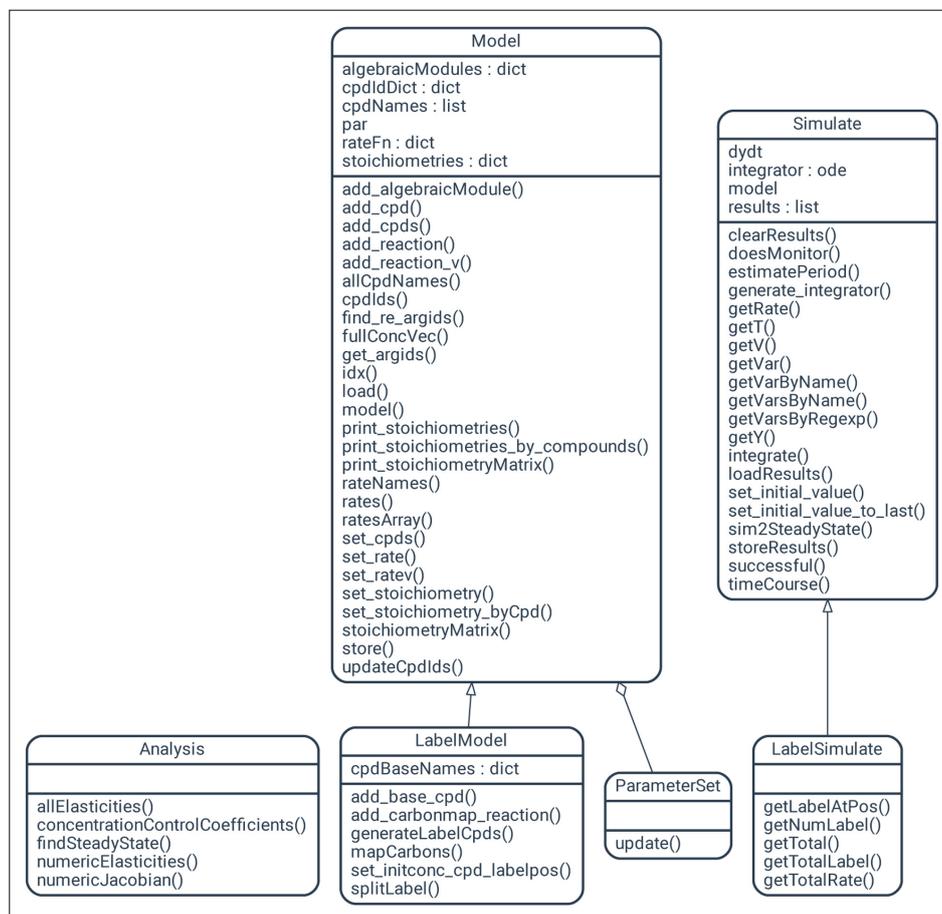
```
m.set_cpds(list_of_compounds)
```

Each of the added compounds becomes a state variable of the system. The full list of all variables is stored in the attribute `m.cpdNames`.

If  $S$  denotes the vector of concentrations of the biochemical reactants (as defined with the method `set_cpds`), the temporal change of the concentrations is governed by:

$$\frac{dS}{dt} = Nv(S, k), \quad (1)$$

where  $N$  denotes the stoichiometric matrix and  $v(S, k)$  the vector of reaction rates as functions of the substrate



**Figure 1:** UML class diagram of `modelbase` software components. It consists of six classes, with `LabelModel` inheriting from `Model` and `LabelSimulate` inheriting from `Simulate`. `ParameterSet` and `Analysis` are special classes containing parameter sets and static methods for analysis respectively.

concentrations  $S$  and parameters  $k$ . The system of ordinary differential equations is assembled automatically after providing all reaction rates and their stoichiometries to the method `m.add_reaction()`. The stoichiometric matrix of a model can be retrieved by the method `m.print_stoichiometries()` or `m.print_stoichiometries_by_compounds()`, for the transposed matrix. A detailed example of instantiating objects and solving a simple biochemical system with three reactions and two metabolites is provided in **Box 1**.

#### Working with algebraic modules

A particularly useful function of the class `Model` has been developed to facilitate the incorporation of algebraic expressions, by which dependent variables can be computed from independent ones. Examples include conserved quantities, such as the sum of adenine phosphates, which is often considered to be constant, and rapid-equilibrium or quasi steady-state approximations (QSSA), which are

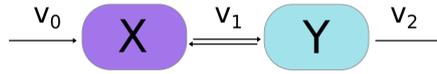
applicable for systems with time-scale separation and allow calculation of fast from slow variables. The function `add_algebraicModule()` accepts as arguments a function describing the rule how the dependent variables are calculated from independent ones, the name of the newly created module, and lists of names of the independent and dependent variables. After definition of an algebraic module, all dependent variables become directly accessible. The full list of independent and dependent variables can be accessed using the method `allCpdNames()`.

#### Various analysis methods

With `import modelbase.Analysis` the user has access to advanced analysis methods on the model object. Currently, it provides methods to numerically calculate elasticities and the Jacobian, find steady states by attempting to solve the algebraic equations, and calculate concentration control coefficients. We expect the range of analysis methods to increase continuously in the future.

**Box 1:** Basic model use

We use `modelbase` to simulate a simple chain of reactions, in which the two state variables X and Y describe the concentrations of the intermediates. We assume a constant influx rate  $v_0$ , a reversible conversion of X to Y, described with mass action kinetics with forward and backward rate constants  $k_{1p}$  and  $k_{1m}$ , respectively, and an irreversible efflux of Y described by mass action kinetics with the rate constant  $k_2$ .



We import the `modelbase` package, `numpy` and `matplotlib.pyplot`, define a list of metabolite species and a dictionary with parameters

```
import numpy as np
import matplotlib.pyplot as plt
import modelbase
cmpds = ['X', 'Y']
p = {'v0':1, 'k1p':0.5, 'k1m':1, 'k2':0.1}
```

We instantiate a model object of class `Model`

```
m = modelbase.Model(p)
```

and pass metabolites to the model (variables are always defined by names)

```
m.set_cpds(cmpds)
```

In the next step we define reaction rate functions. The rate functions always accept the model parameters as first argument, whilst the remaining arguments are metabolite concentrations.

```
v0 = lambda p: p.v0
def v1(p,x,y):
    return p.k1p*x - p.k1m*y
def v2(p,y):
    return p.k2*y
```

and then pass them to the model using `add_reaction()`

```
m.add_reaction('v0', v0, {'X':1})
m.add_reaction('v1', v1, {'X':-1, 'Y':1}, 'X', 'Y')
m.add_reaction('v2', v2, {'Y':-1}, 'Y').
```

To perform the computation we generate an instance of a simulation class using the function `Simulator()`

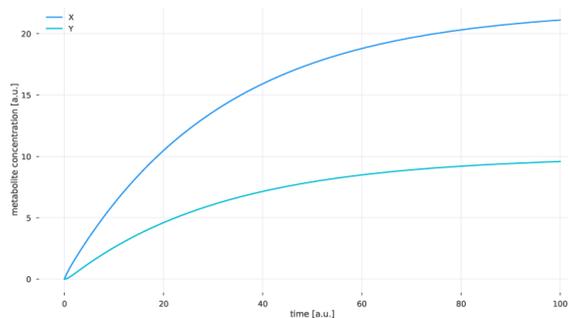
```
s = modelbase.Simulator(m)
```

To integrate the system over a given period of time ( $T=np.linspace(0,100,1000)$ ), with the initial concentrations set to 0 ( $y0=np.zeros(2)$ ), we use the method `timeCourse()`

```
s.timeCourse(T, y0)
```

Convenient access to the results of simulation through various `get*()` methods enables easy graphical display.

```
plt.figure()
plt.plot(s.getT(), s.getY())
plt.legend(m.cpdNames)
plt.show()
```



**Class *LabelModel* for isotope-labelled models**

The `modelbase` package includes a class to construct isotope-labelled versions of developed models. In order to simulate the dynamic distribution of isotopes, `modelbase` defines dynamic variables representing all possible labelling patterns for all intermediates. In contrast to instances of the class *Model*, for instances of the class *LabelModel* the number of potentially labelled atoms (usually carbon) needs to be defined for every compound. This is done with the method `add_base_cpd()`, which accepts the name and the number of labelled atoms of the compound. It automatically creates all  $2^N$  isotope variants of the compound, where  $N$  denotes the number of labelled atoms. Finally, the method `add_carbonmap_reaction()` automatically generates all isotope-specific versions of a reaction. It accepts as arguments the reaction name, rate function, carbon map, list of substrates, list of products and additional variables to be passed.

To instantiate a model object for an isotope-labelled version of developed model simply call

```
m = modelbase.LabelModel(d),
```

where `d` is again a dictionary holding parameters. With an instance of this class a dynamic process, such as the dynamic incorporation of radioactive carbon during photosynthesis, can be easily defined and simulated, using the *Simulator* class described below. An example of how to use this class is provided in **Box 2**.

**Integration methods and simulation subpackages**

Methods for the numeric integration of models are provided by the two subclasses *Simulate* and *LabelSimulate*, where the latter inherits many methods from the first. The first is used for standard kinetic models, the latter for isotope-specific models. Both classes provide computational support for dynamic simulations and methods to numerically simulate the differential equation system and to analyse the results. To provide an automatic instantiation of the correct class, we provide the function *Simulator*. Calling

```
s = modelbase.Simulator(m)
```

returns an instance of either *Simulate* or *LabelSimulate*, depending on the class of model `m`, providing all methods to numerically simulate the differential equation system and to analyse the results. Simple applications to run and plot a time course are given in boxes 1 and 2. By default, the dynamic equations are numerically integrated using a CVODE solver for stiff and non-stiff ordinary differential equation (ODE) systems. The default solver uses the Assimulo simulation package [10], with the most central solver group originating from the SUNDIALS (a Suite of Nonlinear and Differential/Algebraic equation Solvers) package [11]. If Assimulo is not available, standard integration methods from the SciPy library [12] are used. When needed, almost every integrator option can be overridden by the user by simply accessing

```
s.integrator
```

Additionally, the *Simulate* class includes methods to integrate the system until a steady-state is reached (`sim2SteadyState()`), and to estimate the period of smooth limit cycle oscillations (`estimatePeriod()`). The solution arrays are accessed with the methods `getT()`

and `getY()`. The advantage of using this method over using Assimulo's `integrator.ySol` is that `getY()` also returns the result for all the derived variables (for which algebraic modules have been used). In addition, the methods `getVarByName()`, `getVarsByName()` and `getVarsByRegExp()` allow to access the simulated values of one or several variables by their variable names or by regular expressions. Moreover, the method `getV()` gives access to the arrays of reaction rates and `getRate()` allows to access particular rates by the reaction name. The powerful Python plotting library `matplotlib` [13] provides numerous methods for graphical display of the results.

**Systems Biology Markup Language (SBML)**

`modelbase` supports export of a structural (stoichiometric) version of a created model into an XML file in the computer-readable SBML format. To export the model (`m`) simply use the method `m.ModelbaseToSBML(filename)`. A minimal working example can be found in our repository ([https://gitlab.com/ebenhoe/h/modelbase/blob/master/examples/sbml\\_export.py](https://gitlab.com/ebenhoe/h/modelbase/blob/master/examples/sbml_export.py)). Structural and stoichiometric analyses are currently not implemented in `modelbase`, therefore such export allows to take advantage of other SBML compatible modelling environments developed for such tasks (e.g. PySCeS or CobraPy [14]). The import of SBML models into `modelbase` is currently not supported, mainly because of the complementary purpose for which it was developed. The `modelbase` framework simplifies construction of kinetic models, allowing to perform this task with minimal modelling experience. Therefore, the main purpose of `modelbase` is the model design process itself, rather than importing a predefined construct to perform complex computations. However, a full SBML export and import functionality is currently under development to allow model sharing across different environments and platforms.

**Quality control**

`modelbase` has been continuously developed and used within our lab since 2016. It has been successfully applied to study the complexity of photosynthesis and carbon assimilation in plants [7] and is being further maintained and developed.

**(2) Availability****Operating system**

`modelbase` is compatible with all platforms with working Python distribution.

**Programming language**

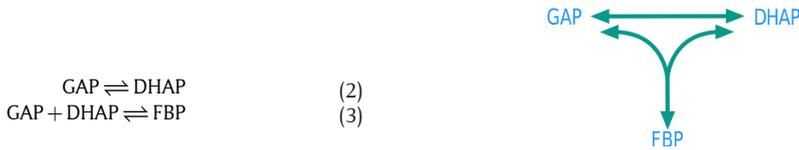
`modelbase` is written in the Python programming language, a general-purpose interpreted, interactive, object-oriented, and high-level programming language. It is available for every major operating system, including GNU/Linux, Mac OSX and Windows and has been tested with Python 3.6.

**Additional system requirements**

None

**Box 2: Isotope-labelled model**

A minimal example of an isotope-label specific model simulates equilibration of isotope distribution in a system consisting of the two reactions of triose-phosphate isomerase and fructose-bisphosphate aldolase:



We import the modelbase package, numpy and matplotlib.pyplot.

```
import numpy as np
import matplotlib.pyplot as plt
import modelbase
```

We define a dictionary of parameters and instantiate the model of class *LabelModel*

```
p = {'kf_TPI': 1, 'Keq_TPI': 21, 'kf_Ald': 2000, 'Keq_Ald': 7000}
m = modelbase.LabelModel(p).
```

Compounds are added with an additional argument defining the numbers of carbons

```
m.add_base_cpd('GAP', 3)
m.add_base_cpd('DHAP', 3)
m.add_base_cpd('FBP', 6)
```

leading to an automatic generation of  $80 = 2^6 + 2^3 + 2^3$  isotope-specific compounds. All reactions are assumed to obey mass-action rate laws. Standard rate laws are defined in the `modelbase.ratelaws` module. Due to simplicity, the following steps are only shown for the forward triose-phosphate isomerase reaction. For more details please see the file `examples/isotopeLabels.py` in the modelbase package.

```
import modelbase.ratelaws as rl
def vlf(p,y):
    return rl.massAction(p.kf_TPI,y)
```

All isotope-specific rates are generated by the `add_carbonmap_reaction()` method, based on a list defining in which positions the carbons appear in the products.

```
m.add_carbonmap_reaction('TPIf',vlf,[2,1,0],['DHAP'],'GAP')
```

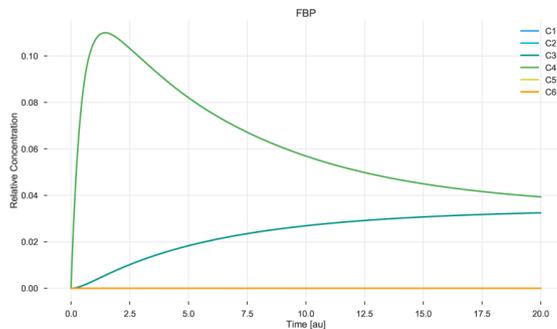
We set the initial conditions such that the total pools are in equilibrium, but carbon 1 of GAP is fully labeled

```
GAP0 = 2.5e-5
DHAP0 = GAP0 * m.par.Keq_TPI
y0d = {'GAP': GAP0,
       'DHAP': DHAP0,
       'FBP': GAP0 * DHAP0 * m.par.Keq_Ald}
y0 = m.set_initconc_cpd_labelpos(y0d,labelpos={'GAP':0})
```

and simulate equilibration of the labels for 20 arbitrary time units

```
s = modelbase.LabelSimulate(m)
T = np.linspace(0,20,1000)
s.timeCourse(T,y0).
```

We plot the result using the `getLabelAtPos()` method (see `examples/isotopeLabels.py`).



**Dependencies**

Dependencies are provided in the `setup.py` file and include:

- `numpy == 1.14.3`
- `scipy == 1.1.0`
- `numdifftools == 0.9.20`
- `assimulo == 2.9`
- `pandas == 0.22.0`
- `python-libsbnl == 5.17.0`

Support for the differential equation solver sundials (CVODE) through the python package `assimulo` requires moreover:

- `Sundials-2.6.0` (for 64bits machines, install Sundials using `-fPIC`)
- `Cython 0.18`
- C compiler
- Fortran compiler

The detailed instruction how to install the prerequisites is included in the repository in our installation guide.

**List of contributors**

In alphabetic order: Marvin van Aalst, Oliver Ebenhöh, Anna Matuszyńska, Nima P. Saadat.

**Software location****Archive**

**Name:** Python Package Index (PyPI)

**Persistent identifier:** <https://pypi.org/project/modelbase/>

**Licence:** GPL3

**Publisher:** Oliver Ebenhöh

**Version published:** 0.2.5

**Date published:** 09/10/18

**Code repository**

**Name:** GitLab

**Persistent identifier:** <https://gitlab.com/ebenhoeh/modelbase>

**Licence:** GNU General Public License v3.0

**Date published:** 09/10/18

**Language**

`modelbase` was entirely developed in English.

**(3) Reuse potential**

The strength of our package lies in its flexibility to be applied to simulate and analyse various distinct biological systems. It can be as efficiently used for the development of new models, as for the reconstruction of existing ones. Here, we demonstrate its power by reimplementing three mathematical models that have been previously published without providing the source code (**Table 1**). This includes i) a model of the photosynthetic electron transport chain (PETC) used to model photoprotective mechanisms in plants and green algae, originating from our lab and initially developed in MATLAB [15]; ii) a model of the Calvin-Benson-Bassham (CBB) Cycle by Poolman *et al.* [16], developed to study the dynamics of the carbon assimilation and iii) a model of the Pentose phosphate pathway (PPP), adapted by Berthon *et al.* [17] to investigate label distribution dynamics in isotope labelling experiments.

**Modelling the PETC to study photoprotective mechanisms**

Part of our research focuses on understanding the dynamics of various photoprotective mechanisms present in photosynthetic organisms [18, 15, 19]. The foundation of our further work constitutes the model of the photosynthetic electron transport chain in green algae *Chlamydomonas reinhardtii* published in 2014 [15]. We have reimplemented the original work in Python and reproduced the results published in the main text (**Figure 2**), providing a photosynthetic electron transport chain core model, compatible with other `modelbase`-adapted modules, to further our studies on the dynamics of light reactions of photosynthesis.

**CBB Cycle and the dynamics of carbon assimilation**

Using `modelbase`, we have reimplemented a model of the CBB Cycle by Poolman *et al.* [16]. The model is a variant of the Pettersson and Ryde-Pettersson [3] model, where the strict rapid equilibrium assumption is relaxed and fast reactions are modelled by simple mass action kinetics. Its main purpose is to study short to medium time scale responses to changes in extra-stromal phosphate concentration and incident light. The concentrations of NADPH, NADP<sup>+</sup>, CO<sub>2</sub> and H<sup>+</sup> are considered constant, leaving the 13 CBB cycle intermediates, ATP, ADP and inorganic phosphate as dynamic variables. The model further incorporates a simplified starch production using glucose 6-phosphate and glucose-1-phosphate and a

**Table 1:** Mathematical models originally published without the source-code, reconstructed in our lab using the `modelbase` package. The source code and examples are available from the GitHub repository of our lab <https://github.com/QTB-HHU/>.

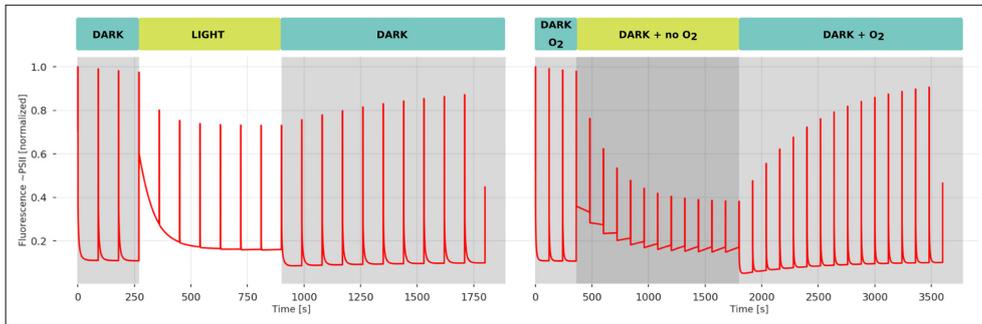
Process	Original publication	GitHub.com/ QTB-HHU/	Developer
Photosynthetic Electron Transport Chain	[15]	<code>./petc-modelbase</code>	A.M.
Calvin-Benson-Bassham Cycle	[16]	<code>./cbb-modelbase</code>	M.v.A.
Pentose Phosphate Pathway	[4, 17]	<code>./ppp-modelbase</code>	T.N.

simple ATP recovery reaction. We used the `modelbase` implementation of the Poolman model to simulate the steady state concentrations of the metabolites depending on the extra-stromal phosphate concentration (Figure 3), reproducing original work by Pettersson and Ryde-Pettersson [3]. We have observed that the system is not stable any more for  $[P_{ext}] > 1.5$ , a feature not discussed in the Poolman paper [16].

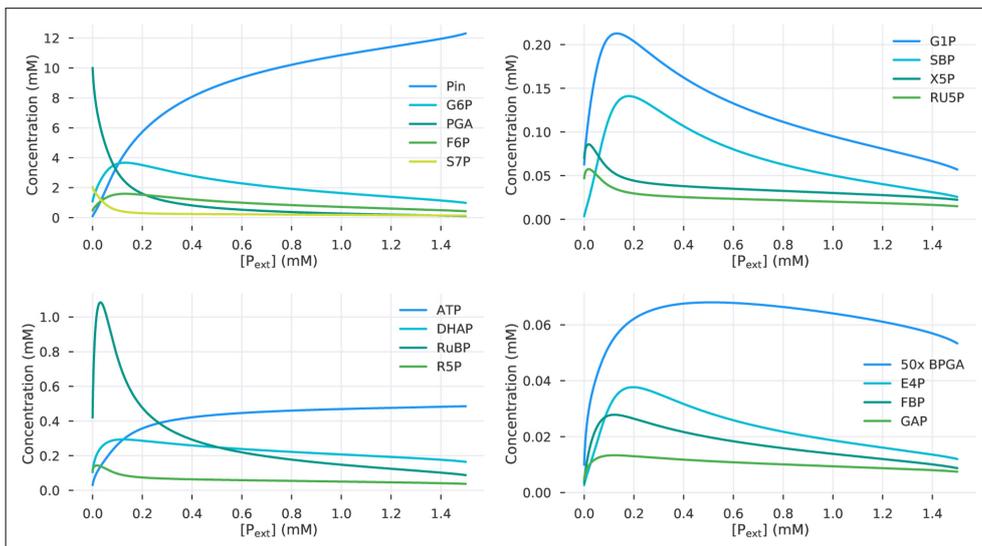
The compatible mathematical representation of the two photosynthetic subsystems, the ATP-producing light reactions and the ATP-consuming CBB cycle, is a prerequisite to merge those two models. Technically, in the `modelbase` framework, this is a straight forward process. Scientifically, it turned out to be not a trivial task (unpublished work).

### PPP and isotope labelling experiments

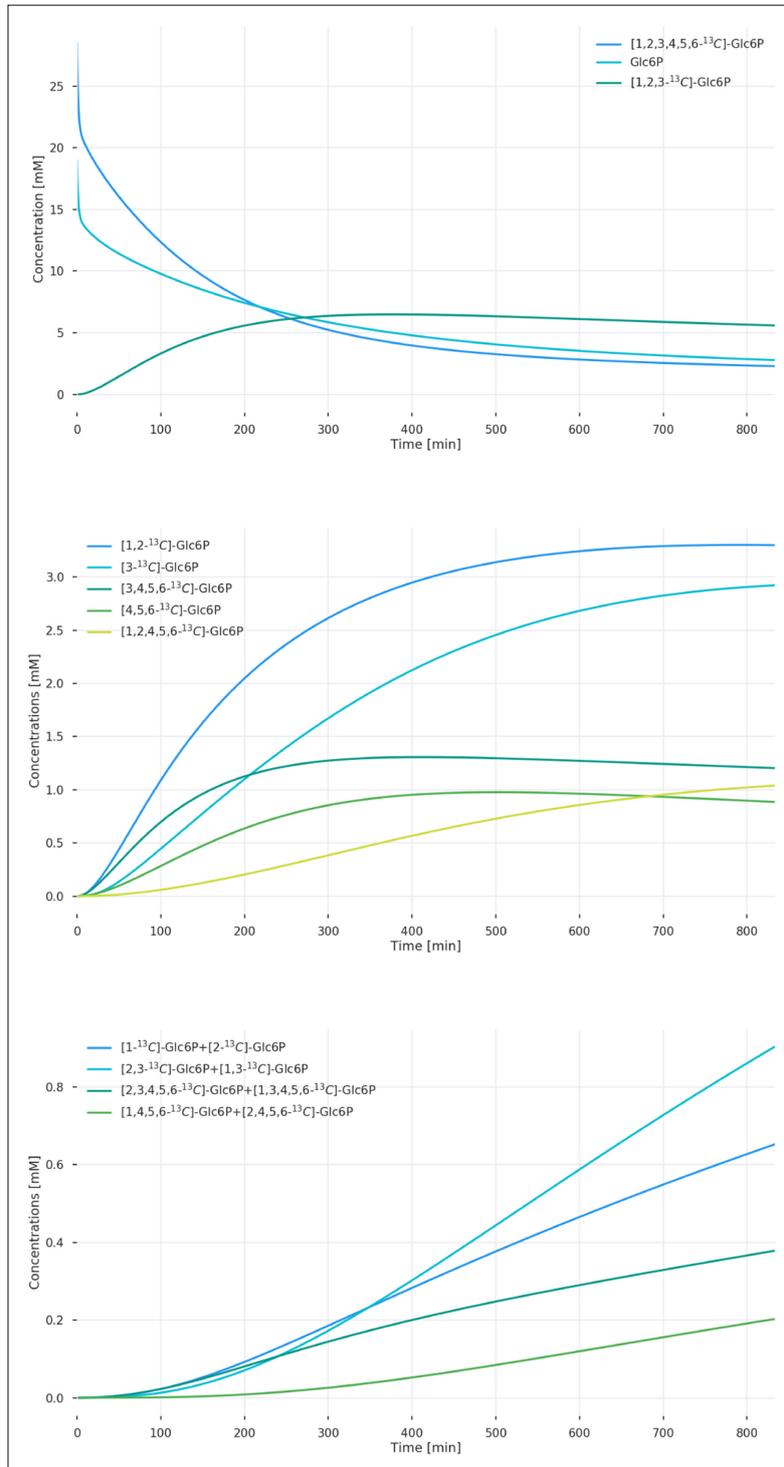
We envisage that especially our `LabelModel` extension will find a wide application in metabolic network analysis. Radioactive and stable isotope labelling experiments constitute a powerful methodology for estimating metabolic fluxes and have a long history of application in biological research [20]. Here, we showcase the potential of `modelbase` for the isotope-labelled experiments by reimplementing the model of the F-type non-oxidative PPP in erythrocytes originally proposed by McIntyre *et al.* [4]. This was later adapted by Berthon *et al.* for label experiments and *insilico* replication of  $^{13}\text{C}$  nuclear magnetic resonance (NMR) studies [17]. We have reproduced the results obtained by the authors, including the time course of diverse Glucose-6-phosphate isotopomers (Figure 4).



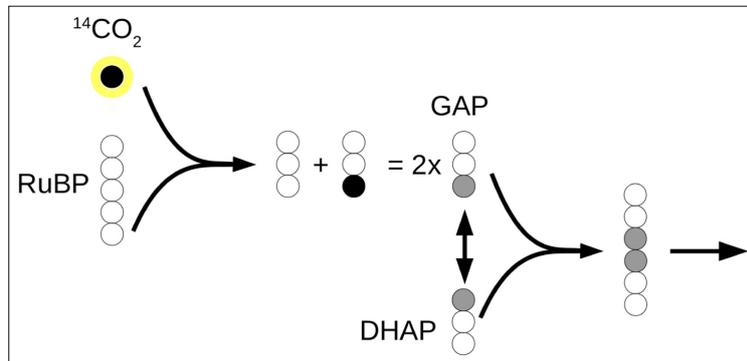
**Figure 2:** Reproduction of the Figures from [15]. Simulated fluorescence trace obtained through Pulse Amplitude Modulation (PAM) under light induced (left) and anoxia induced (right) conditions. The dynamics of the fluorescence decrease corresponds to the activation of a specific photoprotective mechanism called state transitions, while the increase in the signal after the inducer (light or anoxia) is switched off relates to the relaxation of the mechanism.



**Figure 3:** Metabolite steady state concentrations dependent on the extra-stromal phosphate concentration simulated with the Poolman implementation of the Pettersson and Ryde-Pettersson model of the CBB cycle [16].



**Figure 4:** Formation of diverse Glc6P isotopomers in a haemolysate, obtained by solving the adapted model by Berthon *et al.* [17] reimplemented using modelbase.



**Figure 5:** Schematic representation of label incorporation by the CBB cycle intermediates.

#### Other possible applications

Among many other applications, `modelbase` provides tools to reproduce the ‘photosynthetic Gibbs effect’. Gibbs and Kandler described it in 1956 and 1957 [8, 9], when they observed the atypical and asymmetrical incorporation of radioactive  $^{14}\text{CO}_2$  in hexoses. An example of label incorporation by the CBB cycle intermediates is presented schematically in **Figure 5**.

Finally, our package provides a solid foundation for additional extensions to the framework architecture, its classes and modelling routines. To encourage its use and to facilitate the first steps to apply the `modelbase` package, we have prepared an interactive tutorial using a Jupyter Notebook [21], which showcases basic implementation of `modelbase` and each of its classes in easy to follow and thoroughly explained examples (see <https://gitlab.com/ebenhoh/modelbase/blob/master/Tutorial.ipynb>).

#### Abbreviations

**CBB** Calvin-Benson-Bassham; **NMR** Nuclear Magnetic Resonance; **ODE** Ordinary Differential Equations; **PAM** Pulse Amplitude Modulation; **PPP** Pentose Phosphate Pathway; **QSSA** Quasi Steady-State Approximation; **SBML** Systems Biology Markup Language; **UML** Unified Modeling Language.

#### Acknowledgements

We would like to thank the students working on their Bachelor and Masters projects in our lab, who applied and tested this software while investigating their scientific problems.

#### Competing Interests

The authors have no competing interests to declare.

#### Author Informations

- Initiated the project, developed the code and provided teaching examples.
- Developed further the code, prepared the documentation and battery of tests and reimplemented the Calvin-Benson-Bassham Cycle model as an example of `modelbase` utility.
- Provided export support for SBML models.

- Reimplemented the Pentose-Phosphate-Pathway model as an example of `modelbase` utility.
- Developed further the code, prepared the Jupyter Notebook with the tutorial, provided an example of `modelbase` utility and wrote the first draft of the manuscript.

All authors have read the manuscript and contributed to its final version.

#### References

1. **Almquist, J, Cvijovic, M, Hatzimanikatis, V, Nielsen, J and Jirstrand, M** 2014 Kinetic models in industrial biotechnology – Improving cell factory performance. *Metabolic Engineering*, 24: 38–60. DOI: <https://doi.org/10.1016/j.ymben.2014.03.007>
2. **Oliphant, T E** 2007 Python for scientific computing. *Computing in Science and Engineering*, 9(3): 10–20. DOI: <https://doi.org/10.1109/MCSE.2007.58>
3. **Petterson, G and Ryde-Petterson, U** 1988 A mathematical model of the Calvin photosynthesis cycle. *Eur. J. Biochem*, 175(3): 661–672. DOI: <https://doi.org/10.1111/j.1432-1033.1988.tb14242.x>
4. **McIntyre, L M, Thorburn, D R, Bubb, W A and Kuchel, P W** 1989 Comparison of computer simulations of the F-type and L-type non-oxidative hexose monophosphate shunts with  $^{31}\text{P}$ -NMR experimental data from human erythrocytes. *Eur J Biochem*, 180(2): 399–420. DOI: <https://doi.org/10.1111/j.1432-1033.1989.tb14662.x>
5. **Poolman, M G** 2006 ScrumPy: Metabolic modelling with Python. *IEE Proc.-Syst. Biol*, 153(5): 375–378. DOI: <https://doi.org/10.1049/ip-syb:20060010>
6. **Olivier, B G, Rohwer, J M and Hofmeyr, J H S** 2005 Modelling cellular systems with PySCeS. *Bioinformatics*, 21(4): 560–561. DOI: <https://doi.org/10.1093/bioinformatics/bti046>
7. **Ebenhöh, O and Spelberg, S** 2018 The importance of the photosynthetic Gibbs effect in the elucidation of the Calvin-Benson-Bassham cycle. *Biochemical Society Transactions*. DOI: <https://doi.org/10.1042/BST20170245>

8. **Kandler, O** and **Gibbs, M** 1956 Asymmetric Distribution of  $C^{14}$  in the Glucose Phosphates Formed During Photosynthesis. *Plant physiology*, 31(5): 411. DOI: <https://doi.org/10.1104/pp.31.5.411>
9. **Gibbs, M** and **Kandler, O** 1957 Asymmetric distribution of  $C^{14}$  in sugars formed during photosynthesis. *Proc. Natl. Acad. Sci. U.S.A.*, 43: 446–451. DOI: <https://doi.org/10.1073/pnas.43.6.446>
10. **Andersson, C**, **Claus, F** and **Akesson, J** 2015 ScienceDirect Assimulo: A unified framework for ODE solvers. *Mathematics and Computers in Simulation*, 116: 26–43. DOI: <https://doi.org/10.1016/j.matcom.2015.04.007>
11. **Hindmarsh, A C**, **Brown, P N**, **Grant, K E**, **Lee, S L**, **Serban, R**, **Shumaker, D E** and **Woodward, C S** 2005 SUNDIALS: Suite of Nonlinear and Differential/Algebraic Equation Solvers. *ACM Transactions on Mathematical Software*, 31(3): 363–396. DOI: <https://doi.org/10.1145/1089014.1089020>
12. **Jones, E**, **Oliphant, T**, **Peterson, P**, et al. 2001 Scipy: Open source scientific tools for python.
13. **Hunter, J D** 2007 Matplotlib: A 2D graphics environment. *Computing In Science and Engineering*, 9(3): 90–95. DOI: <https://doi.org/10.1109/MCSE.2007.55>
14. **Ebrahim, A**, **Lerman, J A**, **Palsson, B O** and **Hyduke, D R** 2013 COBRApy: COntstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, 7(1): 1. DOI: <https://doi.org/10.1186/1752-0509-7-74>
15. **Ebenhöh, O**, **Fucile, G**, **Finazzi, G G**, **Rochaix, J-D** and **Goldschmidt-Clermont, M** 2014 Short-term acclimation of the photosynthetic electron transfer chain to changing light: A mathematical model. *Philosophical Transactions B*, 369(1640). DOI: <https://doi.org/10.1098/rstb.2013.0223>
16. **Poolman, M G**, **Fell, D A** and **Thomas, S** 2000 Modelling photosynthesis and its control. *Journal of Experimental Botany*, 51(90001): 319–328. DOI: [https://doi.org/10.1093/jexbot/51.suppl\\_1.319](https://doi.org/10.1093/jexbot/51.suppl_1.319)
17. **Berthon, H A**, **Bubb, W A** and **Kuchel, P W** 1993  $^{13}C$  n.m.r. isotopomer and computer-simulation studies of the non-oxidative pentose phosphate pathway of human erythrocytes. *Biochem J*, 296: 379–387. DOI: <https://doi.org/10.1042/bj2960379>
18. **Ebenhöh, O**, **Houwaart, T**, **Lokstein, H**, **Schlede, S** and **Tirok, K** 2011 A minimal mathematical model of nonphotochemical quenching of chlorophyll fluorescence. *Biosystems*, 103(2): 196–204. DOI: <https://doi.org/10.1016/j.biosystems.2010.10.011>
19. **Matuszyńska, A B**, **Heidari, S**, **Jahns, P** and **Ebenhöh, O** 2016 A mathematical model of non-photochemical quenching to study short-term light memory in plants. *Biochimica et Biophysica Acta (BBA) – Bioenergetics*, 1857(12): 1–7. DOI: <https://doi.org/10.1016/j.bbabi.2016.09.003>
20. **Crown, S B**, **Ahn, W** and **Antoniewicz, M R** 2012 Rational design of  $^{13}C$ -labeling experiments for metabolic flux analysis in mammalian cells. *BMC Systems Biology*, 6(1): 43. DOI: <https://doi.org/10.1186/1752-0509-6-43>
21. **Kluyver, T**, **Ragan-Kelley, B**, **Perez, F**, **Granger, B**, **Buisonnier, M**, **Frederic, J**, **Kelley, K**, **Hamrick, J**, **Grout, J**, **Corlay, S**, **Ivanov, P**, **Avila, D**, **Abdalla, S** and **Willing, C** 2016 Jupyter Notebooks – a publishing format for reproducible computational workflows. In: Schmidt, B and Loizides, F (eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, 87–90. IOS Press.

**How to cite this article:** Ebenhöh, O, van Aalst, M, Saadat, N P, Nies, T and Matuszyńska, A 2018 Building Mathematical Models of Biological Systems with mode1base. *Journal of Open Research Software*, 6: 24. DOI: <https://doi.org/10.5334/jors.236>

**Submitted:** 27 June 2018

**Accepted:** 16 October 2018

**Published:** 16 November 2018

**Copyright:** © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Journal of Open Research Software* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 

## **6.2 Network Reconstruction and modeling Made Reproducible with moped**

### **6.2.1 Authors and Details**

Authors: Nima P. Saadat, Marvin v. Aalst, Oliver Ebenhöf  
Authorship: 1. Author (shared)  
Journal: Metabolites  
Status: Published (<https://doi.org/10.3390/metabo12040275>)

### **6.2.2 Contributions**

Nima P. Saadat developed the concept and initial project, developed the initial code, prepared all examples and tests, prepared the original manuscript draft and helped improving the manuscript after review. Marvin van Aalst further improved the code and the functionality, prepared examples and tests and helped improving the manuscript after review. Oliver Ebenhöf helped in concept development and helped in improving and editing the manuscript.

Article

# Network Reconstruction and Modelling Made Reproducible with moped

Nima P. Saadat <sup>1,†</sup> , Marvin van Aalst <sup>1,†</sup>  and Oliver Ebenhöf <sup>1,2,\*</sup> 

<sup>1</sup> Institute of Quantitative and Theoretical Biology, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany; nima.saadat@hhu.de (N.P.S.); marvin.van.aalst@hhu.de (M.v.A.)

<sup>2</sup> Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

\* Correspondence: oliver.ebenhoef@hhu.de

† These authors contributed equally to this work.

**Abstract:** Mathematical modeling of metabolic networks is a powerful approach to investigate the underlying principles of metabolism and growth. Such approaches include, among others, differential-equation-based modeling of metabolic systems, constraint-based modeling and metabolic network expansion of metabolic networks. Most of these methods are well established and are implemented in numerous software packages, but these are scattered between different programming languages, packages and syntaxes. This complicates establishing straight forward pipelines integrating model construction and simulation. We present a Python package moped that serves as an integrative hub for reproducible construction, modification, curation and analysis of metabolic models. moped supports draft reconstruction of models directly from genome/proteome sequences and pathway/genome databases utilizing GPR annotations, providing a completely reproducible model construction and curation process within executable Python scripts. Alternatively, existing models published in SBML format can be easily imported. Models are represented as Python objects, for which a wide spectrum of easy-to-use modification and analysis methods exist. The model structure can be manually altered by adding, removing or modifying reactions, and gap-filling reactions can be found and inspected. This greatly supports the development of draft models, as well as the curation and testing of models. Moreover, moped provides several analysis methods, in particular including the calculation of biosynthetic capacities using metabolic network expansion. The integration with other Python-based tools is facilitated through various model export options. For example, a model can be directly converted into a CobraPy object for constraint-based analyses. moped is a fully documented and expandable Python package. We demonstrate the capability to serve as a hub for integrating reproducible model construction and curation, database import, metabolic network expansion and export for constraint-based analyses.

**Keywords:** metabolic networks; modeling; topological networks; metabolic network expansion; network reconstruction



**Citation:** Saadat, N.P.; van Aalst, M.; Ebenhöf, O. Network Reconstruction and Modelling Made Reproducible with moped. *Metabolites* **2022**, *12*, 275. <https://doi.org/10.3390/metabo12040275>

Academic Editor: Oscar Dias, Hunter N. B. Moseley and Miguel Rocha

Received: 14 December 2021

Accepted: 15 March 2022

Published: 22 March 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Theoretical analysis of metabolic pathways has a longstanding tradition. The early approaches to study glycolysis, for example, have considerably increased our understanding of fundamental regulatory principles in metabolism [1]. In recent approaches, metabolic modeling was employed to study metabolic interdependencies in microbial communities and to identify putative drug targets for microbial pathogens [2,3].

Several theoretical techniques to study metabolism have been established. The most classic technique is the analysis of metabolic networks by representing them as systems of ordinary differential equations (ODEs). This representation heavily depends on detailed knowledge of stoichiometries, parameters of enzyme kinetics and regulatory mechanisms of reactions [4]. This approach is extremely useful for investigating relatively small systems.

The upsurge of novel high-throughput experimental “omics” techniques led to the collection of immense amounts of data, resulting in an ever-increasing number of fully sequenced genomes. The improved quality of annotated genes resulted in a tremendous increase in information of enzymes and the respective metabolic reactions. This information has been collected in biochemical databases such as MetaCyc, BioCyc, KEGG or BiGG [5–8]. Such databases provide information for large-scale metabolic networks of many different organisms. However, analyzing such large-scale metabolic networks using systems of ordinary differential equations is difficult. This is, to a large extent, due to missing information on kinetic parameters of the involved enzymatic reactions [9]. One convenient alternative is constraint-based modeling and its mathematical method flux balance analysis (FBA) [10]. This commonly used approach uses the stoichiometric matrix of a reaction network and finds a steady-state vector of reaction fluxes that maximizes or minimizes an objective function that linearly depends on the reaction rates. Other structural analysis techniques focus on the topology of metabolic networks [11]. One such technique is metabolic network expansion and the related concept of metabolic scopes. The metabolic scope describes the set of metabolites, which are topologically producible by a given network from an initial set of compounds [12,13]. Thus, metabolic network expansion allows to functionally characterize metabolic networks with respect to their biosynthetic capacities [14].

Topological techniques are extremely useful in the process of curating models, in particular to identify and add missing reactions [15]. This process, called gap filling, allows, for example, to complement draft metabolic networks in order to guarantee that observed compounds can be produced from the growth medium [16].

Many of the techniques described above have been implemented as Python packages. However, most of these software packages are not directly compatible with each other.

In this work, we present *moped*, a compact but useful Python package that serves as a hub, offering tools for analysis, development and extension or modification of metabolic models. The integration of BLAST and pathway/genome databases such as MetaCyc and BioCyc into *moped* allows reconstructing metabolic network models directly from genome sequences [17] and ensures that the reconstruction process is fully transparent and reproducible. In addition to the *de novo* construction of models, *moped* provides an interface to import existing metabolic network models in SBML format.

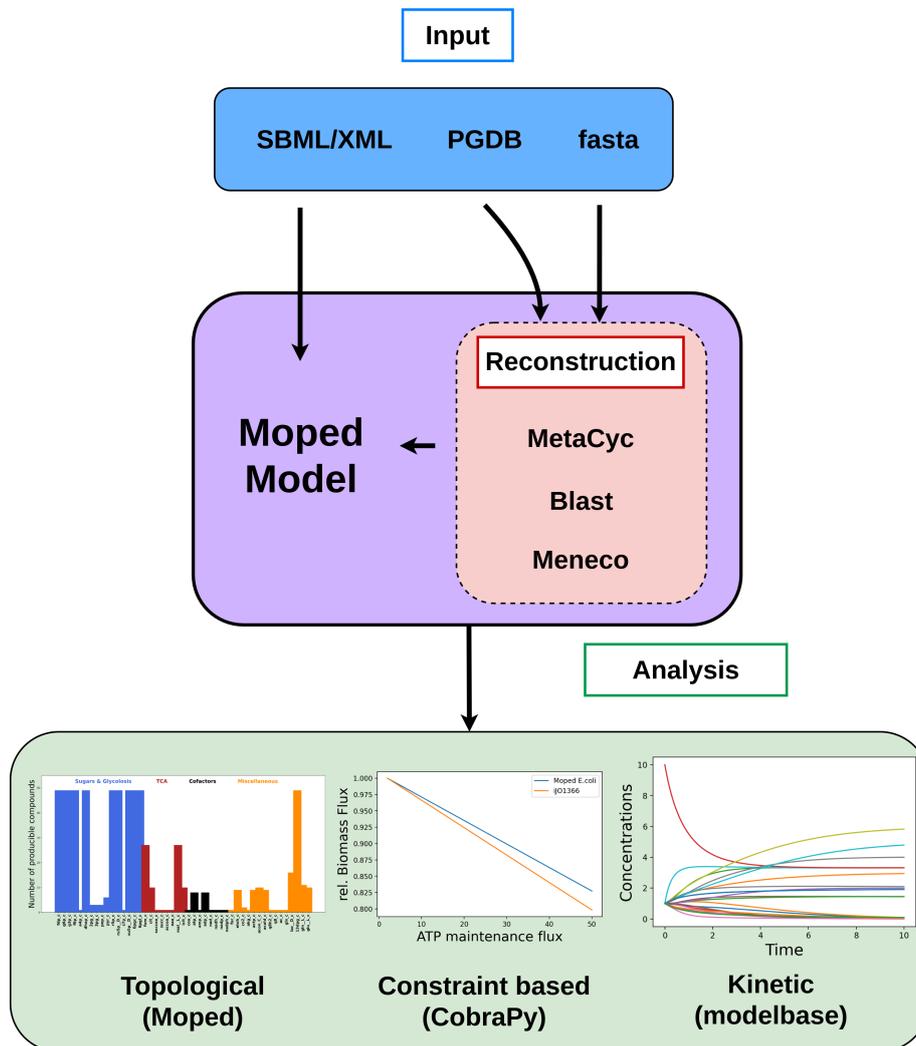
To facilitate curation of metabolic models, *moped* provides an interface to *Meneco*, a topological gap-filling tool based on answer set programming [18]. All models created with *moped* can easily be exported as *CobraPy* objects, thus directly integrating constraint-based with model construction and modification [19]. It is even possible to extract a scaffold model of metabolic pathways for kinetic modeling via *modelbase* [20]. The Python package *moped* presented here is a mathematical modeling hub, which allows constructing reproducible metabolic models *de novo*, integrating existing models in SBML format, curating models by gap filling and performing topological or constraint-based analyses.

## 2. Implementation

### 2.1. Model Import, Extension and Modification

*moped* uses SBML files or PGDB flat files as input for constructing a metabolic network model. PGDBs are organism-specific pathway/genome databases containing annotated reactions and compounds of the metabolism of the organism [6]. These databases further include detailed information about reactions and compounds, such as sum formulas, charges, references to other database entries or subcellular localization. This information is of great importance for a consistent analysis of metabolic networks. SBML files represent metabolic networks in an XML-based format and can be considered as a standard for the exchange of reconstructed and curated metabolic models between tools and platforms [21]. Such files can be, among others, obtained from databases such as BiGG, which provides SBML files of curated metabolic models together with information about the corresponding publications [7].

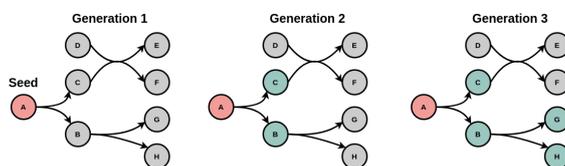
Because of the wide range of import methods (FASTA, PGDBs and SBML), one particular strength of moped is the integration of several analysis tools. An overview of moped's functionalities is shown in Figure 1. Furthermore, moped provides an accessible environment to extend or modify constructed or imported models. Therefore, adding alternative or additional metabolic pathways to pre-existing models, as well as modifying compound and reaction identifiers, is simple and straightforward. Naturally, all moped objects can be exported as SBML. A UML diagram of moped can be found in Figure S1 in the Supplementary Material.



**Figure 1.** The modeling hub moped. moped accepts SBML, FASTA files or MetaCyc and BioCyc PGDBs as inputs. PGDBs and SBML files are directly converted into a moped object. By BLASTing genome/proteome-sequences against MetaCyc, moped models can be constructed utilizing GPR rules. Further reconstruction can be achieved using Meneco for gap filling. Topological model analysis is implemented in moped. For constraint-based and kinetic analysis, moped offers export as CobraPy and modelbase objects, respectively [19,20].

## 2.2. Tools for Metabolic Network Expansion

A useful and valuable feature of moped is the fully implemented network expansion algorithm to perform metabolic network expansions on moped objects. Metabolic network expansion can be used to investigate structural properties of metabolic networks, such as biosynthetic capacities and their robustness against structural perturbations [12]. The core concept of metabolic network expansion is the metabolic scope, which contains all compounds that are producible by a network from a given initial set of compounds, termed the seed (see Figure 2). In the expansion process, the seed is used to find all reactions that can proceed in their annotated direction. The respective products are then added to the seed, forming the new seed for the next expansion step. This process continues until no new compounds can be added to the seed. Thus, scopes characterize biosynthetic capacities of metabolic networks, based exclusively on their topology.



**Figure 2.** Metabolic network expansion. Beginning with an initial set of compounds, the seed (here panel 1), the expansion process detects all producible compounds in a network and adds them to the seed for the next generation until no additional producible compounds are found.

Network expansion depends on a precise definition of reaction reversibilities and involved cofactors. Network expansion uses the stoichiometry of reactions to identify producible compounds. However, stoichiometric coefficients of reactions are annotated for one particular direction. To include the opposite direction (for reversible reactions) into the metabolic network expansion, moped provides a method for reversibility duplication. As illustrated in Figure 3 for triose-phosphoisomerase as an example, this method finds all reversible reactions in a moped object and adds the reversed reaction to the network. The new reaction identifier is identical to the identifier of the original reaction concatenated with the suffix ‘\_rev\_’. This model modification can be reverted if no longer needed.

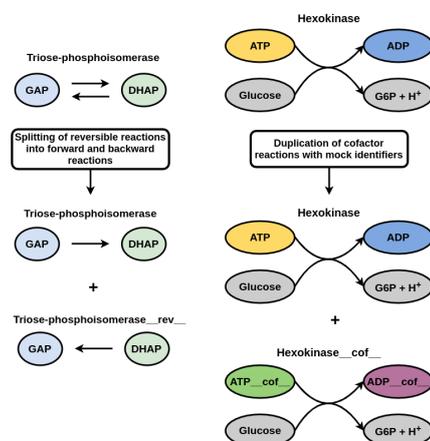
Many reactions depend on specific cofactors. Cofactors usually appear in pairs. One of the most prominent examples is the cofactor pair ATP and ADP. In the majority of reactions with ATP as substrate, ATP serves as a donor of a phosphate group, thus producing ADP. Only a few reactions actually modify the adenosine moiety (for example, in nucleotide de novo synthesis). In network expansion, therefore, no reaction utilizing ATP or ADP as cofactor could proceed, unless these compounds are either included in the seed or can be produced from metabolites within the seed. If the purpose of network expansion is to realistically calculate a set of producible compounds, this behavior is not desired because it leads to a drastic underestimation of the scope. The most naive approach to directly include cofactors in the seed yields misleading results, because in such a case, all compounds that can be generated from digesting, e.g., ATP would be included in the scope.

A pragmatic approach to solve this problem is the duplication of cofactors as proposed in [12]. Here, reactions with cofactor pairs are duplicated, where the copied reactions contain “mock cofactors”. In contrast to the real cofactors, the mock cofactors only occur in reactions, in which they act in their role as cofactors. For ATP, this is the transfer of a phosphate group, for NADH or NADPH the transfer of protons and electrons and for acetyl coenzyme-A, the transfer of the acetyl group. The cofactor duplication allows the use of mock cofactors inside the initial seed. Reactions depending on cofactors might now be able to occur in the expansion process. However, reactions using the cofactors as proper substrates can only occur if the real cofactor can be produced from the seed.

moped provides a convenient method for finding and duplicating all reactions using cofactor pairs. The cofactor pairs can either be automatically determined by moped for

networks imported from BiGG or MetaCyc (see Table S1 and S2 in the Supplementary Material) or they can be declared individually by the user. The identifiers of the duplicated cofactors are replaced by mock identifiers, which contain the suffix ‘\_cof\_’. The same modification is applied to the respective reaction identifiers. This model modification can be reverted if no longer needed.

The implemented methods for cofactor and reversibility duplication are commonly used to obtain biologically meaningful results for metabolic network expansion. However, they are also highly useful for topological gap-filling using Meneco, during model reconstruction. This is further explained in the next section.



**Figure 3.** Topological network modifications moped offers functions for splitting reversible reactions into forward and backward reactions in a network. Adding a copy of each cofactor dependent reaction and replacing cofactors (here ATP and ADP) with mock identifiers allows unblocking cofactor dependent reactions while avoiding degradation products of cofactors contained in the seed. Such modifications enable biologically feasible metabolic network expansion.

### 2.3. Reconstruction of Draft Network Models

Construction of metabolic networks highly depends on reliable databases. In order to enable user-friendly metabolic network reconstruction, moped includes methods for importing data from the MetaCyc and BioCyc databases, identifying homologous sets of genes with BLAST and gap-filling.

MetaCyc is a universal, highly curated reference database of metabolic pathways and biochemical reactions from all domains of life. BioCyc is a database of organism-specific PGDBs containing metabolic network information based on predictions by the PathoLogic component of the Pathway Tools software [22,23]. The MetaCyc and BioCyc databases provide many advantages. Both databases are freely available for academic and nonprofit users. All PGDBs are available in useful flat file formats. Furthermore, these databases include information on the reaction directions based on experimental references and thermodynamics, extensive annotations and, therefore, information about gene–protein–reaction (GPR) associations, as well as thermodynamic information about metabolites and reactions such as the Gibbs energy of formation and the standard Gibbs energy of reactions.

In order to use BioCyc and MetaCyc for metabolic network construction and analysis, moped offers a parser for PGDBs, allowing direct construction of moped objects from MetaCyc or BioCyc flat files. moped objects can directly be used for network analyses including network expansion and constraint-based modeling. Especially for the latter, it is extremely important that all reactions are mass- and charge-balanced to ensure that all solutions obey mass conservation. Therefore, only reactions which are mass- and charge-

balanced are parsed in moped. While this process has the danger of omitting annotated genes, including reactions that are not mass- or charge-balanced would violate fundamental physical principles and lead to unrealistic model properties. This pipeline provided by the database import and parsing of moped makes it straightforward to construct prokaryotic network models. For eukaryotic metabolic networks, however, intensive and careful curation is required due to missing compartment information. More detailed information about the parsing of PGDBs using moped can be found in the documentation.

There exist several pipelines to automatically extract a set of metabolic reactions from a genome or proteome sequence. One popular pipeline is the above mentioned PathoLogic software. moped integrates such a pipeline into the Python programming language, directly converting a genome/proteome sequence into a moped object that can be immediately used for modeling applications. This functionality is provided by an implemented wrapper for local BLAST to find enzyme reactions in genome sequence fasta files or proteome fasta files by similarity search against enzyme reference sequences from the MetaCyc database. This method constructs a new moped object representing a metabolic network of all reactions that are found in a genome sequence or proteome using enzyme monomer amino acid sequences and protein–reaction annotations from MetaCyc to ensure fulfilled gene–protein–reaction associations (GPRs) in all found reactions [24]. This process can be controlled by user-defined thresholds. This integrated pipeline makes the model reconstruction perfectly reproducible and illustrates the functionality of moped as a modeling hub.

The next curation step after the initial automatic network construction is usually gap-filling. This describes a process in which reactions are added to the network in order to ensure that the reconstructed model reflects experimentally observed behavior, such as the production of experimentally measured compounds from the growth medium [25]. There are many available gap-filling methods such as GapFill or MIRAGE [26,27]. Most of these methods are based on constraint-based approaches. A common problem is that these approaches can predict gap-filling solutions that use thermodynamically infeasible cycles. In this sense, these approaches are sensitive to self-producing or energy-generating cycles [18]. Meneco, in contrast, is a topological gap-filling tool based on the network expansion algorithm. Meneco calculates a minimal set of reactions that need to be added to a draft network such that a specified list of target compounds can be produced from a given set of seed compounds. This gap-filling approach offers the advantage that it is inherently impossible for gap-filling solutions to depend on infeasible cycles. Meneco gap-filling can be directly applied as a method to moped objects. One moped object represents the draft network and a second the repair network, from which the added reactions are chosen.

The topological network modifications, i.e., reversibility and cofactor duplication, harmonize ideally with the application of Meneco, resulting in networks with biologically realistic behavior. This again illustrates the integrative nature of the modelling hub moped. For an accurate manual curation, automatically determined gap-filling reactions should always be manually inspected before adding them to the network model.

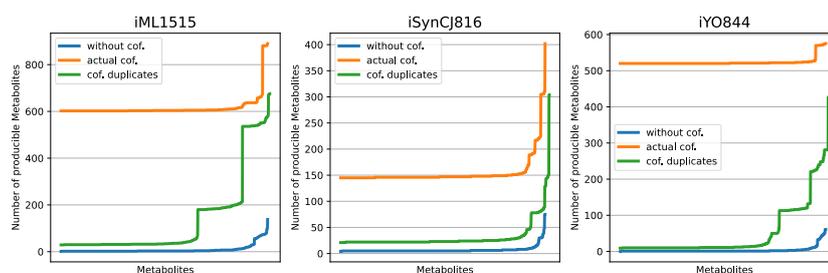
A major advantage and distinguished feature of moped is the complete reproducibility of the construction of draft models, which is much needed in systems biology, and the subsequent manual curation [28]. In moped, the user can add and remove reactions, or even entire pathways, from draft networks. Furthermore, the user can inspect the reactions found by Meneco to fill gaps and decide if these reactions are valid for specific models. All user decisions become part of an executable Python script, making them perfectly reproducible by others. This underlines that in moped, early curation can be integrated closely into the draft model reconstruction process. The importance of such reproducibility and traceability has recently been highlighted [29]. To our knowledge, this feature is unique to is unique to moped and is not yet found in any other reconstruction software. For constraint-based modeling, the user can define which exchange reactions are to be included and, if desired, define their own specific objective functions. moped offers a template biomass function which is based on the iJO1366 and iML1515 biomass functions (see Table S3 in the Supplementary Material); however, users should be encouraged to

design their own specific and precise biomass function for their models as a part of correct manual curation. Reconstructing draft networks in moped lays the ground for model curation without the need to change software environments. In all reconstruction and curation steps, user decisions are documented as commands in an executable Python script, thus making them fully reproducible and transparent.

### 3. Results

#### 3.1. Displaying the Advantage of Cofactor Duplications in Topological Network Analysis

To display the benefits of including the moped cofactor duplication, three established models of *E. coli*, *B. subtilis* and *Synechocystis* sp. PCC 6803 have been parsed into moped for a comparative metabolic network expansion [30–32]. In this analysis, we calculated all single metabolite scopes (i.e., the scopes for the seed consisting only of a single metabolite and water) for the respective models. This has been done in three variations: (i) including no cofactors to the seed, (ii) including the original cofactor compounds and (iii) including on the mock cofactors resulting from cofactor duplication (see above). Figure 4 displays the scope sizes (number of compounds contained in the scope) for each model and each variant to calculate the scopes. Apparently, without cofactors, the scopes are small for most compounds (blue lines). This can be explained by the missing connectivity for reactions that require cofactors. The analysis including the actual cofactor compounds in the seed (orange lines) displays an unrealistically large metabolic scope for every compound, even for inorganic metabolites. This can be explained by the fact that cofactors are usually rather complex metabolites, and now all degradation processes are included during the network expansion. Therefore, the resulting metabolic scopes are no longer reflecting the property of the compound of interest but rather the degradation products of the metabolized cofactor compounds. The corresponding analysis of models using cofactor duplication and mock cofactors duplicates in the seed (green lines) demonstrates that for small or inorganic metabolites, the scope is still relatively small. For more complex organic compounds, the metabolic scope is increasing without artificially increasing the scope size with degradation products of cofactors. This demonstrates the perks of including cofactor duplication and mock cofactors in seeds for biologically more realistic metabolic network expansions.



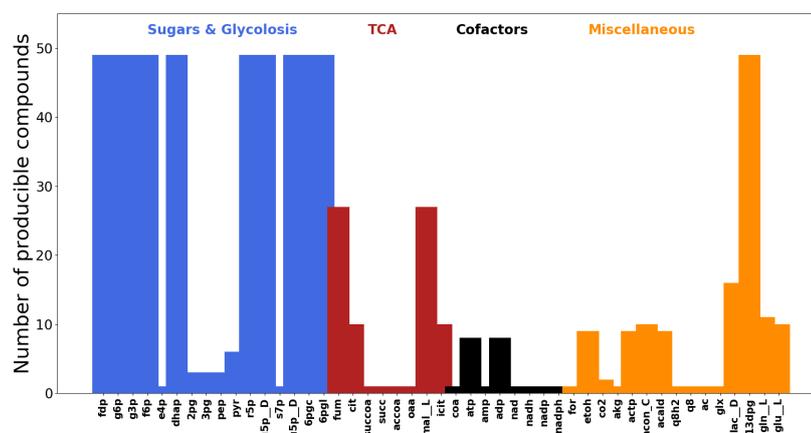
**Figure 4.** Metabolic scopes in established models of *E. coli* (iML1515), *B. subtilis* (iYO844) and *Synechocystis* sp. PCC 6803 (iSynCJ816). The differently colored graphs represent the same analysis but including no cofactors, actual cofactors and cofactor duplicates in the seed.

#### 3.2. Applying Metabolic Network Expansion to a Model of *E. coli* Core Metabolism

We illustrate moped’s metabolic network expansion algorithm with a compact network of *E. coli* core metabolism, which is freely available in SBML format from the BiGG database [33]. After importing the SBML file into moped, we applied cofactor and reversibility duplications as described above.

For each metabolite in the network, we calculate the scope size, i.e., how many new compounds are producible if only this metabolite, water and a set of mock cofactors are available. The results of that analysis are displayed in Figure 5. In this relatively small

metabolic network (72 metabolites and 95 reactions), eleven key compounds, which are mostly part of central metabolism, exhibit a largest observed scope size of 47. Such detailed metabolic network expansion is useful to provide insight about central metabolites, as well as structural and functional characteristics of metabolic networks [14]. Whereas we here only display the scope size, the methods implemented in moped allow a far wider spectrum of analysis methods, including determination the set of producible metabolites, as well as following each step of the expansion process. The code used to produce the results and Figure 5 can be found on <https://gitlab.com/marvin.vanaalst/moped-publication-2021/-/tags/final-publication>, accessed on 13 December 2021.



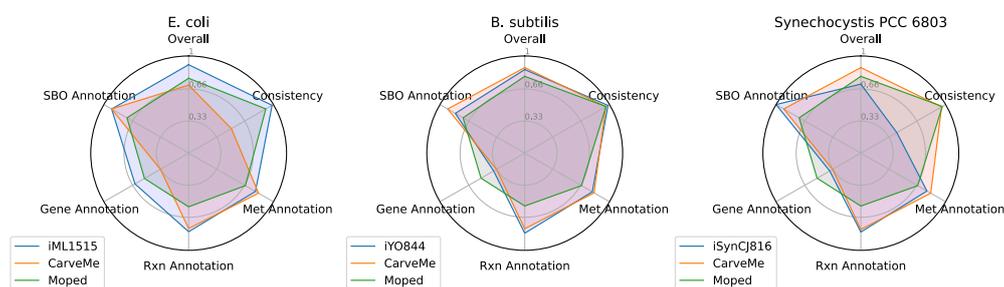
**Figure 5.** Metabolic scopes of all compounds in the *E. coli* core metabolic model calculated using moped. The Y-axis indicates the total amount of compounds producible from every compound, water and a set of acceptor mock cofactors.

### 3.3. Comparison of Draft Reconstructions with Established Models and Softwares

We demonstrate how moped provides a complete and easy-to-use pipeline to construct genome scale models from genome and proteome sequences and how these models can be directly applied for constraint-based analyses. For this, we download the freely available proteome sequences of *Escherichia coli* str. K-12 substr. MG1655, *Synechocystis* sp. PCC 6803 and *Bacillus subtilis* strain 168 [34–36]. We import the MetaCyc PGDB to construct a moped object of the MetaCyc database as a reference network. Applying the BLAST wrapper, which was described above, to the FASTA files and the reference network, we obtained three moped objects, representing the draft network reconstructions. Then, we applied gap filling to ensure that the reconstructed models can produce all basic biomass compounds (inspired by the *E. coli* biomass reaction from iJO1366 [37], including all nucleic acids, amino acids and lipid precursors) from M9 minimal glucose medium. For this analysis, we directly accepted all resulting gap-filling reactions. For a more accurate reconstruction, the proposed gap-filling reactions should be manually inspected before addition to the draft model. We added exchange reactions for all medium compounds and tested if the draft models can exhibit a stationary flux distribution to produce biomass, as determined by flux balance analysis. The construction of these models can be reproduced using the notebooks provided on our accompanying git.

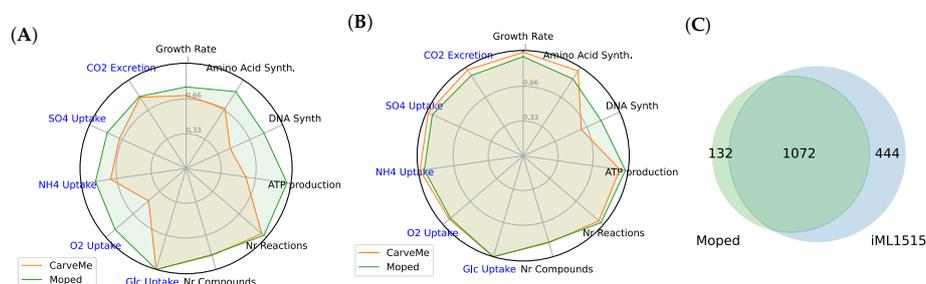
In order to test the quality of our draft models, we compared them with established models for the respective organisms (iML1515, iYO844 and iSynCJ816) [34–36]. Furthermore, we used the same dataset and medium to construct draft models with the established genome scale modeling reconstruction software CarveMe [38]. In order to quantitatively compare all three versions of the organism network reconstructions, we used metabolic model testing (MEMOTE) pipeline to establish a fair and reproducible comparison [39].

MEMOTE calculates scores for genome scale metabolic models to evaluate the stoichiometric consistency, the GPR rules and the quality of annotations for reactions and metabolites in the respective models. A summary of the MEMOTE evaluations for the three models for the three organisms is presented in Figure 6. The MEMOTE evaluation shows that the stoichiometric consistency of draft models produced by moped is always of high quality. Figure 6 shows that draft models reconstructed by CarveMe and moped display generally good overall scores and annotations. While CarveMe draft model reconstructions show the tendency to provide better reaction annotations, moped draft model reconstructions display a generally better annotation of genes and GPR rules.



**Figure 6.** MEMOTE evaluations for draft model reconstructions produced by CarveMe and moped, as well as established models, for *E. coli*, *Bacillus subtilis* and *Synechocystis* sp. PCC 6803. MEMOTE evaluations include the stoichiometric consistency and the annotation level of models.

The functionality and predictive power of draft models constructed by moped has been compared for *Escherichia coli* str. K-12 substr. MG1655 with a similarly constructed draft model using CarveMe, and the iML1515 model. For this analysis, the models automatically constructed moped and CarveMe were analysed without further modification. We calculated maximal growth rates, respective ATP production rates and exchange fluxes for compounds in the medium. Furthermore, we calculated optimal production rates for amino acids and nucleic acids. These model functionalities have been compared to the predictions of iML1515. Figure 7A displays the predicted fluxes of the draft models constructed by moped and by CarveMe relative to the predictions of iML1515. In the radar plots, the relative distance is indicated. For two flux values  $v_1$  and  $v_2$ , the distance  $\min(v_1/v_2, v_2/v_1)$  is plotted. The draft model constructed with moped shows a higher similarity to the behaviour of iML1515 in almost all functionalities, especially in oxygen uptake rate, ATP production rate and nucleic acid synthesis. Some discrepancies between the model behaviours can be linked to slightly differing biomass compositions and lower bounds for exchange fluxes. In order to reduce such bias, we performed the same analysis but with such adjustments that biomass compositions and all lower and upper bounds are identical. Extended MEMOTE evaluations can be found in Figure S2 in the Supplementary Material. Figure 7B shows that now draft models produced with moped and CarveMe exhibit very similar behaviour to iML1515 in all functionalities, except in nucleic acid synthesis, in which moped draft models are more similar to iML1515. The overlap of GPR annotations of the draft model constructed with moped and iML1515 is shown in Figure 7C. The vast majority of genes in the draft model constructed with moped can be found in iML1515 and therefore illustrates the quality of the automated reconstruction. This analysis has only been performed with the draft model constructed with moped because the draft model constructed with CarveMe and iML1515 do not share any common database links. These results show that draft model reconstructions made with moped exhibit a high quality that is able to keep up with the quality of established models and software tools.



**Figure 7.** Functional comparison of the draft model reconstructions using moped and CarveMe with iML1515. We calculated maximal growth rates, respective ATP production rates and exchange fluxes for compounds in the medium, as well as optimal production rates for amino acids and nucleic acids for completely unmodified draft models (A) and models with identical biomass functions and reaction bounds (B). In the radar plots, the relative distance between the two values are reported. Panel (C) shows the overlap of GPR annotations found in the draft model constructed with moped and iML1515.

#### 4. Conclusions

Here, we present moped, a Python package representing a hub connecting the construction, modification and curation of genome scale metabolic networks with various analysis methods, which support studies of metabolic networks. moped supports the de novo construction of metabolic networks by importing databases, providing homology searches, including GPR associations and integrating an established gap-filling routine without the need to change software environments. Existing models from external sources can be imported using the standardized format SBML. Metabolic network models are represented as moped objects, which can be modified by easy-to-use and intuitive methods. moped models can be exported into various formats, thus integrating a diverse set of established analysis tools. Metabolic network expansion and constraint-based optimization can be easily performed for any model represented as a moped object.

Examination of moped draft model reconstructions using MEMOTE demonstrated that the resulting models are generally of a high quality. The strength of draft model reconstructions with moped is the direct integration into the Python programming language: Every decision in the automatic and manual reconstruction process is documented in executable Python scripts. Therefore, the whole reconstruction process becomes fully transparent and is easily reproducible by any interested user.

The modular architecture of the open source package moped is particularly designed for allowing further extensions to enhance its functionality, such as the integration of additional software tools. We provide an extensive documentation for moped, as well as troubleshooting guides, unit-tests for all provided methods and example notebooks illustrating the usage of moped at <https://gitlab.com/marvin.vanaalst/moped-publication-2021> (accessed on 13 December 2021).

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/metabo12040275/s1>, Figure S1: UML diagram of core packages in moped, Figure S2: Extended MEMOTE evaluations for draft model reconstructions, Table S1: Cofactor pairs of MetaCyc identifiers, Table S2: Cofactor pairs of BiGG identifiers, Table S3: Default biomass composition

**Author Contributions:** Conceptualization, N.P.S.; implementation, N.P.S. and M.v.A.; writing—original draft preparation, N.P.S.; writing—review and editing, N.P.S., M.v.A. and O.E. All authors have read and agreed to the published version of the manuscript.

**Funding:** Funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany’s Excellence Strategy EXC 2048/1, Project ID: 390686111 (O.E.) and EU’s Horizon 2020 research and innovation programme under the Grant Agreement 862087 (M.v.A.).

**Data Availability Statement:** Operating systems: Linux, OS X Programming language: Python License: GPLv3 Any restrictions to use by nonacademics: For nonprofit use only All source code including scripts to produce all manuscript figures can be found at <https://gitlab.com/marvin.vanaalst/moped> (accessed on 13 December 2021) and <https://gitlab.com/marvin.vanaalst/moped-publication-2021> (accessed on 13 December 2021).

**Acknowledgments:** We thank Anna Matuszyńska, St. Elmo Wilken and Ovidiu Popa for critically reading our manuscript draft, Clémence Frioux for providing support for Meneco integration and Dipali Singh for giving valuable advice on MetaCyc and PDGB issues.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

The following abbreviations are used in this manuscript:

PGDB	Pathway/Genome Database
FBA	Flux Balance Analysis
GPR	Gene–Protein–Reaction
SBML	Systems Biology Markup Language
ODE	Ordinary Differential Equations

### References

- Rapoport, T.A.; Heinrich, R.; Jacobasch, G.; Rapoport, S. A linear steady-state treatment of enzymatic chains. A mathematical model of glycolysis of human erythrocytes. *Eur. J. Biochem.* **1974**, *42*, 107–120. [[CrossRef](#)] [[PubMed](#)]
- Zomorodi, A.R.; Segrè, D. Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities. *Nat. Commun.* **2017**, *8*, 1563. [[CrossRef](#)] [[PubMed](#)]
- Hartman, H.B.; Fell, D.A.; Rossell, S.; Jensen, P.R.; Woodward, M.J.; Thorndahl, L.; Jelsbak, L.; Olsen, J.E.; Raghunathan, A.; Daefler, S.; et al. Identification of potential drug targets in *Salmonella enterica* sv. Typhimurium using metabolic modelling and experimental validation. *Microbiology* **2014**, *160*, 1252–1266. [[CrossRef](#)] [[PubMed](#)]
- Heinrich, R.; Schuster, S. *The Regulation of Cellular Systems*; Chapman and Hall: New York, NY, USA, 1996.
- Caspi, R.; Billington, R.; Keseler, I.M.; Kothari, A.; Krummenacker, M.; Midford, P.E.; Ong, W.K.; Paley, S.; Subhraveti, P.; Karp, P.D. The MetaCyc database of metabolic pathways and enzymes—a 2019 update. *Nucleic Acids Res.* **2020**, *48*, D445–D453. [[CrossRef](#)] [[PubMed](#)]
- Karp, P.D.; Billington, R.; Caspi, R.; Fulcher, C.A.; Latendresse, M.; Kothari, A.; Keseler, I.M.; Krummenacker, M.; Midford, P.E.; Ong, Q.; et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* **2019**, *20*, 1085–1093. [[CrossRef](#)] [[PubMed](#)]
- King, Z.A.; Lu, J.; Dräger, A.; Miller, P.; Federowicz, S.; Lerman, J.A.; Ebrahim, A.; Palsson, B.O.; Lewis, N.E. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **2016**, *44*, D515–D522. [[CrossRef](#)]
- Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)]
- Raman, K.; Chandra, N. Flux balance analysis of biological systems: Applications and challenges. *Brief. Bioinform.* **2009**, *10*, 435–449. [[CrossRef](#)]
- Orth, J.D.; Thiele, I.; Palsson, B.Ø. What is flux balance analysis? *Nat. Biotechnol.* **2010**, *28*, 245–248. [[CrossRef](#)]
- Wunderlich, Z.; Mirny, L.A. Using the topology of metabolic networks to predict viability of mutant strains. *Biophys. J.* **2006**, *91*, 2304–2311. [[CrossRef](#)]
- Handorf, T.; Ebenhöf, O.; Heinrich, R. Expanding metabolic networks: Scopes of compounds, robustness, and evolution. *J. Mol. Evol.* **2005**, *61*, 498–512. [[CrossRef](#)] [[PubMed](#)]
- Romero, P.; Karp, P. Nutrient-related analysis of pathway/genome databases. In *Biocomputing 2001*; World Scientific: Singapore, 2000; pp. 471–482.
- Ebenhöf, O.; Handorf, T. Functional classification of genome-scale metabolic networks. *EURASIP J. Bioinform. Syst. Biol.* **2009**, *2009*, 570456. [[CrossRef](#)] [[PubMed](#)]
- Christian, N.; May, P.; Kempa, S.; Handorf, T.; Ebenhöf, O. An integrative approach towards completing genome-scale metabolic networks. *Mol. Biosyst.* **2009**, *5*, 1889–1903. [[CrossRef](#)]
- Orth, J.D.; Palsson, B.Ø. Systematizing the generation of missing metabolic knowledge. *Biotechnol. Bioeng.* **2010**, *107*, 403–412. [[CrossRef](#)] [[PubMed](#)]
- Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]

18. Prigent, S.; Frioux, C.; Dittami, S.M.; Thiele, S.; Larhlimi, A.; Collet, G.; Gutknecht, F.; Got, J.; Eveillard, D.; Bourdon, J.; et al. Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. *PLoS Comput. Biol.* **2017**, *13*, e1005276. [[CrossRef](#)]
19. Ebrahim, A.; Lerman, J.A.; Palsson, B.O.; Hyduke, D.R. COBRApy: CONstraints-based reconstruction and analysis for python. *BMC Syst. Biol.* **2013**, *7*, 74. [[CrossRef](#)] [[PubMed](#)]
20. van Aalst, M.; Ebenhöf, O.; Matuszyńska, A. Constructing and analysing dynamic models with modelbase v1. 2.3: A software update. *BMC Bioinform.* **2021**, *22*, 203. [[CrossRef](#)]
21. Hucka, M.; Finney, A.; Sauro, H.M.; Bolouri, H.; Doyle, J.C.; Kitano, H.; Arkin, A.P.; Bornstein, B.J.; Bray, D.; Cornish-Bowden, A.; et al. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics* **2003**, *19*, 524–531. [[CrossRef](#)]
22. Karp, P.D.; Paley, S.; Romero, P. The pathway tools software. *Bioinformatics* **2002**, *18*, S225–S232. [[CrossRef](#)]
23. Karpe, P.D.; Latendresse, M.; Caspi, R. The pathway tools pathway prediction algorithm. *Stand. Genom. Sci.* **2011**, *5*, 424–429. [[CrossRef](#)] [[PubMed](#)]
24. Machado, D.; Hergård, M.J.; Rocha, I. Stoichiometric representation of gene–protein–reaction associations leverages constraint-based analysis from reaction to gene-level phenotype prediction. *PLoS Comput. Biol.* **2016**, *12*, e1005140. [[CrossRef](#)] [[PubMed](#)]
25. Thiele, I.; Palsson, B.Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **2010**, *5*, 93. [[CrossRef](#)]
26. Kumar, V.S.; Dasika, M.S.; Maranas, C.D. Optimization based automated curation of metabolic reconstructions. *BMC Bioinform.* **2007**, *8*, 212.
27. Vitkin, E.; Shlomi, T. MIRAGE: A functional genomics-based approach for metabolic network model reconstruction and its application to cyanobacteria networks. *Genome Biol.* **2012**, *13*, R111. [[CrossRef](#)] [[PubMed](#)]
28. Tiwari, K.; Kananathan, S.; Roberts, M.G.; Meyer, J.P.; Sharif Shohan, M.U.; Xavier, A.; Maire, M.; Zyouf, A.; Men, J.; Ng, S.; et al. Reproducibility in systems biology modelling. *Mol. Syst. Biol.* **2021**, *17*, e9982. [[CrossRef](#)]
29. Mendoza, S.N.; Olivier, B.G.; Molenaar, D.; Teusink, B. A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biol.* **2019**, *20*, 158. [[CrossRef](#)]
30. Monk, J.M.; Lloyd, C.J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **2017**, *35*, 904–908. [[CrossRef](#)]
31. Oh, Y.K.; Palsson, B.O.; Park, S.M.; Schilling, C.H.; Mahadevan, R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.* **2007**, *282*, 28791–28799. [[CrossRef](#)]
32. Joshi, C.J.; Peebles, C.A.; Prasad, A. Modeling and analysis of flux distribution and bioproduct formation in *Synechocystis* sp. PCC 6803 using a new genome-scale metabolic reconstruction. *Algal Res.* **2017**, *27*, 295–310. [[CrossRef](#)]
33. Orth, J.D.; Fleming, R.M.; Palsson, B.O. Reconstruction and use of microbial metabolic networks: The core *Escherichia coli* metabolic model as an educational guide. *EcoSal Plus* **2010**, *4*, 1. [[CrossRef](#)] [[PubMed](#)]
34. Blattner, F.R.; Plunkett, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; et al. The complete genome sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453–1462. [[CrossRef](#)] [[PubMed](#)]
35. Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.; Sasamoto, S.; et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **1996**, *3*, 109–136. [[CrossRef](#)] [[PubMed](#)]
36. Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.; Alloni, G.; Azevedo, V.; Bertero, M.; Bessières, P.; Bolotin, A.; Borchert, S.; et al. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **1997**, *390*, 249–256. [[CrossRef](#)]
37. Orth, J.D.; Conrad, T.M.; Na, J.; Lerman, J.A.; Nam, H.; Feist, A.M.; Palsson, B.Ø. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.* **2011**, *7*, 535. [[CrossRef](#)]
38. Machado, D.; Andrejev, S.; Tramontano, M.; Patil, K.R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **2018**, *46*, 7542–7553. [[CrossRef](#)]
39. Lieven, C.; Beber, M.E.; Olivier, B.G.; Bergmann, F.T.; Ataman, M.; Babaei, P.; Bartell, J.A.; Blank, L.M.; Chauhan, S.; Correia, K.; et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **2020**, *38*, 272–276. [[CrossRef](#)]

## **6.3 Balancing energy supply during photosynthesis - a theoretical perspective**

### **6.3.1 Authors and Details**

Authors: Anna Matuszyńska, Nima P. Saadat, Oliver Ebenhöf  
Authorship: 1. Author (shared)  
Journal: *Physiologia Plantarum*  
Status: Published (doi:10.1111/ppl.12962)

### **6.3.2 Contributions**

Nima P. Saadat helped in merging the models, developed research hypotheses, performed all computational analyses, collected all results and interpreted the results, and wrote the result and discussion sections of the draft manuscript. Anna Matuszyńska merged the models, provided all conversions and mathematical descriptions, wrote the introduction and discussion sections of the manuscript and helped in interpreting the results. Oliver Ebenhöf helped developing research hypotheses, interpreting the results and helped improving the manuscript.

## Balancing energy supply during photosynthesis – a theoretical perspective

Anna Matuszyńska<sup>a,b,\*</sup>, Nima P. Saadat<sup>a,†</sup> and Oliver Ebenhöh<sup>a,b</sup>

<sup>a</sup>Institute of Quantitative and Theoretical Biology, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

<sup>b</sup>CEPLAS Cluster of Excellence on Plant Sciences, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

### Correspondence

\*Corresponding author,  
e-mail: anna.matuszynska@hhu.de

Received 13 November 2018;  
revised 22 February 2019

doi:10.1111/ppl.12962

The photosynthetic electron transport chain (PETC) provides energy and redox equivalents for carbon fixation by the Calvin-Benson-Bassham (CBB) cycle. Both of these processes have been thoroughly investigated and the underlying molecular mechanisms are well known. However, it is far from understood by which mechanisms it is ensured that energy and redox supply by photosynthesis matches the demand of the downstream processes. Here, we deliver a theoretical analysis to quantitatively study the supply–demand regulation in photosynthesis. For this, we connect two previously developed models, one describing the PETC, originally developed to study non-photochemical quenching, and one providing a dynamic description of the photosynthetic carbon fixation in C3 plants, the CBB Cycle. The merged model explains how a tight regulation of supply and demand reactions leads to efficient carbon fixation. The model further illustrates that a stand-by mode is necessary in the dark to ensure that the carbon fixation cycle can be restarted after dark–light transitions, and it supports hypotheses, which reactions are responsible to generate such mode in vivo.

### Introduction

Decades of multidisciplinary research of photosynthesis resulted in a detailed understanding of the molecular, regulatory and functional mechanisms of light-driven carbon fixation. Yet, still much is to uncover, especially in terms of identifying processes limiting photosynthetic productivity, and further basic research will be necessary to redesign and potentially optimize photosynthesis (Ort et al. 2015, Cardona et al. 2018). Historically, the process of photosynthesis has been divided into two parts. The so-called ‘light reactions’ of the photosynthetic electron transport chain (PETC) convert light into chemical

energy, supplying ATP and NADPH. This energy is used to drive the carbon dioxide reduction and fixation processes known as the ‘dark reactions’. Thus, the photosynthetic light and dark reactions can be viewed as a molecular economy supply–demand system (Hofmeyr and Cornish-Bowden 2000, Rohwer and Hofmeyr 2008, Christensen et al. 2015).

Despite this clear interdependence, these processes are often studied in isolation. This approach permits a detailed and in-depth analysis of particular components at the cost of simplifying others. This separation is also reflected in theoretical research. Numerous approaches in the past decades aimed at translating the complexity

*Abbreviations* – CBB, Calvin-Benson-Bassham; MCA, metabolic control analysis; NPQ, non-photochemical quenching; ODE, ordinary differential equations; PETC, photosynthetic electron transport chain; PPP, pentose phosphate pathway; PQ, plastoquinone; PS, photosystem; Ru5P, ribulose-5-phosphate; SBPase, seduheptulose-1,7-bisphosphatase; TPT, triose phosphate transporters.

<sup>†</sup>These authors equally contributed to this work.

of photosynthesis into a mathematical language, resulting in an impressive portfolio of kinetic models. The majority of these models focus either on the supply or on the demand side. Many classical models of the Calvin-Benson-Bassham (CBB) cycle, such as the biochemical models for C3 photosynthetic CO<sub>2</sub> assimilation (Hahn 1986, 1987, Pettersson and Ryde-Pettersson 1988, Poolman et al. 2000, Farquhar et al. 2007, Zhu et al. 2007, 2009), made no attempt to model the processes of the PETC in any detail. Instead, they simplify the rate of electron transport supplying ATP and NADPH in often just one lumped reaction (e.g. non-rectangular hyperbola as a function of absorbed irradiance in the study by Morales et al. (2018a)), or even considered key components as constant (NADPH in the study by Pettersson and Ryde-Pettersson (1988)). Likewise, many models of the PETC made no attempt to include details of the energy consuming reactions and describe ATP and NADPH demand by simple lumped reactions. Such an understandable simplification resulted from the fact that these models were created to study specific light harvesting mechanisms, such as state transitions (Ebenhöh et al. 2014), non-photochemical quenching (NPQ) (Ebenhöh et al. 2011, Zaks et al. 2012, Matuszyńska et al. 2016) or the role of antenna complexes in photosynthetic productivity (Rubin and Riznichenko 2009).

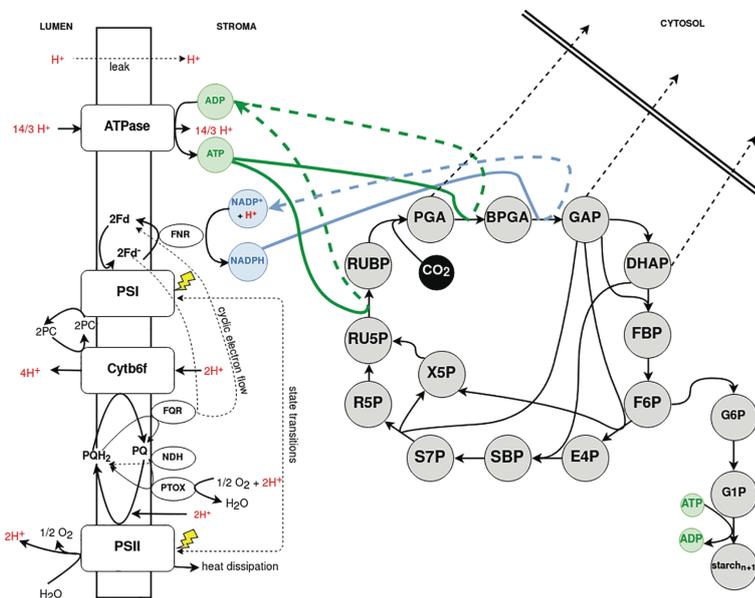
The purpose of this study is to provide a theoretical understanding of the interactions and interdependencies of the PETC and the carbon fixation cycle, with a focus on investigating the supply–demand control of photosynthesis. For such an exercise, mathematical models are ideally suited, because they allow systematic alterations of parameters, which are not easily accessible experimentally, and thus to draw general conclusions about regulatory principles. Apparently, investigating the delicate supply–demand system of photosynthesis requires a mathematical model that contains both processes. Noteworthy, there exist a few successful attempts to include both electron transport and carbon assimilation processes into a unified mathematical framework. The model proposed by Laisk et al. (2006) provides a solid summary of our knowledge on photosynthesis. The model was constructed with an emphasis on including the electron transport through photosystems PSII and PSI, together with a detailed description of the downstream metabolism. As a result, the model can represent steady state photosynthesis and chlorophyll fluorescence, but is insufficient to reproduce dark–light induction of photosynthesis, a property that is critical in the context of our proposed supply–demand analysis. The ‘e-photosynthesis’ model by Zhu et al. (2013) is a comprehensive description including ‘as many photosynthesis-related reactions as possible’. Because of

its complexity, using the e-photosynthesis model (Zhu et al. 2013) for a systematic supply–demand analysis is challenging. Moreover, the highly detailed description of the molecular processes included in the model makes it hard to draw conclusions of general validity. Finally, Morales et al. (2018b) recently developed a thorough model of the PETC, including all relevant processes at the chloroplast and leaf level. Nevertheless, as the emphasis of this model was on the PETC regulation, the CBB cycle has been simplified into two steps. This imbalance in the levels of detail describing the two sub-processes is the main reason why we decided against using it.

We have therefore developed a new photosynthesis model that contains the key components of both subsystems, yet is simple enough to allow for systematic investigations. The model has been constructed by merging a model of the PETC, originally designed to study photoprotective mechanisms (Ebenhöh et al. 2014, Matuszyńska et al. 2016), with a kinetic model of C3 carbon fixation (Pettersson and Ryde-Pettersson 1988, Poolman et al. 2000). We demonstrate that coupling these two models into a connected supply–demand system is possible, but far from trivial, and results in new emergent properties. Using metabolic control analysis (Kacser and Burns 1973, Heinrich and Rapoport 1974, Heinrich and Schuster 1996) and metabolic supply–demand analysis (Hofmeyr and Cornish-Bowden 2000), we provide a quantitative description how the control over the overall photosynthetic flux is distributed under various conditions. Moreover, our model analysis illustrates the need for a stand-by mode of the carbon fixation cycle in the dark to ensure that it can be restarted after prolonged dark periods. Our model results demonstrate that the oxidative pentose phosphate pathway (PPP) can provide exactly this functionality. Remarkably, deactivation of CBB enzymes in the dark alone is insufficient to enable reactivation. These insights could not have been obtained without a model that merges light-dependent and -independent reactions.

We further expect that our model presented here serves as a basis for future developments. We have specifically constructed the model in a modular architecture, which makes it technically straight-forward to include other relevant interacting pathways, such as photorespiration or other ATP consuming processes. Together with quantitative experimental data, it will be possible to parameterize the model to specific organisms and conditions. Thus validated, we expect that the model becomes a useful tool to predict how photosynthetic efficiency is affected upon environmental or genetic perturbations. We therefore envision that our model, with suitable modifications,

**Fig. 1.** Schematic representation of the photosynthetic processes described by our merged mathematical model. The reactions take place in two compartments. In the lumen, the four protein supercomplexes (PSII, PSI, Cytb6f and ATPase) are embedded, which drive the electron transport in two modes, linear and cyclic; the stroma provides the compartment of C3 photosynthetic carbon fixation. The cytosol defines the system boundary. In color (green and blue) we have highlighted the reactions linking the two sub-models: The production and consumption of ATP and NADPH, respectively.



will provide a sound theory that supports attempts to improve photosynthetic performance in a targeted manner.

### The model

We are presenting here the result of connecting two previously developed kinetic models of photosynthesis, both based on ordinary differential equations (ODEs). The first model describes the primary photosynthetic reactions through the PETC, leading to the production of ATP and NADPH. The CBB cycle is considered as the main consumer of the produced energy and reducing equivalents. Therefore in this model, the downstream metabolism has been simplified to two consuming reactions governed by mass action kinetics. It has been developed based on our previous work: the core model of the PETC by Ebenhöf et al. (2014) and the model of high-energy dependent quenching in higher plants developed by Matuszyńska et al. (2016). Using this model, we are able to compute the fluorescence emission under various light protocols, monitor the redox state of the thylakoids and the rate of ATP and NADPH synthesis. The second model is the Poolman (Poolman et al. 2000) implementation of the carbon fixation model by Pettersson and Ryde-Pettersson (1988), reproduced in our institute using the modelbase software (Ebenhöf et al. 2018). In contrast to the original model

(Pettersson and Ryde-Pettersson 1988), in the Poolman representation the rapid equilibrium assumptions were not solved explicitly, but instead approximated by mass-action kinetics with very large rate constants. Solving the system of ODEs allows computation of different carbon fixation rates and reaction activities at varying concentrations of external orthophosphate. In the original model, the input of the ETC has been simplified by a single lumped reaction of ATP synthesis ( $v_{16}$  in the study by Pettersson and Ryde-Pettersson 1988), while NADPH has been kept as a constant parameter.

### Included processes and the stoichiometry

The model, schematically represented in Fig. 1, comprises 35 reaction rates and follows the dynamics of 24 independent state variables (Appendix S1, Supporting Information, for a full list of reaction rates and ODEs). In addition, we compute a number of values such as emitted fluorescence or variables derived from conserved quantities. Light is considered as an external, time-dependent variable. As the focus of this model is to study basic system properties, such as the response to relative changes in the light intensity, we did not calibrate our simulations to experimentally measured light intensities. Therefore in this work, light is expressed in micromoles of photons per square meter

per second ( $\mu\text{mol m}^{-2} \text{s}^{-1}$ ) and reflects the quantity of light efficiently used, but the conversion factor to the photon flux density of the incident light is unknown. We included two compartments in our model, the thylakoid lumen and the chloroplast stroma. In the lumen, the reaction kinetics for oxidized plastoquinone (PQ), oxidized plastocyanin, oxidized ferredoxin, lumenal protons (H) and non-phosphorylated antenna (light harvesting complexes) were taken from Ebenhöf et al. (2014). The four-state description of the quencher activity, based on the protonation of the PsbS protein and activity of the xanthophyll cycle, was taken from our mathematical model of NPQ, initially developed to study short-term light memory in plants (Matuszyńska et al. 2016). The previous description of ATP and NADPH consuming reactions is supplemented by the detailed description of the CBB cycle, taking place in the stroma. Processes of the CBB cycle have been implemented as in the mathematical model of C3 photosynthesis by Poolman et al. (2000), based on the original work of Pettersson and Ryde-Pettersson (1988). The original model reproduces different carbon fixation rates and reaction activities at different concentrations of external orthophosphate, and includes the conversion of fixed carbon into either triose phosphates or sugar and starch. This model has been parametrized for  $\text{CO}_2$  saturating conditions and we kept the same assumption for all our analyses. The previous description of ATP synthesis is supplemented in our model with the new rate  $v_{\text{ATPsynthase}}$  which depends on the proton motive force built up by the PETC activity. Moreover, the stromal concentration of NADPH is dynamic.

### Model compartments and units

The original models were initially developed for different organisms (Ebenhöf et al. (2014) for *Chlamydomonas reinhardtii*, Matuszyńska et al. (2016) for *Arabidopsis thaliana* and Pettersson and Ryde-Pettersson (1988), Poolman et al. (2000) based on data for isolated spinach chloroplasts). Moreover, concentrations and rates were expressed in different units. This patchwork of parameters motivated us to create a general model of photosynthesis, which is not restricted to a single organism. To keep the original structure of the models, but provide consistency, we have kept the original units for each of the compartments and used a conversion factor ( $p_{\text{conv}}$ , Appendix S1) to convert quantities where needed. Thus, concentrations of proteins and pool sizes inside the lumen are expressed as in previous models of the electron transport (Ebenhöf et al. 2014, Matuszyńska et al. 2016) in  $\text{mmol} (\text{mol Chl})^{-1}$ , and the first order rates in  $\text{mmol} (\text{mol Chl})^{-1} \text{s}^{-1}$ .

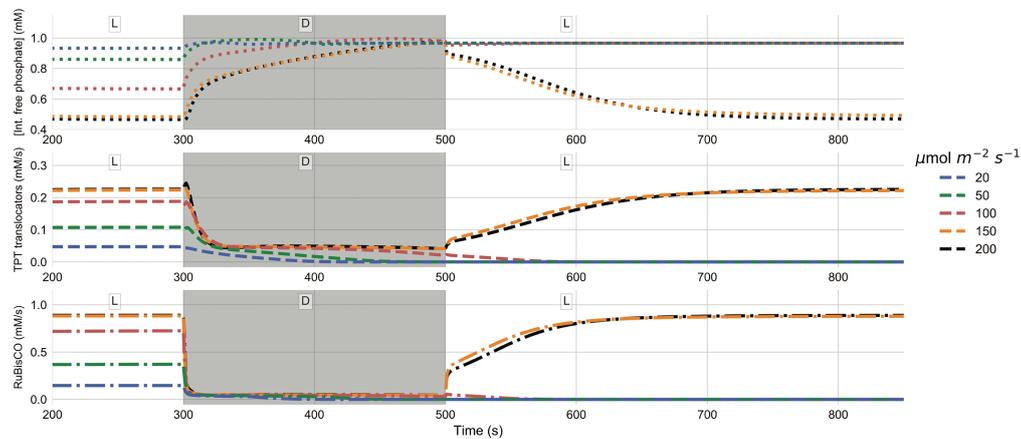
Concentrations of metabolites and pools inside the stroma are expressed in mM, as in (Pettersson and Ryde-Pettersson 1988; Poolman et al. 2000). To convert the concentration of ATP produced through the electron transport chain activity, expressed in  $\text{mmol} (\text{mol Chl})^{-1}$ , to mM, used to express concentrations in the stroma, we made several assumptions, as in our previous models of photosynthesis (Ebenhöf et al. 2011, 2014, Matuszyńska et al. 2016), which were originally derived from Laisk et al. (2006)): (1) chlorophyll content is assumed to be fixed and equal to  $350 \cdot 10^{-6} \text{ mol m}^{-2}$  thylakoid membrane, (2) the volume of thylakoid stroma and lumen are 0.0112 and  $0.0014 \text{ l m}^{-2}$ , respectively. Thus,  $1 \text{ mmol} (\text{mol Chl})^{-1}$  corresponds to  $2.5 \cdot 10^{-4} \text{ M}$  in the lumen and  $3.2 \cdot 10^{-5} \text{ M}$  in the stroma. Although the results presented here have been obtained for these particular values describing the surface-to-volume ratios inside the chloroplast, it is in principle easy to change the according parameters to reflect different experimental conditions (Matuszyńska et al. 2016).

### Computational analysis

The model has been implemented using the modelbase software, a console-based application written in Python, recently developed by us (Ebenhöf et al. 2018). Stoichiometry and parameters are provided in Appendix S1, to be found on our GitHub repository ([www.github.com/QTBB-HHU/photosynthesismodel](http://www.github.com/QTBB-HHU/photosynthesismodel)). Moreover, we provide a Jupyter Notebook that allows the user to repeat all the simulations leading to the production of the figures presented in this manuscript.

### Reliability of the model

We have assembled the model of photosynthesis adapting previously validated and published mathematical models of two interdependent processes. We have used the same parameters as reported in the previous work and did not perform any further parameter fits (the full list of parameters is provided in Tables S1–S5 in Appendix S1). We have monitored the evolution of several critical parameters to evaluate physiological plausibility of our computational results, including lumenal pH (kept under moderate light around 6), RuBisCO rate (in the order of magnitude of measured values) and the redox state of the PQ pool, used as an estimate of the overall redox state. Moreover, systematic steady state analysis of the model under different light conditions lead to plausible concentrations of CBB cycle intermediates and fluxes, as reported in the literature (Pettersson and Ryde-Pettersson 1988).



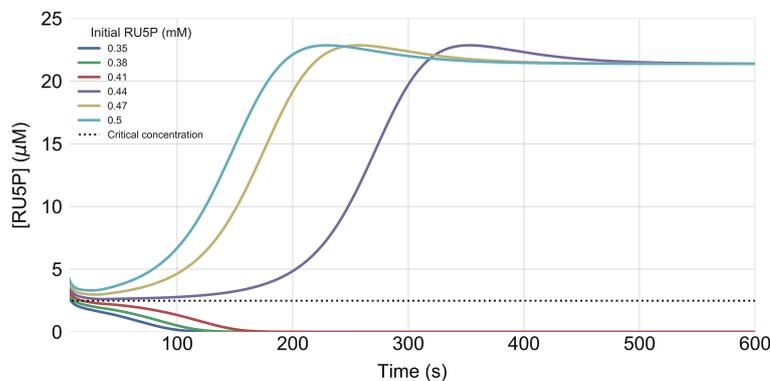
**Fig. 2.** Simulations of light–dark–light transitions for different light intensities, ranging from 20 to 200  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . Shown are the dynamics of internal orthophosphate concentration, triose phosphate transporter (TPT) export and carbon fixation rates. The simulated time-courses are shown from 200s, when the system has reached a stationary state. From 300 to 500s (gray area), the external light has been set to 5  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . The figure illustrates that for low light intensities the CBB cycle fails to restart in the second light period.

## Results and discussion

We used our merged model of photosynthesis and carbon fixation to perform a systematic supply–demand analysis of the coupled system. First, we have integrated the system for various constant light intensities until it reached steady state. Examples are provided in Fig. S1 in Appendix S1. We observed reasonable stationary values of intermediates and fluxes for most of the light intensities. However, under very low light intensities (below 5  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ), the phosphorylated CBB cycle intermediates dropped to zero, and ATP reached the maximal concentration equalling the total pool of adenosine phosphates. Depending on the initial conditions, either a non-functioning state, characterized by zero carbon fixation rate, or a functioning state, characterized by a positive stationary flux, was reached. This observation of bistability constituted the starting point of our analysis of the tight supply–demand relationship.

In order to analyze this behavior in more detail, we performed time course simulations, in which the light was dynamically switched from constant sufficient light (between 20 and 200  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ), to a ‘dark phase’ of 200s duration with a light intensity of 5  $\mu\text{mol m}^{-2} \text{s}^{-1}$ , back to high light, and observed the dynamics of the model variables. In Fig. 2 we display the dynamics of the internal orthophosphate concentration, the sum of all three triose phosphate transporter (TPT) export rates and the RuBisCO rate (from top to bottom, respectively) during such light–dark–light simulations.

In agreement with the steady-state simulations, higher light intensities result in a higher overall flux during the initial light phase. Higher carbon fixation and export fluxes are accompanied by lower orthophosphate concentrations, which reflect higher levels of CBB cycle intermediates. In the dark phase, the non-functional state with zero carbon flux is approached. While rates decrease, orthophosphate increases, reflecting a depletion of the CBB intermediate pools. In the second light phase, only the simulated transitions to light intensities of 150 and 200  $\mu\text{mol m}^{-2} \text{s}^{-1}$  could recover a functional state under the chosen conditions. For lower light intensities, apparently the CBB intermediate pool was depleted to a level, at which re-illumination fails to recover the CBB cycle activity. Obviously, this behavior disagrees with everyday observations in nature (plant leaves recover from dark periods also under low light intensities). Nevertheless, the model is useful to generate novel insights. First, it illustrates that a critical threshold of intermediate concentrations exists. If levels drop below this threshold, the cycle cannot be re-activated. Second, it explains the mechanisms leading to intermediate depletion. Under low light conditions, insufficient energy supply results in reduced activity of ATP and NADPH dependent reactions in the carbon fixation cycle, leading to a reduced regeneration rate of ribulose 1,5-bisphosphate from ribulose-5-phosphate (Ru5P). Simultaneously, the reversible (ATP independent) reactions remain active. As triose phosphates are products of reversible reactions, these continue to be exchanged via the TPT



**Fig. 3.** Simulations in light intensity of  $500 \mu\text{mol m}^{-2} \text{s}^{-1}$  for different initial concentrations of Ru5P, ranging from 0.35 to 0.5 mM. The Ru5P abundance is shown after 10 s, when the system is approximately equilibrated. The dashed line displays the critical concentration for sufficient cyclic activity after equilibrating. The figure displays that initial Ru5P concentrations below 0.44 mM result in a loss of Ru5P abundance.

export reactions with free phosphate, which leads to a depletion of the CBB cycle intermediates and a concomitant increase of the orthophosphate pool. This further illustrates that even deactivating key light-regulated CBB enzymes in the dark will not prevent the collapse of the cycle, because the continued activity of the reversible reactions and the triose phosphate translocator will still lead to depleted cycle intermediates (Fig. S2 in Appendix S1).

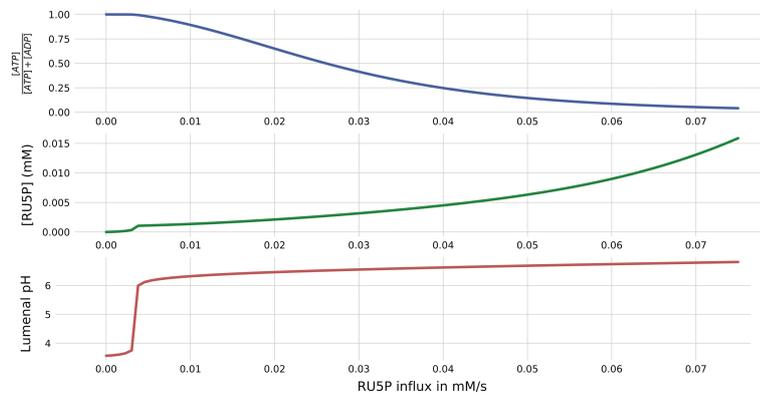
Clearly, the model is missing important mechanisms that prevent such a functional failure. In particular, we are interested in how a stand-by mode can be realized, in which intermediate levels are maintained above the critical threshold, while at the same time the resources required to do so, are minimized. A possible strategy to prevent the collapse of the carbon fixation cycle is to resupply important intermediates. One biochemical process in plants that is known to produce Ru5P is the oxidative phase of the PPP, in which one glucose-6-phosphate molecule is oxidized and decarboxylated to Ru5P, while producing NADPH and  $\text{CO}_2$  (Kruger and Von Schaewen 2003). In order to estimate critical intermediate levels required to prevent the collapse of the carbon fixation cycle, we performed simulations under sufficient light ( $500 \mu\text{mol m}^{-2} \text{s}^{-1}$ ), with different initial conditions: the initial concentrations of all carbon fixation intermediates are set to zero, except for Ru5P. The simulated Ru5P concentration, depicted in Fig. 3, displays a characteristic dynamic. In the first seconds, the CBB cycle intermediates are equilibrated by the fast reversible reactions. If the equilibrated Ru5P concentration remains above the critical threshold of approximately  $2.5 \mu\text{M}$ , the cycle reaches a functional state, if it falls below, it will collapse. Interestingly, the threshold concentration is rather independent of the light intensity (Fig. S3 in Appendix S1).

To simulate a simple mechanism implementing a stand-by mode, which maintains sufficient CBB cycle intermediate levels, we introduced a trivial conceptual reaction, exchanging inorganic phosphate with Ru5P. Fig. 4 displays simulated steady state values of the relative stromal ATP concentrations, Ru5P concentrations and luminal pH in insufficient light conditions ( $5 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) as a function of the Ru5P influx. Again, a clear threshold behavior can be observed. If the Ru5P influx exceeds approximately  $4 \mu\text{M s}^{-1}$ , not only CBB intermediates assume non-zero concentrations, but also the luminal pH reaches realistic and non-lethal levels.

As expected, increased Ru5P influx results in increased stationary Ru5P concentrations, which is accompanied by an increased flux through RuBisCO and the TPT exporter (Fig. S4 in Appendix S1), indicating a higher stand-by flux, and therefore, a higher requirement of resources to maintain this mode.

These results suggest that a constant flux providing Ru5P in the dark with a rate just above the critical threshold of  $4 \mu\text{M s}^{-1}$  should maintain intermediate CBB levels sufficiently high, while at the same time minimize the required investment. Indeed, with a constant supply of Ru5P with  $5 \mu\text{M s}^{-1}$ , the system can be restarted and reaches a functional stationary state after a prolonged dark period (Fig. S5 in Appendix S1). Per carbon, this rate translates to  $25\text{--}30 \mu\text{M carbon/s}$ , depending whether the pentoses are directly imported or derived from hexoses. Comparing this to stationary carbon fixation in the light of  $0.1\text{--}1 \text{ mM s}^{-1}$  (for light intensities between  $20$  and  $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ , Fig. 2 and Fig. S1 in Appendix S1) shows that resupply under these conditions would consume a considerable fraction of the previously fixed carbon. This calculation demonstrates the importance of down-regulating the CBB

**Fig. 4.** Steady state simulations in low light intensity of  $5 \mu\text{mol m}^{-2} \text{s}^{-1}$  and systematically increasing influxes of Ru5P from 0 to  $0.08 \text{ mM s}^{-1}$ . The figure displays normalized ATP abundance, Ru5P concentration and luminal pH.



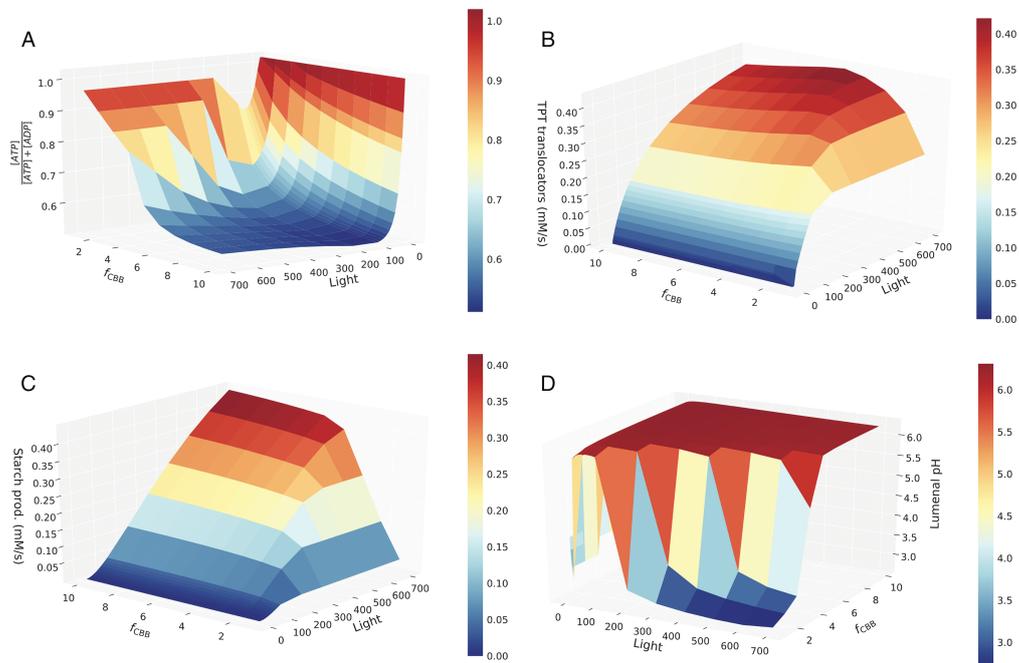
cycle in dark conditions for a positive carbon fixation balance over a day/night cycle. Indeed, key enzymes in the carbon fixation cycle are known to be regulated by the pH and the redox state of the chloroplast stroma. For example, RuBisCO activity is controlled by proton levels and magnesium ions (Tapia et al. 2000, Andersson 2008). Fructose-1,6-bisphosphatase, seduheptulose-1,7-bisphosphatase (SBPase) and Phosphoribulokinase are all controlled by the redox state through the thioredoxin-ferredoxin system, and also by pH (Chiadmi et al. 1999, Raines et al. 2000, Raines 2003). Furthermore, Hendriks et al. showed the light dependency of the ADP-glucose pyrophosphorylase (Hendriks et al. 2003), which is part of the lumped reaction  $v_{\text{Starch}}$  in our model. All these mechanisms will lead to a considerable reduction of the required stand-by flux of the CBB cycle, but are not yet included in our simple merged model.

In the original formulation of our model without constant Ru5P supply or light-dependent regulation of CBB enzymes, low light intensities lead to a rapid collapse of the cycle. However, in sufficient light ATP levels are very high and carbon fixation rates are already saturated in moderate light conditions (Fig. 2 and Fig. S1 in Appendix S1). These findings indicate that the sets of parameters for the carbon fixation enzymes and the light reactions, derived from the respective original publications, might not be suitably adapted when employed in a merged, cooperating, system. This is not surprising considering that they originate from completely different systems and conditions. However, we wish to highlight here that systems biology models are known to include a ‘sloppy’ spectrum of parameter sensitivities, and yet still provide robust predictions (Gutenkunst et al. 2007).

In order to systematically investigate the supply–demand behavior of the coupled system in

different light conditions, we introduce a ‘regulation factor’  $f_{\text{CBB}}$  of the CBB cycle, by which all  $V_{\text{max}}$ -values of the light-regulated enzymes (see above) are multiplied. This allows for a systematic variation of the energy demand by simulating accelerated or decelerated carbon fixation activity. Performing this variation under different light conditions gives insight into the synchronization of ATP and NADPH production and consumption rates, and thus enables a more profound analysis of the supply–demand regulation of photosynthesis (Brandes et al. 1996, Chiadmi et al. 1999, Raines et al. 2000). For the following steady-state analysis, the conceptual Ru5P influx reaction is not included.

Fig. 5 displays stationary values of key model variables for different light intensities and regulation factors. In agreement with the observations presented above, that very low light intensities lead to a collapse of the cycle, ATP concentrations (Fig. 5A) are maximal (zero ADP), triose phosphate export (Fig. 5B) and starch production (Fig. 5C) are zero, and the luminal pH (Fig. 5D) is very low (around 4). The latter is readily explained by the fact that the pH gradient built up by the low light cannot be reduced by the ATPase, which lacks the substrate ADP. Further, it becomes clear that the regulation factor of  $f_{\text{CBB}} = 1$ , corresponding to the original parameters, is far from optimal. The ATP:ADP ratio remains very high, and TPT export and starch production rates are well below their optimum, regardless of the light intensities. The stationary luminal pH further illustrates that parameters are not ideally adjusted. Not only for very low light, but also for moderate to high light conditions (above  $300 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) the lumen is dramatically acidic, indicating a mismatch in production and consumption processes. Increasing the regulation factor to values  $f_{\text{CBB}} \approx 4$  leads to a dramatic improvement of the performance of the system. The ATP:ADP ratio assumes



**Fig. 5.** Steady state analysis of the merged photosynthesis model under varying light intensities (x-axis) and carbon fixation velocities (y-axis). On the z-axis (A) the relative ATP abundance, (B) TPT export flux, (C) starch production rate and (D) luminal pH are displayed.

realistic and healthy values around one, triose phosphate export approximately doubles, and starch production increases by one order of magnitude compared with the original parameter values. Concomitantly, the luminal pH remains moderate (pH 5.8, as suggested by Kramer et al. 1999). An advantage of mathematical modelling is that one can also predict the behavior of system variables, which are not easy to obtain experimentally. In Fig. S6 in Appendix S1, we exemplarily depict oxidized ferredoxin, oxidized PQ, relative NADP<sup>+</sup> and violaxanthin levels.

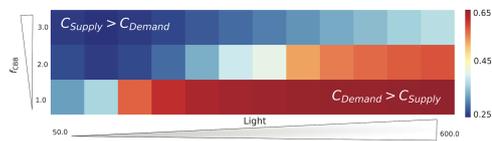
Quantitative analysis of the supply–demand behavior of the system can be performed by calculating flux control coefficients (Kacser and Burns 1973, Heinrich and Rapoport 1974). To investigate the relative overall flux control of supply and demand reactions, we first divide the set of all reactions in the model (R) into two non-overlapping sets S and D. S represents the supply set containing all PETC reactions and D represents the demand reaction set including all CBB cycle reactions. We define the overall control of supply ( $C_{\text{Supply}}$ ) and demand ( $C_{\text{Demand}}$ ) reactions as the sum of the absolute values of all control coefficients of reactions from S

and D, respectively, on the steady-state flux through the RuBisCO reaction,

$$C_{\text{Demand}} = \sum_{k \in D} |C_k^J| \quad (1)$$

$$C_{\text{Supply}} = \sum_{k \in S} |C_k^J|, \quad (2)$$

where  $C_k^J$  denotes the normalized control coefficient of reaction  $k$  on the steady-state carbon fixation rate. Fig. 6 displays the normalized overall control of demand reactions  $C_{\text{Demand}}/(C_{\text{Demand}} + C_{\text{Supply}})$ , in dependence on different light intensities and carbon fixation regulation factors. Low light intensities and fast carbon fixation reactions shift the overall flux control to the supply reactions. This can readily be explained because under these conditions (low light and fast CBB enzymes) energy and redox provision by the light reactions are the limiting factor. Interestingly, PSII and PSI contribute strongest to the overall flux control on the supply side (Fig. S7 in Appendix S1). Conversely, high light intensities and slow carbon fixation reactions shift the overall flux control to the demand side, because under these conditions, the



**Fig. 6.** Normalized overall control of the demand reactions ( $C_{\text{Demand}}$ ) under different light intensities (x-axis) and CBB cycle activities (y-axis). The results show how the control shifts from the demand reactions under high light conditions, but low CBB activity, to the supply, under low light conditions but faster CBB cycle.

system is energetically saturated, and the bottleneck is in the CBB cycle consuming the energy and redox equivalents. Noteworthy, it is the SBPase reaction that exhibits the highest overall flux control (Fig. S8 in Appendix S1), while RuBisCO has only minor control.

## Conclusions

Merging mathematical models is a highly non-trivial task. Even if two individual models yield plausible results, there is no guarantee that this is also true after mathematically combining these models. Besides pure technicalities, such as converting concentrations to appropriate units, there are a number of issues that make merging models challenging. Commonly, individual models have been developed with quite different scientific questions in mind, and may therefore display drastically different degrees of details of the involved processes. Moreover, parametrization is often performed for different organisms, tissues or conditions. Most importantly, increasing the system size by integrating two or more models may lead to novel emergent properties that were not observable in the individual models.

In this work, we have successfully merged a model of the PETC, supplying ATP and NADPH, to a model of the CBB cycle, consuming ATP and NADPH. The successful merge was largely facilitated by ensuring a comparable level of simplification of the two individual models (PETC described by 9 ODEs and CBB cycle by 15 ODEs). Our merged model represents a supply–demand system and as such exhibits systemic properties that did not exist in each of the individual models. Linking supply and demand processes into one functional model allowed us to employ metabolic control analysis for a systematic investigation of the regulatory dependence between the PETC and CBB cycle. By simulating the light–dark–light transitions, we could rationalize the importance of the oxidative PPP in providing substrates as a mechanism to operate the CBB cycle in a stand-by mode. Simultaneously, we illustrate that regulating the activity of the

CBB cycle in very low light is critical to avoid excessive investment into the stand-by mode. Moreover, the model demonstrates that regulation adapting to different light intensities is important to balance the supply by the PETC to the downstream demand. Using metabolic control analysis (MCA), we quantified the control distribution of supply and demand in the system for different light conditions and for varying CBB cycle activities. By introducing a regulation factor, corresponding to the CBB cycle enzyme activities, we demonstrate that the system requires higher input of light to obtain saturation for faster carbon fixation. Our MCA analysis showed that supply reactions exhibit high overall flux control when light is limited. Conversely, the demand reactions control the flux in light-saturating conditions. Among the supply reactions, the activity of PSII and PSI exhibit the highest overall flux control, while among the demand reactions, SBPase maintains the highest overall flux control (Figs S7 and S8 in Appendix S1). Interestingly, the often considered bottleneck enzyme RuBisCO exhibits only little overall flux control. This observation can be explained by the fact that the model assumes saturated  $\text{CO}_2$  conditions.

Our model is freely available as open source software, and we ensure that the results presented here can easily be reproduced. Because of its balanced simplicity and clear modular structure, we envisage that it serves as a platform for future development. Our model results have been obtained for specific experimental values of chlorophyll content and surface-to-volume ratios of the chloroplast stroma and lumen. However, these values strongly depend on the growth conditions and vary between plant species. The clear structure of the model and the documented code make it straight-forward to change model parameters for alternative experimentally determined values. Only relatively minor modifications of the model structure will be necessary to employ it for further analyses of the relationship between the PETC and other subprocesses. For instance, by describing starch as a dynamic variable and by providing a simplified representation of the oxidative PPP, one could improve our understanding of the light dependent turnover of starch (Stitt and Zeeman 2012) and rationalize the resupply of pentoses from hexoses in the chloroplast by the oxidative PPP (Neuhaus and Emes 2000, Kruger and Von Schaewen 2003) and investigate the role of alternative shunts (Preiser et al. 2018). In principle it is also straight-forward to simulate non-saturated carbon dioxide concentrations by modifying the RuBisCO rate equation accordingly (e.g. Witzel et al. 2010). However, under these conditions photorespiration can no longer be neglected. Therefore, for a realistic simulation of such scenarios, a simplified representation of

the photorespiratory pathway should be included in the model. With such an extension of the model, one could further investigate the energy balance (Igamberdiev et al. 2001) and the distribution of flux control between the PETC, the CBB cycle and the photorespiration reactions. Another model assumption is that all ATP produced by the light reactions is consumed by the CBB cycle. This, however, is of course only an approximation. Under severe stress conditions, or in different organisms, such as C4 plants or nitrogen-assimilating algae, this approximation is certainly not justified. Our implementation of the model in the modelbase environment (Ebenhöh et al. 2018) is designed to facilitate modifications in an intuitive way. Therefore, adding additional ATP consuming processes is technically simple, allowing theoretical investigations how such an additional demand will influence the behavior of the photosynthetic supply–demand system. The challenge here will be the derivation of realistic rate equations that describe the dependence of the additional ATP consumption rate on the ATP concentration and possibly other system variables.

The process of integrating two models described here illustrates the strength of theoretical approaches. Linking two processes leads to novel properties (here supply–demand balancing), which can be investigated to provide new fundamental insight. The merged model can rationalize the importance of systemic properties, and thus explain why certain mechanisms exist. In particular, none of the individual models could have explained the relevance of the stand-by mode or the role of adaptive regulation in maximizing efficiency, and thus explain the functional importance of the oxidative PPP or the redox and pH sensitivity of key CBB enzymes in a dynamic environment.

### Author contributions

A.M. merged the models and provided all mathematical descriptions. N.P.S. performed the computational analyses and prepared the first draft of the Results. All authors were involved in the interpretation of the results and preparation of the manuscript.

*Acknowledgements*—This study was funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy EXC 2048/1, Project ID: 390686111.

### References

Andersson I (2008) Catalysis and regulation in rubisco. *J Exp Bot* 59: 1555–1568  
 Brandes HK, Larimer FW, Hartman FC (1996) The molecular pathway for the regulation of

phosphoribulokinase by Thioredoxin f. *J Biol Chem* 271: 3333–3335  
 Cardona T, Shao S, Nixon PJ (2018) Enhancing photosynthesis in plants: the light reactions. *Essays Biochem* 62: 85–94  
 Chiadmi M, Navaza A, Miginiac-Maslow M, Jacquot JP, Cherfils J (1999) Redox signalling in the chloroplast: structure of oxidized pea fructose-1,6-bisphosphate phosphatase. *EMBO J* 18: 6809–6815  
 Christensen CD, Hofmeyr JHS, Rohwer JM (2015) Tracing regulatory routes in metabolism using generalised supply-demand analysis. *BMC Syst Biol* 9: 1–18  
 Ebenhöh O, Houwaart T, Lokstein H, Schleder S, Tirok K (2011) A minimal mathematical model of nonphotochemical quenching of chlorophyll fluorescence. *Biosystems* 103: 196–204  
 Ebenhöh O, Fucile G, Finazzi GG, Rochaix JD, Goldschmidt-Clermont M (2014) Short-term acclimation of the photosynthetic electron transfer chain to changing light: a mathematical model. *Philos Trans R Soc Lond B Biol Sci* 369: 20130223  
 Ebenhöh O, van Aalst M, Saadat NP, Nies T, Matuszyńska A (2018) Building mathematical models of biological systems with modelbase. *J Open Res Softw* 6: 24. <http://doi.org/10.5334/jors.236>  
 Farquhar GD, von Caemmerer S, Berry JA (2007) A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species. *Planta* 149: 78–90  
 Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol* 3: 1871–1878  
 Hahn BD (1986) A mathematical model of the Calvin cycle: analysis of the steady state. *Ann Bot* 57: 639–653  
 Hahn BD (1987) A mathematical model of photorespiration and photosynthesis. *Ann Bot* 60: 157–169  
 Heinrich R, Rapoport TA (1974 Feb) A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur J Biochem* 42: 89–95  
 Heinrich R, Schuster S (1996) *The Regulation of Cellular Systems*. Chapman & Hall, London  
 Hendriks JH, Kolbe A, Gibon Y, Stitt M, Geigenberger P (2003) ADP-glucose pyrophosphorylase is activated by posttranslational redox-modification in response to light and to sugars in leaves of *Arabidopsis* and other plant species. *Plant Physiol* 133: 838–849  
 Hofmeyr JHS, Cornish-Bowden A (2000) Regulating the cellular economy of supply and demand. *FEBS Lett* 476: 47–51  
 Igamberdiev AU, Bykova NV, Lea PJ, Gardeström P (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: 427–438

- Kacser H, Burns JA (1973) The control of flux. *Symp Soc Exp Biol* 27: 65–104
- Kramer DM, Sacksteder CA, Cruz JA (1999) How acidic is the lumen? *Photosynth Res* 60: 151–163
- Kruger NJ, Von Schaewen A (2003) The oxidative pentose phosphate pathway: structure and organisation. *Curr Opin Plant Biol* 6: 236–246
- Laisk A, Eichelmann H, Oja V (2006) C3 photosynthesis in silico. *Photosynth Res* 90: 45–66
- Matuszyńska AB, Heidari S, Jahns P, Ebenhöf O (2016) A mathematical model of non-photochemical quenching to study short-term light memory in plants. *Biochim Biophys Acta* 1857: 1860–1869
- Morales A, Kaiser E, Yin X, Harbinson J, Molenaar J, Driever SM, Struik PC (2018a) Dynamic modelling of limitations on improving leaf CO<sub>2</sub> assimilation under fluctuating irradiance. *Plant Cell Environ* 41: 589–604
- Morales A, Yin X, Harbinson J, Driever SM, Molenaar J, Kramer DM, Struik PC (2018b) In silico analysis of the regulation of the photosynthetic electron transport chain in C3 plants. *Plant Physiol* 176: 1247–1261
- Neuhaus HE, Emes MJ (2000) Nonphotosynthetic metabolism in plastids. *Annu Rev Plant Biol* 51: 111–140
- Ort DR, Merchant SS, Alric J, Barkan A, Blankenship RE, Bock R, et al. (2015) Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc Natl Acad Sci* 112: 201424031
- Pettersson G, Ryde-Pettersson U (1988) A mathematical model of the Calvin photosynthesis cycle. *Eur J Biochem* 175: 661–672
- Poolman MG, Fell DA, Thomas S (2000 feb) Modelling photosynthesis and its control. *J Exp Bot* 51: 319–328
- Preiser AL, Banerjee A, Fisher N, Sharkey TD (2018) Supply and consumption of glucose 6-phosphate in the chloroplast stroma. *bioRxiv*: 442434. <https://doi.org/10.1101/442434>
- Raines CA (2003) The Calvin cycle revisited. *Photosynth Res* 75: 1–10
- Raines CA, Harrison EP, Ölçer H, Lloyd JC (2000) Investigating the role of the thiol-regulated enzyme sedoheptulose-1,7-bisphosphatase in the control of photosynthesis. *Physiol Plant* 110: 303–308
- Rohwer JM, Hofmeyr JHS (2008) Identifying and characterising regulatory metabolites with generalised supply-demand analysis. *J Theor Biol* 252: 546–554
- Rubin A, Riznichenko G (2009) Modeling of the primary processes in a photosynthetic membrane. In: Laisk A, Nedbal L, Govindjee (eds) *Photosynthesis in silico: Understanding Complexity from Molecules to Ecosystems*, Vol. 29. Springer, Dordrecht, pp 151–176
- Stitt M, Zeeman SC (2012) Starch turnover: pathways, regulation and role in growth. *Curr Opin Plant Biol* 15: 282–292
- Tapia O, Oliva M, Safont VS, Andrés J (2000) A quantum-chemical study of transition structures for enolization and oxygenation steps catalyzed by Rubisco: on the role of magnesium and carbamylated Lys-201 in opening oxygen capture channel. *Chem Phys Lett* 323: 29–34
- Witzel F, Götze J, Ebenhöf O (2010) Slow deactivation of ribulose 1, 5-bisphosphate carboxylase/oxygenase elucidated by mathematical models. *FEBS J* 277: 931–950
- Zaks J, Amarnath K, Kramer DM, Niyogi KK, Fleming GR (2012) A kinetic model of rapidly reversible nonphotochemical quenching. *Proc Natl Acad Sci* 109: 15757–15762
- Zhu XG, de Sturler E, Long SP (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
- Zhu XG, Alba R, de Sturler E (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: 1490–1499
- Zhu XG, Wang Y, Ort DR, Long SP (2013 sep) E-photosynthesis: a comprehensive dynamic mechanistic model of C3 photosynthesis: from light capture to sucrose synthesis. *Plant Cell Environ* 36: 1711–1727

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Appendix S1.** Supplement.

Edited by J. Messinger

## **6.4 Computational Analysis of Alternative Photosynthetic Electron Flows Linked With Oxidative Stress**

### **6.4.1 Authors and Details**

Authors: Nima P. Saadat, Tim Nies, Marvin van Aalst, Brandon Hank, BÜsra Demirtas, Oliver Ebenhöh and Anna Matuszyńska  
Authorship: 1. Author  
Journal: Frontiers in Plant Science  
Status: Published (doi: 10.3389/fpls.2021.750580)

### **6.4.2 Contributions**

Nima P. Saadat helped building and implementing the model, producing results, writing the original draft results and discussion, editing the manuscript after review and visualized the results. Anna Matuszyńska initiated the project idea and conceptzualization, acquired funding, helped visualizing the results, building the model, performing formal analysis, writing the original draft discussion and editing the manuscript after review. Marvin van Aalst helped building the model and visualizing the results. Oliver Ebenhöh helped acquiring funding, building the model and writing original draft results. Tim Nies helped performing formal analyses, building the model and writing the original draft introduction. BÜsra Demirtas helped writing the original draft introduction. Brandon Hank helped performing formal analyses and writing the original draft methods.



# Computational Analysis of Alternative Photosynthetic Electron Flows Linked With Oxidative Stress

Nima P. Saadat<sup>1</sup>, Tim Nies<sup>1</sup>, Marvin van Aalst<sup>1</sup>, Brandon Hank<sup>1</sup>, Büsra Demirtas<sup>1</sup>, Oliver Ebenhöf<sup>1,2</sup> and Anna Matuszyńska<sup>1,2\*</sup>

<sup>1</sup> Institute of Quantitative and Theoretical Biology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany, <sup>2</sup> CEPLAS - Cluster of Excellence on Plant Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

## OPEN ACCESS

### Edited by:

Virginie Mengin,  
University of Essex, United Kingdom

### Reviewed by:

Jean-David Rochaix,  
Université de Genève, Switzerland  
Wenqiang Yang,  
Institute of Botany (CAS), China

### \*Correspondence:

Anna Matuszyńska  
anna.matuszynska@hhu.de

### Specialty section:

This article was submitted to  
Plant Metabolism and Chemodiversity,  
a section of the journal  
Frontiers in Plant Science

**Received:** 30 July 2021

**Accepted:** 21 September 2021

**Published:** 22 October 2021

### Citation:

Saadat NP, Nies T, van Aalst M, Hank B, Demirtas B, Ebenhöf O and Matuszyńska A (2021) Computational Analysis of Alternative Photosynthetic Electron Flows Linked With Oxidative Stress. *Front. Plant Sci.* 12:750580. doi: 10.3389/fpls.2021.750580

During photosynthesis, organisms respond to their energy demand and ensure the supply of energy and redox equivalents that sustain metabolism. Hence, the photosynthetic apparatus can, and in fact should, be treated as an integrated supply-demand system. Any imbalance in the energy produced and consumed can lead to adverse reactions, such as the production of reactive oxygen species (ROS). Reaction centres of both photosystems are known sites of ROS production. Here, we investigate in particular the central role of Photosystem I (PSI) in this tightly regulated system. Using a computational approach we have expanded a previously published mechanistic model of C3 photosynthesis by including ROS producing and scavenging reactions around PSI. These include two water to water reactions mediated by Plastid terminal oxidase (PTOX) and Mehler and the ascorbate-glutathione (ASC-GSH) cycle, as a main non-enzymatic antioxidant. We have used this model to predict flux distributions through alternative electron pathways under various environmental stress conditions by systematically varying light intensity and enzymatic activity of key reactions. In particular, we studied the link between ROS formation and activation of pathways around PSI as potential scavenging mechanisms. This work shines light on the role of alternative electron pathways in photosynthetic acclimation and investigates the effect of environmental perturbations on PSI activity in the context of metabolic productivity.

**Keywords:** reactive oxygen species, cyclic electron flow, mathematical model, photosynthesis, electron transport (photosynthetic)

## 1. INTRODUCTION

Photosynthetic organisms are the primary producers of biomass available in the biosphere. By employing complex biophysical processes, which act on multiple temporal and spatial scales, they perform highly efficient energy converting reactions (see for example Ksenzhek and Volkov, 1998). The basic machinery behind these reactions consists of two parts. The first one is the photosynthetic electron transport chain (PETC). Embedded in the thylakoid membrane, the PETC mediates the transfer of electrons, extracted from water molecules, over the complexes of Photosystem II (PSII), Cytochrome<sub>b<sub>6</sub>f</sub>, and Photosystem I (PSI) to the final electron acceptor NADP<sup>+</sup> via the mobile electron carriers plastoquinone (PQ), plastocyanin (PC), and ferredoxin (Fd). Thereby a proton gradient is formed, which is used to drive the synthesis of ATP by the ATP synthase. The second part of the photosynthetic process is the Calvin-Benson-Bassham (CBB) cycle, regulated by the

thioredoxin system (Geigenberger et al., 2017). NADPH and ATP produced by the PETC are used during the CBB cycle to fix CO<sub>2</sub> into organic compounds. Any imbalance between production and consumption can lead to adverse reactions, such as the production of reactive oxygen species (ROS) (Asada, 2006; Suzuki et al., 2012; Schwarzlander and Finkemeier, 2013) and affect the overall photosynthetic efficiency. Several sub-processes exist, distributed over the whole PETC, that contribute to the production of potentially toxic ROS compounds (Maurino and Flügge, 2008; Dietz et al., 2016; Khorobrykh et al., 2020).

To fine-adjust the formation of ATP and NADPH in the PETC, alternative electron transport pathways evolved (Curien et al., 2016). These alternative electron transport pathways are used to react immediately to changing environmental conditions (Alric and Johnson, 2017). Foremost, the cyclic electron flow (CEF) around PSI including the PGR5-PGRL1 mediated pathway is worth mentioning (Johnson, 2011). Studies have shown that CEF is essential for the functioning of photosynthesis (Munekage et al., 2004) and acts as a protective mechanism in fluctuating light conditions (Kono et al., 2014; Kono and Terashima, 2016). Alternative electron transport pathways balance the ATP and NADPH ratio to prevent an overexcitation of photosystems and redox imbalance in the PETC. Thus, the chance of forming toxic ROS is lowered. The Mehler reaction at PSI, which forms superoxide radicals O<sub>2</sub><sup>-</sup>, was extensively investigated in multiple species (Makino et al., 2002; Curien et al., 2016). Scavenging of ROS, for instance via the ascorbate-glutathione (ASC-GSH) cycle, is potentially an energy-demanding process (Das and Roychoudhury, 2014). However, it prevents physical damage inflicted on the molecular machinery of photosynthesis, which would be even more severe for the energy balance (for an analysis of costs associated with photoinhibition, see for example Raven, 2011). Multiple sophisticated regulatory mechanisms evolved to prevent the formation of ROS beforehand by lowering the energy pressure that acts on the PETC, such as non-photochemical quenching (NPQ) (see Müller et al., 2001).

Because of the existence and possible interaction of numerous mechanisms acting on different parts of the PETC, a system-wide investigation of the dynamics of photosynthesis is necessary. Existing evidence of the beneficial role of various water to water (W-W) cycles during photosynthesis (Curien et al., 2016) inspired us to investigate their impact on balancing the ATP to NADPH ratio. Computational kinetic models of photosynthesis have been proven to be useful for such analyses (Stirbet et al., 2020). Yet, none of these models investigated the role of ROS formation and scavenging. Our goal was to expand the existing model (Matuszyńska et al., 2019) of photosynthesis with key steps of both ROS formation and scavenging (via the ASC-GSH cycle) around PSI as well as linking the W-W

cycle with acclimation mechanisms. Moreover, based on our previous supply-demand analyses (Matuszyńska et al., 2019), we have included the regulation of key CBB enzymes through the thioredoxin system. This model thus provides the theoretical background to investigate non-trivial connections of the different components and to study complex systemic behaviour.

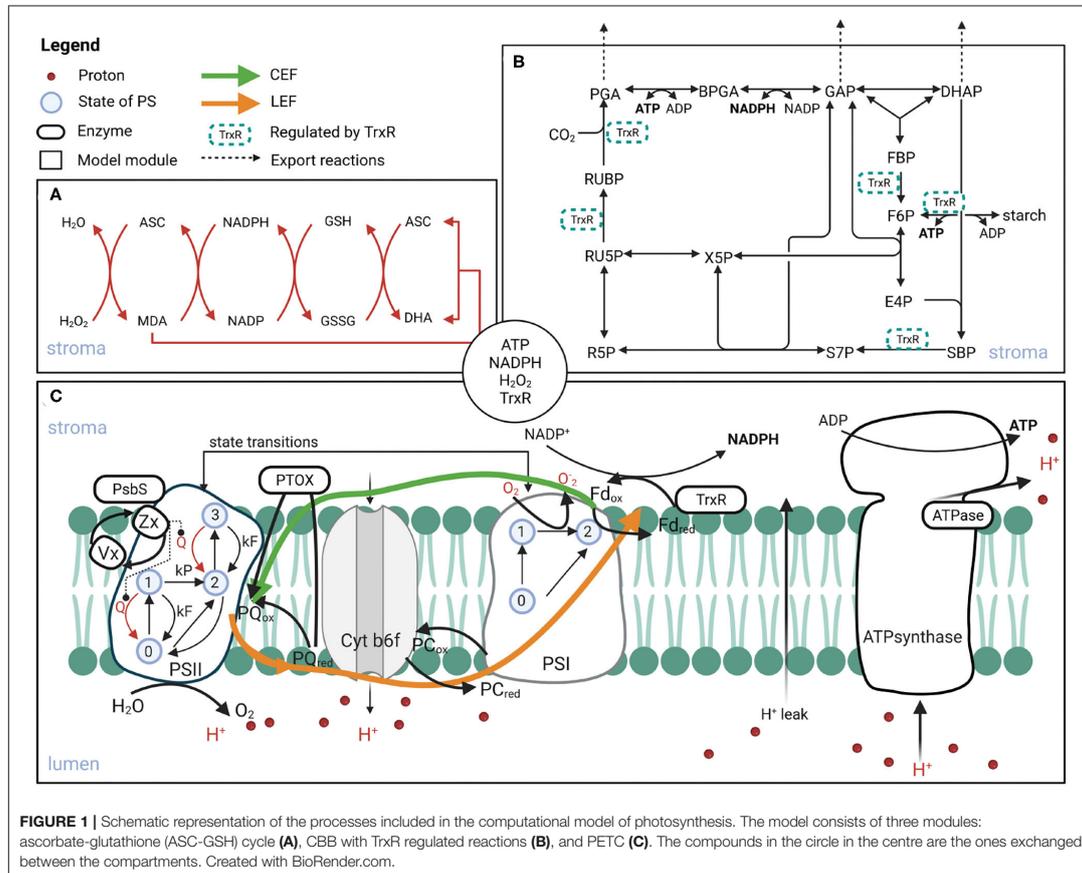
In this work we present the results of multiple analyses that allowed us to investigate the importance of alternative electron flows around PSI. We systematically investigated the impact of the Mehler reaction and the CEF on intermediate concentrations of both PETC and CBB cycle. We found out that some of the fluxes in the PETC are drastically influenced by the CEF. Therefore, we performed a Metabolic Control Analysis (MCA) that clearly showed a high impact of the Sedoheptulose-bisphosphate enzyme (SBPase) on the ROS scavenging mechanism, CBB and the PETC. Finally, the role of the SBPase was further elucidated. With this scientific work, we formalised a connection between the CBB cycle, PETC, and ASC-GSH cycle. We showed the interconnection between these parts of photosynthesis and also shed light on the control each part has over others via mathematical modelling. We therefore expanded our understanding of the complex interplay between different acclimatory processes in photosynthesis and created a computational framework to stimulate future scientific efforts in this direction.

## 2. METHODS

### 2.1. Model Description

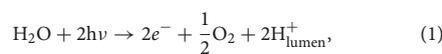
We have developed further the previously published mechanistic model of photosynthesis (Matuszyńska et al., 2019). The description of the demand side (**Figure 1B**) has been firstly complemented by including the thioredoxin reductase (TrxR) regulation. TrxR regulates the activation of the CBB-enzymes, depending on reduced Fd. Next, considering that the CBB cycle is the main, but not the only consumer of the energy equivalents produced by the PETC (**Figure 1C**), we included two reactions representing additional consumption of ATP and NADPH. Finally, the focus was put on adding two mechanisms responsible for the production and scavenging of ROS around PSI. An alternative electron transfer from PSI to oxygen has been included, leading to the production of superoxide which is rapidly converted to hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) by the superoxide dismutase (SOD). This implementation required changing the description of the PSI mechanism from the original model (Matuszyńska et al., 2019). Because of the rapid velocity of the SOD enzyme, the H<sub>2</sub>O<sub>2</sub> production is modelled as a single step, representing the Mehler reaction. We based our simplified description of the ROS scavenging reactions on the published kinetic models of the ASC-GSH cycle by Valero et al. (2009, 2015). Our description of the cycle is represented by four saturating enzymatic reactions [mediated by ascorbate peroxidase (APX), monodehydroascorbate reductase (MDAR), dehydroascorbate reductase (DHAR), glutathione reductase (GR)] and one spontaneous disproportionation of monodehydroascorbate radicals (MDA), see **Figure 1A**. The pools of ASC and GSH are considered constant.

**Abbreviations:** ASC, ascorbate; CBB, Calvin-Benson-Bassham; CEF, cyclic electron flow; OE, overexpressor; Fd, ferredoxin; GSH, glutathione; LEF, linear electron flow; KD, knock down; KO, knock out; MCA, Metabolic Control Analysis; MDA, monodehydroascorbate radicals; PETC, photosynthetic electron transport chain; PC, plastocyanin; PPF, photosynthetic photon flux density; PSI, photosystem I; PSII, photosystem II; PTOX, Plastid terminal oxidase; PQ, plastoquinone; ROS, reactive oxygen species; SOD, superoxide dismutase; TrxR, thioredoxin reductase; W-W, water-water.



### 2.1.1. Linear and Alternative Electron Flows

The rates of electron flow through various pathways are directly calculated from the rates through PSII and FNR. In the model, the stoichiometry of the rate of PSII is



which produces 2 electrons. Therefore, the rate of linear electron flow (LEF) is twice the simulated rate through PSII. Likewise, the rate of CEF is twice the rate mediated by FNR.

### 2.1.2. Units

The choice of units is the same as in Matuszyńska et al. (2019), keeping the original units of stromal and luminal compartments. The concentrations in the lumen are expressed in mmol (mol Chl)<sup>-1</sup> and inside the stroma in mM. To convert the concentrations of ATP, NADPH and H<sub>2</sub>O<sub>2</sub> produced in the lumen to the unit of the stroma, where these metabolites are consumed/scavenged, we employ a conversion factor where

1 mmol (mol Chl)<sup>-1</sup> corresponds to 3.2 · 10<sup>-5</sup> M in the stroma (Laisk et al., 2006).

## 2.2. Computational Analysis

The mathematical model is a system of 30 ordinary differential equations with 46 reaction rates. The model was integrated with Assimulo (Andersson et al., 2015) via the Python-based software modelbase version 1.3.8 (van Aalst et al., 2021). Python files containing the model and Jupyter notebooks with our simulations used to produce all figures are provided on our GitLab repository <https://gitlab.com/qtb-hhu/models/cyclicphotosyn-2021>.

### 2.2.1. Metabolic Control Analysis

Flux ( $C_{v_k}^J$ ) and concentration ( $C_{v_k}^S$ ) control coefficients are defined as

$$C_{v_k}^J = \frac{v_k}{J} \frac{\partial J_j / \partial p}{\partial v_k / \partial p} \quad (2)$$

$$C_{v_k}^{S_j} = \frac{v_k}{S_j} \frac{\partial S_j / \partial p}{\partial v_k / \partial p}, \quad (3)$$

where  $J_j$  and  $S_j$  are respectively the steady-state fluxes and concentrations of the system,  $p$  is a kinetic parameter which affects directly only reaction  $k$  with the rate  $v_k$  (see Kacser and Burns, 1973; Heinrich and Rapoport, 1974; Heinrich and Schuster, 1996). We approximated these formulas numerically using the central difference, varying the parameters by  $\pm 1\%$ . Control coefficients quantify the relative effect of a parameter perturbation on steady state fluxes and concentrations.

### 3. RESULTS

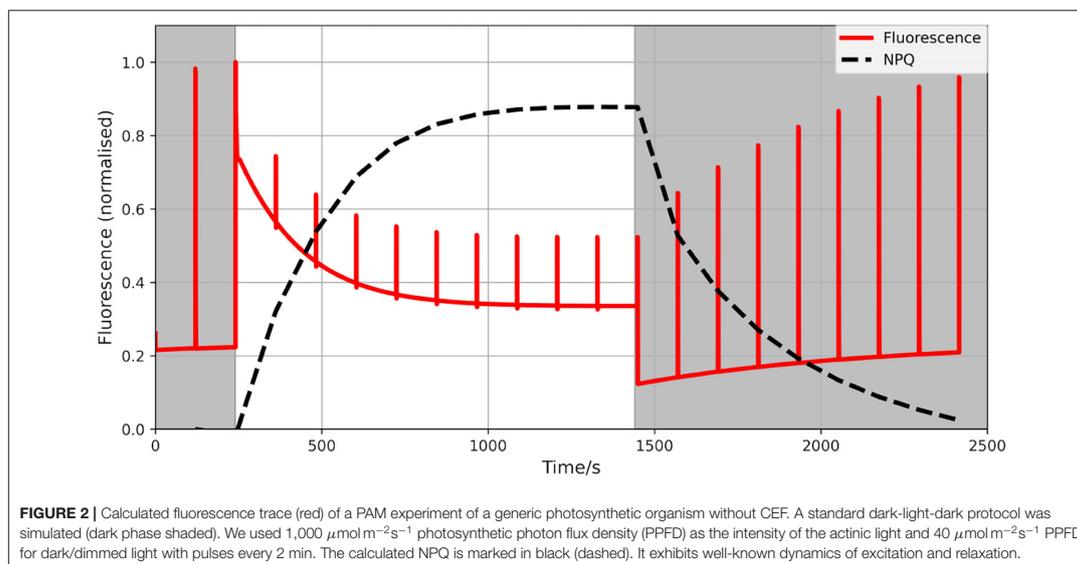
The model has been used to study electron flows around PSI and their relevance to the overall performance of the photosynthetic machinery under both steady-state and dynamic conditions. To confirm that our improved model can indeed be used beyond steady-state and can realistically reproduce short-term acclimation responses we simulated a standard PAM fluorescence trace. The results exhibit typical fluorescence dynamics under high light conditions (Figure 2). It should be however noted that quantities discussed here should not be understood as precise predictions of a specific experimental observations, but are rather meant to illustrate the general plausibility of the model behaviour.

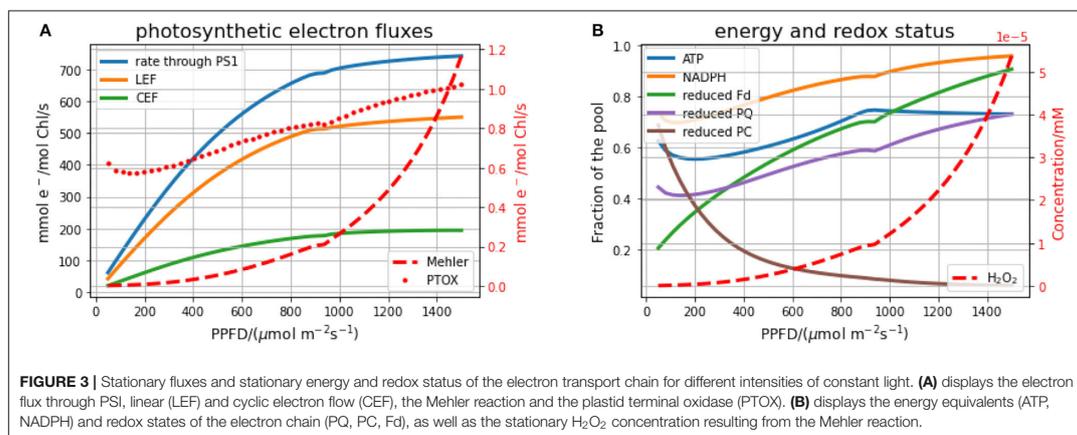
#### 3.1. Steady-State Behaviour Under Continuous Light

We first investigate the steady-state behaviour of the model under various light intensities (Figure 3). In the left panel (Figure 3A), the stationary electron fluxes over different light

intensities through the PSI, LEF, CEF, the Mehler reaction and the plastid terminal oxidase (PTOX) are depicted. The rate of the electron transport chain increases linearly for low light conditions and saturates in high light. Carbon fixation rates follow the same general pattern (see **Supplementary Material**), which has been repeatedly confirmed in experiments for a wide range of photosynthetic organisms (Hesketh and Baker, 1967; Huang et al., 2016). In our simulations, the transition to the light-saturated regime occurs around a photosynthetic photon flux density (PPFD) of  $900 \mu\text{mol m}^{-2}\text{s}^{-1}$ , which is in good agreement with previously observed and modelled values (Kromdijk et al., 2019). In contrast to the electron transport chain, the rate of the Mehler reaction strongly increases in high light conditions, leading also to increased stationary hydrogen peroxide ( $\text{H}_2\text{O}_2$ ) concentrations (Figure 3B). Nevertheless, even in high light, the rate of the electron transfer to oxygen by the Mehler reaction reaches only around 0.2% of the electrons transferred by PSI. This means that even under high light, less than 1% of the NADPH produced by the electron transport chain is required to scavenge the ROS produced in PSI through the Mehler reaction. Most redox carriers are more reduced in high light, with the exception of PC, which is more oxidised in higher light. This observation can be explained by the fact that more light increases the rate of PSI, which directly removes electrons from the PC pool. This explanation is supported by the results of the MCA, indicating that increased PSI leads to a more oxidised PC pool (see also Figure 6).

A key enzyme in the ASC-GSH cycle is the MDA reductase, which reduces MDA back to ASC using NADPH as an electron donor (Figure 1A). Interestingly, a simulated knock-down of this enzyme to 1% of its original value does not affect the overall electron fluxes. However, in high light, the





deficiency in MDA reductase is compensated by the spontaneous disproportionation of MDA into DHA and ASC, which leads to approximately 100-fold increased levels of the MDA radical (see **Supplementary Material**). Because overall electron fluxes and  $H_2O_2$  production rates are not affected, also the ratio of NADPH required for scavenging ROS is unaltered in the MDA reductase knock-down.

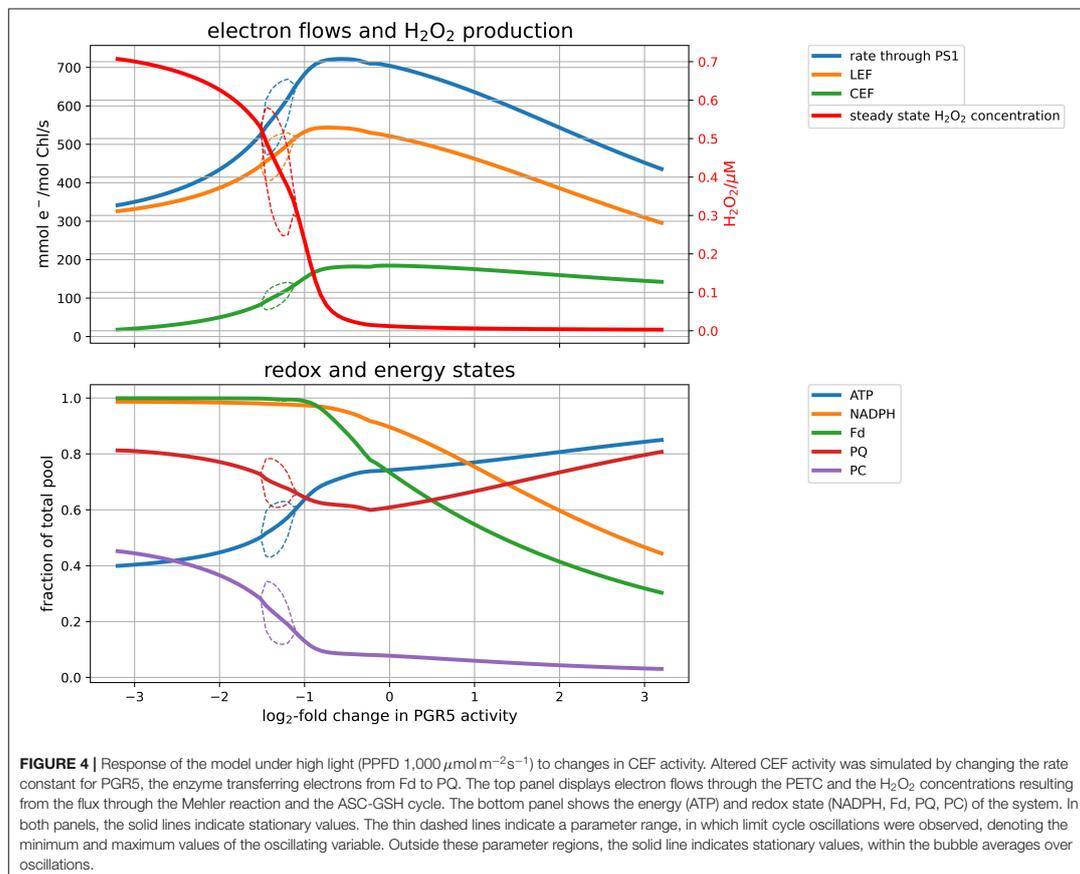
### 3.2. Performance of PGR5 Mutants Under Continuous Light

By transferring electrons from Fd back into the PQ pool, the protein PGR5 mainly mediates the CEF. We employed our model to study how altering the CEF affects electron flows and downstream metabolism, by systematically varying the corresponding enzyme activity (**Figure 4**) under simulated high light conditions ( $PPFD\ 1000\ \mu mol\ m^{-2}\ s^{-1}$ ). These simulations correspond to knocking down (KD) or overexpressing (OE) the PGR5 protein, which catalyses the reduction of PQ by Fd. Slowing down CEF does not only result in a slower CEF rate but also leads to a reduced overall photosynthetic electron flux and carbon fixation rate (top panel of **Figure 4**). This behaviour illustrates the physiological role of CEF to adjust the ATP/NADPH ratio produced by the PETC to the downstream demand. Because the provided ratio does not align with downstream demand, electrons accumulate in the final products of the PETC, leading to over-reduced Fd and NADPH pools (lower panel of **Figure 4**). Over-reduced Fd, in turn, reduces the availability of electron acceptors for PSI, which leads to an increased rate of the Mehler reaction and  $H_2O_2$  levels. The reduced photosynthetic capacity of PGR5 mutant plants has been demonstrated experimentally (DalCorso et al., 2008). A simulated knockout (KO) quantitatively reproduces the observation that maximal PSII rate is approximately half of the wildtype ( $\sim 300$  vs.  $520\ mmol\ e^{-}/mol\ Chl/s$  in **Figure 4**), and that light saturation is reached at lower intensities compared to the wildtype (approximately at  $PPFD\ 500\ \mu mol\ m^{-2}\ s^{-1}$ —see **Supplementary Material**). Also increasing the CEF has negative

effects on the performance. If more electrons are re-inserted into the PETC, the overall ATP level increases and electron carriers are less reduced, but the overall production rate of NADPH and ATP decreases, leading again to a reduced carbon fixation rate. It seems, therefore, that there exists an optimal PGR5 activity, that maximises photosynthetic efficiency and carbon fixation by avoiding over-reduction of the electron chain, while at the same time redirecting not more electrons than necessary back into the chain. Under low light (for figures, see **Supplementary Material**), the CEF plays a less important role. Under these conditions, increasing PGR5 activity increases the ratio of CEF to LEF and slightly decreases carbon fixation rates. The simulations suggest that, whereas under high light CEF is clearly beneficial for the photosynthetic efficiency, under low light conditions a low PGR5 activity is favourable for  $CO_2$  fixation rates.

### 3.3. Importance of Alternative Electron Flows Under Fluctuating Light

It was repeatedly demonstrated experimentally that the CEF is particularly important to maintain photosynthetic activity under fluctuating light conditions (Yamori et al., 2016; Yamamoto and Shikanai, 2018). Comparing simulations of wildtype with PGR5 mutant shows that carbon fixation is indeed drastically reduced when no CEF operates (**Figure 5**). These results are in qualitative agreement with experimental findings (Yamori et al., 2016). However, the experimentally observed dynamics are quantitatively different from our simulations. In particular, the reactivation dynamics of RuBisCO in the transition to high light are considerably slower in the experiment as compared to the model simulations. This indicates that the mechanisms activating the CBB cycle in such transitions are not yet represented in the model in a quantitatively correct way. Still, the model provides a theoretical explanation for the reduced photosynthetic efficiency by illustrating that the PGR5 mutant is unable to establish a healthy redox balance in light periods.

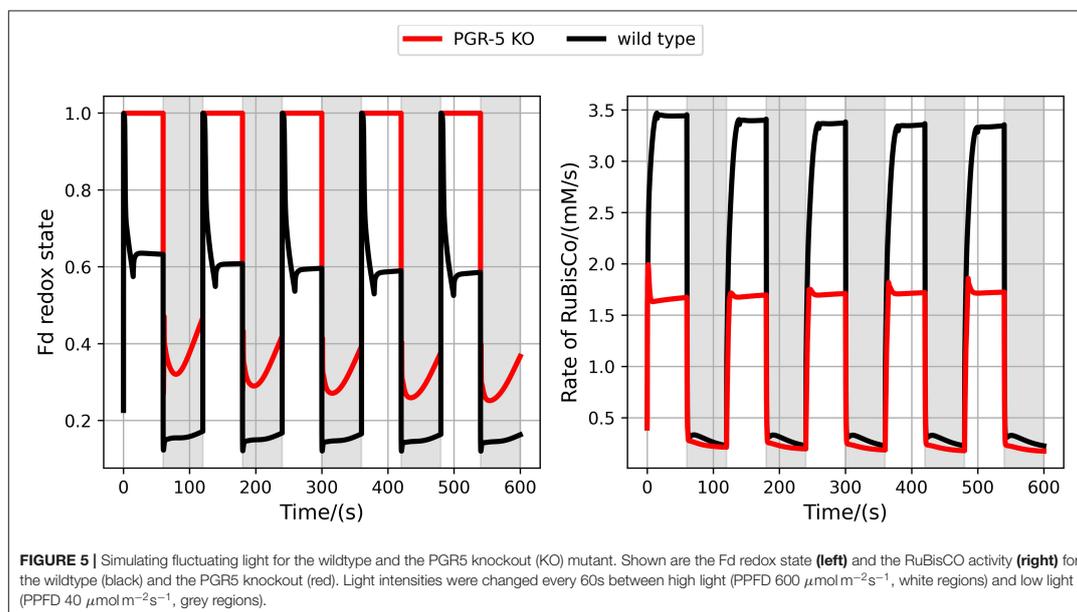


### 3.4. SBPase Exhibits Striking Control Over Photosynthesis Under High Light

The above investigations illustrate that electron flow around PSI apparently affects not only the PETC itself but also downstream metabolism, in particular carbon fixation. In order to understand which processes carry the strongest control in this complex supply-demand system, we performed MCA and systematically determined flux and concentration control coefficients for high (PPFD  $1000 \mu\text{mol m}^{-2} \text{s}^{-2}$ ) and low (PPFD  $100 \mu\text{mol m}^{-2} \text{s}^{-2}$ ) light conditions. A selection of flux and concentration control coefficients are depicted in **Figure 6**. Additionally to get a global picture of the model's behaviour we performed a simple global sensitivity analysis using Latin Hyperspace Sampling and Partial Rank Correlation Coefficients that can be found in the **Supplementary Material**.

In agreement with the analysis of the effects of perturbing PGR5 activity, and thus CEF (**Figure 4**), it is observed that increasing PGR5 leads to slightly decreased fluxes in the PETC and the CBB cycle. In contrast, increasing CEF strongly decreases

the Mehler reaction and the associated scavenging pathways. Remarkably, under high light, the strongest control on PETC and CBB cycle fluxes is exhibited by the SBPase, whereas RuBisCO carries almost no flux control. This observation confirms previous theoretical results obtained from a model simulating the CBB cycle alone (Poolman et al., 2000). Increasing SBPase results in a significant increase of both PETC and CBB cycle rates, and strongly suppresses the Mehler reaction and associated scavenging reactions, while the redox pools except PC are more oxidised, and ATP levels are decreased. PSII is the initial complex of the PETC and thus a natural candidate for high flux control. Indeed, it exerts positive control over PETC and CBB cycle fluxes in high light, but with a much lower control strength compared to SBPase. Increasing PSII (and PSI and to a lesser extent the cytochrome  $b_6f$  complex) predominantly increases the Mehler reaction. This behaviour changes dramatically under low light. Here, CBB enzymes exert almost no flux control, but electron transport and carbon fixation rates are mostly controlled by the activities of the photosystems. Increasing PSII leads



to more reduced redox pools and lower ATP levels, whereas increasing PSI leads to more oxidised redox pools and higher ATP levels. Both photosystems have a positive control on CBB cycle intermediates RuBP and PGA, while only PSI positively affects the bisphosphates FBP and SBP. Altogether these analyses confirm the previous observation (Matuszyńska et al., 2019) that under low light control resides predominantly on the supply side (PETC), while under high light control is shifted toward the demand side (CBB).

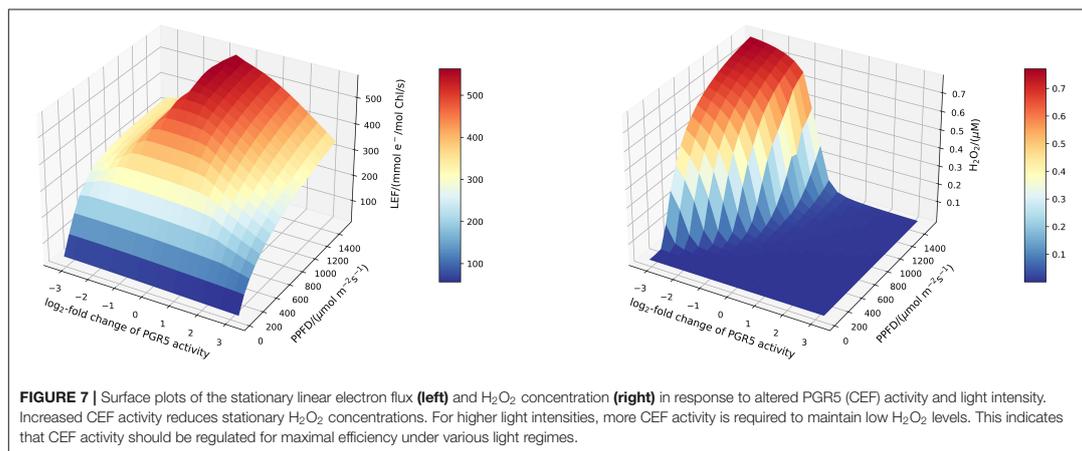
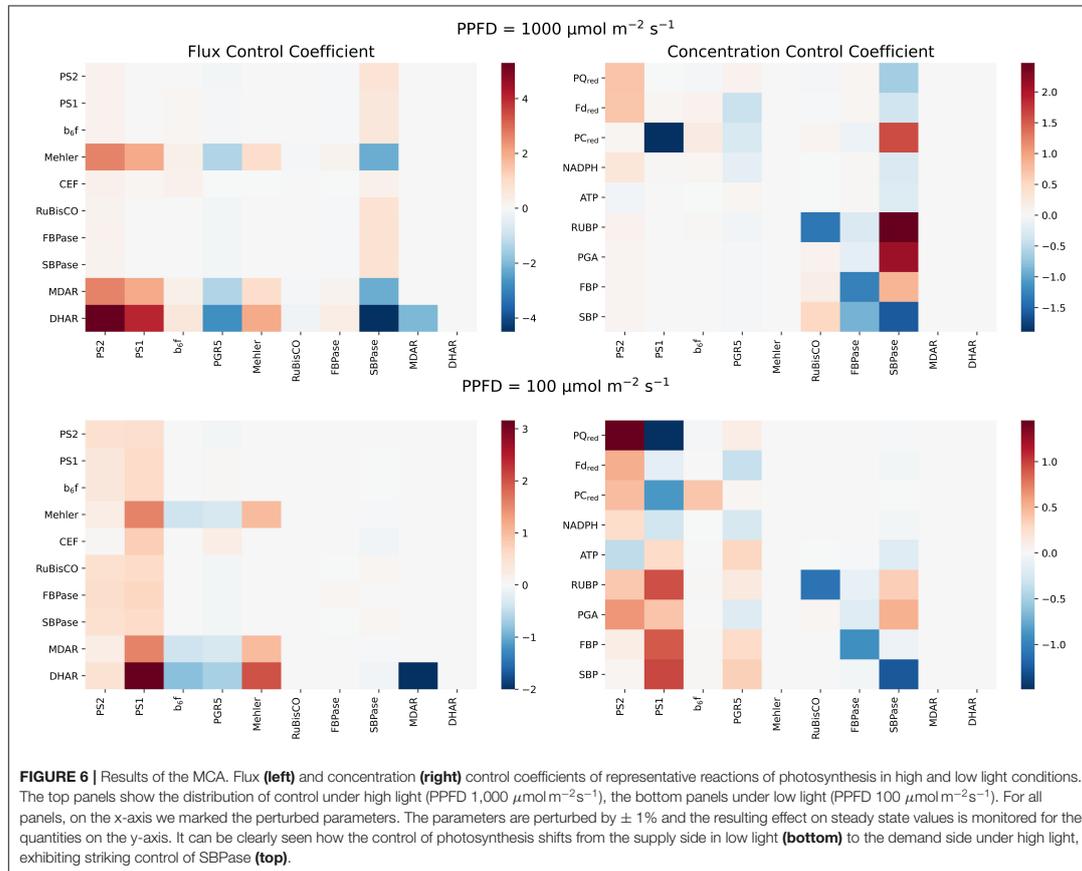
### 3.5. ROS Production as a Balancing Mechanism

To increase our understanding of the antagonistic behaviour of the Mehler reaction and the CEF, and to account for the changing relative importance of these processes under low and high light, we systematically investigated the efficiency of photosynthesis for altered CEF under different light intensities. **Figure 7** displays simulated linear electron fluxes and  $\text{H}_2\text{O}_2$  concentrations in response to changed light intensities and PGR5 activities. Whereas under low light conditions (of less than approximately 500  $\mu\text{mol m}^{-2}\text{s}^{-1}$ , the photosynthetic efficiency is rather independent of the PGR5 activity, this is dramatically different in high light. Both, too low and too high CEF activity leads to a reduced photosynthetic flux, but for different reasons. Impaired CEF results in drastically elevated  $\text{H}_2\text{O}_2$  levels because ATP and NADPH production ratios cannot be adapted to the downstream requirements. In contrast, increased activity of PGR5 mediated CEF simply leads to more oxidised NADPH and

Fd (see **Figure 5**), and redirects electron flux from linear to cyclic, thus reducing the overall net carbon fixation rate.

## 4. DISCUSSION

In oxygenic photosynthesis, LEF is considered the basic driver of photosynthetic carbon fixation. Yet alone, it does not provide the exact ratio of ATP to NADPH that is necessary to drive carbon assimilation (Kramer and Evans, 2009). Hence, alternative circuits of the electron flow are considered to balance the production of ATP per NADPH (Curien et al., 2016). In this work the presented computational model has been developed to investigate the alternative electron circuits around PSI that produce a proton gradient without NADPH synthesis, therefore altering this ratio. These include the CEF around PSI and two of the W-W cycle including the Mehler reaction at PSI and the PTOX downstream PSII (Curien et al., 2016). Additionally, we have provided an important link between ROS formation and metabolism regulation by including a simple description of ROS scavenging around PSI via the ASC-GSH cycle. This allowed us to further investigate the role of the cycle in keeping photosynthetic activity at medium and higher light intensities (Muller-Schusselle et al., 2020). Although it is only one of the many known pathways (Maurino and Flüggé, 2008), it is considered as the first step in the long process of including redox balance through ROS production into computational models of photosynthesis, in an attempt to support the synthetic redesign of photosynthetic systems (Zhu et al., 2020).



We have argued before that photosynthesis shall be viewed as a supply-demand system because of the connection between the ATP and NADPH production and consumption (Matuszyńska et al., 2019). Considering the tight regulation of such a system, we investigated the influence of alternative electron pathways on the rate of CO<sub>2</sub> assimilation, with a particular focus on their photoprotective behaviour and the role of the CEF (see change in rate of RuBisCO in **Figure 5**). The presented model is intended to serve as a theoretical workbench that is not only valid for a single experiment or plant species but is in principle adaptable to a wide range of scenarios and photosynthetic organisms. While not precisely calibrated to a particular experimental dataset, we ensured that the model displays realistic behaviour. In particular, the steady-state of key variables, such as the redox state of electron carriers as well as carbon fixation fluxes are plausible, and the simulated PAM experiments show characteristic NPQ dynamics (**Figure 2**). The model allows moreover the simulation of genetic perturbations, such as KO, KD and OE, which has been demonstrated extensively on the PGR5 mutant, impaired with its capacity of a CEF. The focus on the PGR5/PGRL1 pathway was motivated by its particular role in regulating proton motive force around PSI (Wang et al., 2015). **Figure 4** highlights the critical role of the CEF by displaying a strongly reduced LEF, highly oxidised redox state of the electron carriers and a very strong increase in hydrogen peroxide concentration. Interestingly, our computational analysis systematically displayed the dependency of the system behaviour in PGR5 KO and OE to different light intensities. The differences between PGR5 mutants are mostly visible in higher light conditions, as shown in **Figure 7**.

Light, although necessary to drive photosynthesis, can be also harmful to the organism. NPQ is a central part of the first line of defence of plants against damaging effect of light. In order to prevent high ROS levels, plants developed mechanisms allowing dissipation of excess light energy as heat (Ruban, 2016). Our simulations demonstrate that in high light intensity the whilst Mehler and PTOX reactions continue to increase, contributing significantly to the photoprotection and overall redox balance (**Figure 3**). These results are in line with the previously proposed role of the W-W cycle acting as a relaxation system to suppress the photoproduction of <sup>1</sup>O<sub>2</sub> in PSII (Asada, 2006). We expect that the model presented in this work will be useful for a systematic assessment of the possible beneficial effect of ROS formation in a physiological context (Foyer and Noctor, 2005; Foyer, 2018; Mhamdi and Van Breusegem, 2018).

Within our expanded model of photosynthesis we have performed MCA and confirmed the pivotal role of SBPase in control over the system, as in our previous work (Poolman et al., 2000; Matuszyńska et al., 2019). SBPase has been shown to control both supply and demand of photosynthesis and, consequently, in this expanded model, it exhibits a strong influence on the electron flows. **Figure 6** displays that in high light conditions, an increase of SBPase activity strongly decreases the Mehler reaction rate and as a consequence the rate of the main scavenging reactions DHAR and MDAR. This phenomenon can be explained by the increase in efficiency of the CBB cycle, which causes faster ATP consumption and prevents over-reduction of the PETC, therefore reducing the rate and impact

of the Mehler reaction. It is important to consider that this behaviour is observed in scenarios with saturated carbon dioxide conditions. However, the model can in principle be directly applied to other, more natural, conditions. For example, it would be interesting to compare the electron flux distribution under non-saturating conditions. Further, although we have varied oxygen systematically, to mimic conditions under which oxygen becomes limiting, all our analysis have been performed under saturated CO<sub>2</sub> conditions.

A natural further step of expanding this work would be to include the mechanism of photorespiration, mainly because it plays a physiological role in reducing the redox pressure in the stroma under conditions leading to low carbon fixation (Ort and Baker, 2002) and because it is a major source of ROS associated with the photosynthetic activity (Dietz et al., 2016). A reliable mathematical model of photorespiration to be considered has been proposed by Yokota et al. (1985).

With this work we provide a tool to further study the dynamics and cross-talk between the multiple regulatory mechanisms activated by photosynthetic organisms in response to changes in light. With our model, we could demonstrate how electron flows around PS1 affect photosynthetic efficiency and how increasing CBB cycle activity decreases Mehler reaction activity. Moreover, the model allowed us to rationalise that CEF should be regulated with changing light intensities as a trade-off between optimising electron flux efficiency and minimising ROS production. We envisage that this model helps to further investigate the tight relation between ROS scavenging in the chloroplast and the dynamic adaptation of photosynthesis to changing conditions.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

AM: initial idea and conceptualisation. AM and OE: funding acquisition. AM, MA, and NS: visualisation. AM, OE, MA, NS, and TN: model building. AM, BH, OE, MA, NS, and TN: formal analyses. BD and TN: writing—original draft and introduction. BH: writing—original draft and methods. OE and NS: writing—original draft and results. AM and NS: writing—original draft, discussion, and writing—review and editing. All authors read and accepted the final version of the manuscript.

## FUNDING

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2048/1 - project ID 390686111 (AM and OE), the Deutsche Forschungsgemeinschaft Research Grant MA 8103/1-1 (AM), the Deutsche Forschungsgemeinschaft (DFG), project ID 391465903/GRK 2466 (TN), and the EU's

Horizon 2020 research and innovation programme under the Grant Agreement 862087 (MA).

## ACKNOWLEDGMENTS

We would like to thank Veronica Maurino for a fruitful discussion on ROS at the beginning of this project and all members of the Photosynthesis Task Force at the Institute

of Quantitative and Theoretical Biology whose ideas and engagement was essential for this work.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.750580/full#supplementary-material>

## REFERENCES

- Alic, J., and Johnson, X. (2017). Alternative electron transport pathways in photosynthesis: a confluence of regulation. *Curr. Opin. Plant Biol.* 37, 78–86. doi: 10.1016/j.pbi.2017.03.014
- Andersson, C., Claus, F., and Akesson, J. (2015). ScienceDirect Assimulo: a unified framework for ODE solvers. *Math. Comput. Simul.* 116, 26–43. doi: 10.1016/j.matcom.2015.04.007
- Asada, K. (2006). Production and scavenging of reactive oxygen species in chloroplasts and their functions. *Plant Physiol.* 141, 391–396. doi: 10.1104/pp.106.082040
- Curien, G., Flori, S., Villanova, V., Magneschi, L., Giustini, C., Forti, G., et al. (2016). The water to water cycles in microalgae. *Plant Cell Physiol.* 57, 1354–1363. doi: 10.1093/pcp/pcw048
- DalCorso, G., Pesaresi, P., Masiero, S., Aseva, E., Schunemann, D., Finazzi, G., et al. (2008). A complex containing PGRL1 and PGR5 is involved in the switch between linear and cyclic electron flow in Arabidopsis. *Cell* 132, 273–285. doi: 10.1016/j.cell.2007.12.028
- Das, K., and Roychoudhury, A. (2014). Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Front. Environ. Sci.* 2:53. doi: 10.3389/fenvs.2014.00053
- Dietz, K.-J., Turkan, I., and Krieger-Liszka, A. (2016). Redox- and reactive oxygen species-dependent signaling into and out of the photosynthesizing chloroplast. *Plant Physiol.* 171, 1541–1550. doi: 10.1104/pp.16.00375
- Foyer, C. H. (2018). Reactive oxygen species, oxidative signaling and the regulation of photosynthesis. *Environ. Exp. Bot.* 154, 134–142. doi: 10.1016/j.envexpbot.2018.05.003
- Foyer, C. H., and Noctor, G. (2005). Oxidant and antioxidant signalling in plants: a re-evaluation of the concept of oxidative stress in a physiological context. *Plant Cell Environ.* 28, 1056–1071. doi: 10.1111/j.1365-3040.2005.01327.x
- Geigenberger, P., Thormählen, I., Daloso, D. M., and Fernie, A. R. (2017). The unprecedented versatility of the plant thioredoxin system. *Trends in Plant Science* 22, 249–262. doi: 10.1016/j.tplants.2016.12.008
- Heinrich, R., and Rapoport, T. A. (1974). A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur. J. Biochem.* 42, 89–95. doi: 10.1111/j.1432-1033.1974.tb03318.x
- Heinrich, R., and Schuster, S. (1996). *The Regulation of Cellular Systems*. London: Chapman & Hall. doi: 10.1007/978-1-4613-1161-4
- Hesketh, J., and Baker, D. (1967). Light and carbon assimilation by plant communities I. *Crop Sci.* 7, 285–293. doi: 10.2135/cropsci1967.0011183X000700040002x
- Huang, W., Hu, H., and Zhang, S.-B. (2016). Photosynthesis and photosynthetic electron flow in the alpine evergreen species *Quercus guyavifolia* in winter. *Front. Plant Sci.* 7:1511. doi: 10.3389/fpls.2016.01511
- Johnson, G. N. (2011). Physiology of psi cyclic electron transport in higher plants. *Biochim. Biophys. Acta Bioenerg.* 1807, 384–389. doi: 10.1016/j.bbabi.2010.11.009
- Kacser, H., and Burns, J. A. (1973). The control of flux. *Symp. Soc. Exp. Biol.* 27, 65–104.
- Khorobrykh, S., Havurinne, V., Mattila, H., and Tyystjärvi, E. (2020). Oxygen and ROS in photosynthesis. *Plants* 9:91. doi: 10.3390/plants9010091
- Kono, M., Noguchi, K., and Terashima, I. (2014). Roles of the cyclic electron flow around PSI (CEF-PSI) and O<sub>2</sub>-dependent alternative pathways in regulation of the photosynthetic electron flow in short-term fluctuating light in *Arabidopsis thaliana*. *Plant Cell Physiol.* 55, 990–1004. doi: 10.1093/pcp/pcu033
- Kono, M., and Terashima, I. (2016). Elucidation of photoprotective mechanisms of PSI against fluctuating light photoinhibition. *Plant Cell Physiol.* 57, 1405–1414. doi: 10.1093/pcp/pcw103
- Kramer, D. M., and Evans, J. R. (2009). The importance of energy balance in improving photosynthetic productivity. *Plant Physiol.* 155, 70–78. doi: 10.1104/pp.110.166652
- Kromdijk, J., Glowacka, K., and Long, S. P. (2019). Predicting light-induced stomatal movements based on the redox state of plastoquinone: theory and validation. *Photosynth. Res.* 141, 83–97. doi: 10.1007/s11120-019-00632-x
- Ksenzhek, O. S., and Volkov, A. G. (1998). *Plant Energetics*. San Diego, CA: Academic Press, Elsevier.
- Laisk, A., Eichelmann, H., and Oja, V. (2006). C3 photosynthesis *in silico*. *Photosynth. Res.* 90, 45–66. doi: 10.1007/s11120-006-9109-1
- Makino, A., Miyake, C., and Yokota, A. (2002). Physiological functions of the water–water cycle (Mehler reaction) and the cyclic electron flow around PSI in rice leaves. *Plant Cell Physiol.* 43, 1017–1026. doi: 10.1093/pcp/pcf124
- Matuszyńska, A., Saadat, N. P., and Ebenhö, O. (2019). Balancing energy supply during photosynthesis—a theoretical perspective. *Physiol. Plant.* 166, 392–402. doi: 10.1111/ppl.12962
- Maurino, V., and Flügge, U.-I. (2008). Experimental systems to assess the effects of reactive oxygen species in plant tissues. *Plant Signal. Behav.* 3, 923–928. doi: 10.4161/psb.7036
- Mhamdi, A., and Van Breusegem, F. (2018). Reactive oxygen species in plant development. *Development* 145:dev164376. doi: 10.1242/dev.164376
- Müller, P., Li, X.-P., and Niyogi, K. K. (2001). Non-photochemical quenching. A response to excess light energy. *Plant Physiol.* 125, 1558–1566. doi: 10.1104/pp.125.4.1558
- Muller-Schussele, S., Wang, R., Gutle, D., Romer, J., Rodriguez-Franco, M., Scholz, M., et al. (2020). Chloroplasts require glutathione reductase to balance reactive oxygen species and maintain efficient photosynthesis. *Plant J.* 103, 1140–1154. doi: 10.1111/tpj.14791
- Munekage, Y., Hashimoto, M., Miyake, C., Tomizawa, K.-I., Endo, T., Tasaka, M., et al. (2004). Cyclic electron flow around photosystem I is essential for photosynthesis. *Nature* 429, 579–582. doi: 10.1038/nature02598
- Ort, D., and Baker, N. (2002). A photoprotective role for O(2) as an alternative electron sink in photosynthesis? *Curr. Opin. Plant Biol.* 5, 193–198. doi: 10.1016/S1369-5266(02)00259-5
- Poolman, M. G., Fell, D. A., and Thomas, S. (2000). Modelling photosynthesis and its control. *J. Exp. Bot.* 51, 319–328. doi: 10.1093/jexbot/51.suppl\_1.319
- Raven, J. A. (2011). The cost of photoinhibition. *Physiol. Plant.* 142, 87–104. doi: 10.1111/j.1399-3054.2011.01465.x
- Ruban, A. (2016). Nonphotochemical chlorophyll fluorescence quenching: mechanism and effectiveness in protecting plants from photodamage. *Plant Physiol.* 170, 1903–1916. doi: 10.1104/pp.15.01935
- Schwarzlander, M., and Finkemeier, I. (2013). Mitochondrial energy and redox signaling in plants. *Antioxid. Redox Signal.* 18, 2122–2144. doi: 10.1089/ars.2012.5104
- Stirbet, A., Lazar, D., Guo, Y., and Govindjee, G. (2020). Photosynthesis: basics, history and modelling. *Ann. Bot.* 126, 511–537. doi: 10.1093/aob/mcz171

- Suzuki, N., Koussevitzky, S., Mittler, R., and Miller, G. (2012). ROS and redox signalling in the response of plants to abiotic stress: ROS and redox signalling in plants. *Plant Cell Environ.* 35, 259–270. doi: 10.1111/j.1365-3040.2011.02336.x
- Valero, E., González-Sánchez, M. I., Maciá, H., and García-Carmona, F. (2009). Computer simulation of the dynamic behavior of the glutathione-ascorbate redox cycle in chloroplasts. *Plant Physiol.* 149, 1958–1969. doi: 10.1104/pp.108.133223
- Valero, E., Maciá, H., De la Fuente, I. M., Hernandez, J.-A., Gonzalez-Sanchez, M.-I., and Garcia-Carmona, F. (2015). Modeling the ascorbate-glutathione cycle in chloroplasts under light/dark conditions. *BMC Syst. Biol.* 10:11. doi: 10.1186/s12918-015-0239-y
- van Aalst, M., Ebenhöf, O., and Matuszyńska, A. (2021). Constructing and analysing dynamic models with modelbase v1. 2.3: a software update. *BMC Bioinformatics* 22:203. doi: 10.1186/s12859-021-04122-7
- Wang, C., Yamamoto, H., and Shikanai, T. (2015). Role of cyclic electron transport around photosystem I in regulating proton motive force. *Biochim. Biophys. Acta Bioenerg.* 1847, 931–938. doi: 10.1016/j.bbabi.2014.11.013
- Yamamoto, H., and Shikanai, T. (2018). PGR5-dependent cyclic electron flow protects photosystem I under fluctuating light at donor and acceptor sides. *Plant Physiol.* 179, 588–600. doi: 10.1104/pp.18.01343
- Yamori, W., Makino, A., and Shikanai, T. (2016). A physiological role of cyclic electron transport around photosystem I in sustaining photosynthesis under fluctuating light in rice. *Sci. Rep.* 6:20147. doi: 10.1038/srep20147
- Yokota, A., Kitaoka, S., Miura, K., and Wadano, A. (1985). Reactivity of glyoxylate with hydrogen peroxide and simulation of the glycolate pathway of C3 plants and Euglena. *Planta* 165, 59–67. doi: 10.1007/BF00392212
- Zhu, X.-G., Ort, D. R., Parry, M. A. J., and von Caemmerer, S. (2020). A wish list for synthetic biology in photosynthesis research. *J. Exp. Bot.* 71, 2219–2225. doi: 10.1093/jxb/eraa075

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Saadat, Nies, van Aalst, Hank, Demirtas, Ebenhöf and Matuszyńska. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## **6.5 Bioenergetics of the secondary metabolites production in photosynthetic glandular trichomes**

### **6.5.1 Authors and Details**

Authors: Nima P. Saadat, Marvin v. Aalst, Oliver Ebenhöh,  
Alain Tissier, Anna Matuszyńska  
Authorship: 1. Author (shared)  
Journal: Submitted to New Phytologist  
Status: Submitted

### **6.5.2 Contributions**

Nima P. Saadat produced the results, helped building the model, produced original draft methods, results and discussion, helped visualizing results and improving the manuscript. Marvin van Aalst helped producing results, building the model and improving the original draft manuscript. Anna Matuszyńska developed the original idea and concept of the project, helped acquiring funding, produced the original draft introduction and helped improving the manuscript. Oliver Ebenhöh helped with building the model and interpreting the results, as well as improving the manuscript. Alain Tissier helped with interpreting the results and improving the manuscript.

---

## ORIGINAL ARTICLE

Journal Section

# Bioenergetics of the secondary metabolites production in photosynthetic glandular trichomes

Nima P. Saadat<sup>1\*</sup> | Marvin van Aalst<sup>1\*</sup> | Oliver Ebenhöf<sup>1,2</sup> | Alain Tissier<sup>3</sup> | Anna B. Matuszyńska<sup>2,4</sup>

<sup>1</sup>Institute of Theoretical and Quantitative Biology, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

<sup>2</sup>Cluster of Excellence on Plant Sciences, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

<sup>3</sup>Leibniz-Institut für Pflanzenbiochemie, Weinberg 3, 06120 Halle, Germany

<sup>4</sup>Computational Life Science, Department of Biology, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany

### Correspondence

Nima Saadat, Institute of Theoretical and Quantitative Biology, Heinrich Heine University Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany  
Email: nima.saadat@hhu.de

### Funding information

This work was funded by the Deutsche Forschungsgemeinschaft Research (DFG, German Research Foundation) Grant MA 8103/1-1 project ID 420069095 (AM) and EU's Horizon 2020 research and innovation programme under the Grant Agreement 862087 (MvA).

Several commercially important secondary metabolites are produced and accumulated in high amounts by glandular trichomes (GTs), giving the prospect of exploiting trichomes as metabolic cell factories. Due to extremely high metabolic fluxes through GT particular attention was drawn to how they are achieved. Question regarding bioenergetics of these dedicated structures becomes more interesting when one realizes that some cells are photosynthetically active. Despite recent advances, a profound understanding of how primary metabolism contributes to GTs high metabolic fluxes is not fully elucidated. We provide the first reconstruction of specialised metabolism in photosynthetic glandular trichomes of *Solanum lycopersicum*. The outcome of this research closes an important gap in knowledge of the role of chloroplast in the synthesis of specialised metabolites in plants' GTs.

### KEYWORDS

bioenergetics, glandular trichomes, photosynthesis, stoichiometric model, secondary metabolites

**Abbreviations:** CBB, Calvin-Benson-Bassham cycle; DMAPP, dimethylallyl diphosphate; FBA, flux-balance analysis; F6P, Fructose-6-Phosphate; GEM, genome-scale metabolic models; GT, glandular trichome; IPP, isopentenyl diphosphate; LP, linear programming; MEP, methyl-erythritolphosphate pathway; MEV, mevalonate pathway; PETC, photosynthetic electron transfer chain; TCA, citric acid cycle

Equally contributing authors.

## 1 | INTRODUCTION

Most plant species exhibit cellular outgrowths of their epidermis called trichomes. Due to their often species-specific characteristic, many criteria for classification exist, the most popular one being the division into non-glandular and **glandular trichomes** (GT) [1]. Whilst non-glandular trichomes serve more as a physical and mechanical defence against biotic and abiotic stresses, all GTs are characterised by the ability to synthesise and accumulate vast amounts of valuable specialised (secondary) metabolites. Due to extremely high metabolic fluxes in these organs, production of some metabolites can reach up to 20% of the leaf dry weight, qualifying GTs as true **metabolic cell factories**. Products of GTs include terpenoids, phenylpropanoids, flavonoids, fatty acid derivatives and acyl sugars [2] exhibiting antifungal, insecticide or pesticide properties. Thereby GTs are not only incredibly important to plant fitness, as they contribute to the chemical arsenal of plants, but are also of relevance to multiple industries.

The key carbon source in most GTs of tomatoes is sucrose which is converted into a multitude of organism-specific metabolites in the glands [3]. The massive productivity of hydrocarbon compounds implies however a supply of adequate amounts of not only carbon, energy and reducing power, but also precursors, produced by intermediate pathways. Terpenoids represent the largest and structurally most diverse class of plant metabolites and are major products of GT biosynthesis. Despite their multiplicity, with over 30 000 well-known structures, they are all assemblies of C5 isoprene units built from isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). There are two identified pathways for IPP and DMAPP production: i) the plastidial 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway from pyruvate and glyceraldehyde-3-phosphate or ii) the cytosolic mevalonate (MVA) pathway from acetyl-CoA [4]. Although the pathways are thought to be largely independent, some exchange of precursors may occur [5], and such (and others [6]) **cross-talk** require further investigation. For instance, is there some cross-talk of plastidial and cytosolic pathways providing the 5-carbon precursors? And if so, what effect does it have on overall productivity? Beyond this, a major issue is the source of energy and its distribution to understand how GTs achieve their high productivity. The question becomes more intriguing when one realises that some of the GTs contain photosynthetically active chloroplasts (as the type VI GT in *S. lycopersicum* [7]). It is still unclear what the advantages and disadvantages of photosynthetic GTs, are in contrast to non-photosynthetic GTs. The separation of cytosolic and chloroplast bound pathways, as well as the utility of photosynthesis, are until now only vaguely understood, and the most recent summary of current advances has been recently provided [8].

To shed light on the advantages of photosynthetic GT for terpenoid synthesis and secondary metabolism, investigations of the systems bioenergetics and reaction flux distributions are needed. Mathematical, computational models provide a coherent framework to study metabolism. Constraint-based stoichiometric models [9] are particularly adequate for exploratory studies of the systemic properties of a metabolic network and investigations of the flux distributions. Such models are static and represent mathematically the network of biochemical reactions of an organism in the form of a matrix [10]. They can focus on various scales, with genome-scale metabolic models (GEMs) aiming at representing the whole biochemical network of an individual organism. GEMs are constructed by assigning biochemical functions to enzymes encoded in the genome, and due to the expansion of the whole genome sequencing, many plant GEMs are currently available, with *Oryza sativa indica* [11], *Arabidopsis thaliana* [12] and *Solanum lycopersicum* L. [13] among many others. Flux Balance Analysis (FBA) [14, 15], a mathematical method that allows calculating the flow of metabolites through the network, is a popular tool to predict the production rate of the compound of interest. FBA requires two assumptions: i) the experimental system is at a steady state, and ii) the network is optimised to maximise or minimise certain biological outcomes, for instance, its biomass. The so-called, cell-specific, objective functions in GEMs are optimized in a linear programming (LP) approach in which all reaction fluxes are constrained within given boundaries. This constraint-based analysis of GEMs allows the calculation of optimal flux solutions in

different conditions, therefore allowing investigations on the metabolic fluxes and bioenergetics of systems.

In this work, we have reconstructed the metabolism in the photosynthetic glandular trichome of a *Solanum lycopersicum* LA4024 using previously published transcriptome and metabolome data [3]. With a general, mathematical framework, we investigated the effect of having photosynthetically active machinery inside of a trichome. In our simulations, we observed the increase in terpenoid production under increasing light intensities. Increased photosynthetic activity shifts the partitioning of uptaken carbons from catabolism to anabolism due to increased energy levels. Bioenergetics and energy levels determine which of the known terpenoid precursor production pathways (MEV, MEP) is more desirable/optimal in different light/stress conditions. Our model can explain the benefits of having chloroplasts in GTs and serves as a groundwork for further investigations of the possible cross-talks between the two pathways of terpenoid precursor synthesis.

## 2 | METHODS

### 2.1 | Choice of the model organism

In this study, we have chosen to investigate GT in the tomato genus. *Solanum lycopersicum* serves as an excellent model organism for glandular trichome study due to the availability of i) high-quality complete genome sequence [16], ii) excellent genetic resources [17], iii) comparative multi-omics data [3], iv) several mathematical models available, including whole genome metabolic network reconstruction [13], and v) in contrast to other well studied organisms like peppermint [18, 19], possession of only photosynthetic GT.

### 2.2 | Modelling environment

Our model is implemented in Python, using our in-house developed package `moped`. With `moped` all decision processes and taken steps are well documented in a transparent and repeatable fashion [20]. The annotated script detailing every step we used to construct the model can be found in our GitLab repository at <https://gitlab.com/qtbbhu/models/glandular-trichomes>.

### 2.3 | Model design and assumptions

Although a genome-scale model of tomato metabolism is available (iHY3410 model [13]), we decided to use the bottom-up approach and perform a reconstruction ourselves, as we were not able to reconstruct the steps of manual curation performed by the authors. We based the model reconstruction on available transcriptomics and metabolomics data [3], the LycoCyc database (tomato metabolic pathway database, version 3.3 [21], available from Solanaceae Genomics Network, <http://www.sgn.cornell.edu>) and biochemical knowledge in plants from scientific publications. The resulting model consists of 1307 reactions and 1371 metabolites and its behaviour have been ensured to match reported observations [3]. Furthermore, the model quality and consistency have been thoroughly inspected using MEMOTE to ensure the highest quality [22]. There are nine exchange reactions, allowing free exchange of metabolites such as oxygen, as well as light absorption. Light is represented as photosynthetically active photons absorbed by the photosystems. The units of light absorption are represented in  $\frac{\mu\text{molPhotons}}{\text{s}\cdot\text{m}^2}$  and the detailed calculation of our calculations are provided in the Supporting Information. In this model, we decided not to include a free carbon dioxide influx as it has been reported that carbon dioxide exchange is 100 times lower in photosynthetic GTs than in leaves [3].

In most constraint-based models and their analyses, the maximization objective is the production of biomass [23]. While this may be applicable for prokaryotic organisms, we doubt that photosynthetic glandular trichome cells are maximising the increase of their replication rate, and rather maximise terpenoid synthesis while also having a mandatory production rate of macromolecules to keep cells intact. For this, our model includes an objective function to produce terpenoids while requiring a fixed flux through a function of biomass synthesis, consuming typical components like amino acids, sugars, nucleotides and fatty acid precursors. For this, we used an *Escherichia coli* (*E. coli*) biomass function, as our main concern was to capture the necessity for growth and self-repair, while the primary objective function of a glandular trichome is terpenoid production. To describe additional energy required for the maintenance of cells, we implemented a representative reaction for ATP maintenance, as it is common practice in metabolic modelling [24]. After subsequent gap-filling using Meneco, a tool for metabolic network completion [25], our model can simulate the synthesis of all compounds found within the metabolomics data [3], all terpenoids found in photosynthetic GTs of tomato [26] as well as all compounds within biomass from sucrose, light, orthophosphate, ammonia, sulfate, protons and water. The resulting model is a data-driven, yet simplified, constraint-based model which is tested against infeasible energy and mass generating cycles. Within our model simplifications, we found that a model consisting of three essential compartments (cytosol, intermembrane space and extracellular space) sufficiently represents photosynthetic GT metabolic profiles (Fig. 1). While detailed compartmental separation is common practice in large genome-scale metabolic models, it would not make any difference to the results of our model simulations due to the fact that there are several intercompartmental transporters between the chloroplast and the cytosol for energy equivalents like ATP and other key metabolites [27]. Adding over-detailed compartmentalisation to the model would therefore not alter any of our results and is left out for the sake of model simplicity and preventing unfavourable model modifications. All details and information about the exact construction process of the model, as well as all investigations and analyses, can be found in our provided scripts at <https://gitlab.com/qtb-hhu/models/glandular-trichomes>.

### 3 | RESULTS

We used our model to perform a general analysis in which we simulate the rate of terpenoid synthesis over systematically increasing light intensities via parsimonious Flux Balance Analysis (pFBA). Fig. 2 displays that with increasing light absorptions, the rate of terpenoid synthesis in photosynthetic glandular trichomes increases up until approximately  $10 \frac{\mu\text{mol Photons}}{\text{s}\cdot\text{m}^2}$ . This increase in terpenoid synthesis rate with increasing absorbed light is particularly interesting due to the fact that the model can not utilize atmospheric carbon dioxide, and Sucrose being the only carbon source. This means that there is a change in metabolic fluxes which enables this increase in terpenoid synthesis rate. To further investigate what changes in the metabolism of photosynthetic glandular trichomes in increasing light intensities, we inspect the respective changes in the exchange fluxes of the model. Fig 2 shows the exchange fluxes of carbon dioxide and oxygen in our pFBA model simulations over increasing light absorptions. Noticeably, the excretion of carbon dioxide systematically decreases up until approximately  $10 \frac{\mu\text{mol Photons}}{\text{s}\cdot\text{m}^2}$ . Interestingly, the consumption of oxygen decreases to zero at approximately  $7 \frac{\mu\text{mol Photons}}{\text{s}\cdot\text{m}^2}$ . From this light intensity on, oxygen excretion begins and increases until  $10 \frac{\mu\text{mol Photons}}{\text{s}\cdot\text{m}^2}$ . These observations are crucial for a general understanding of the model behaviour. An increase in absorbed light causes higher photosynthetic activity, resulting in oxygen production. This explains the decreasing oxygen uptake and the switch to oxygen excretion at  $7 \frac{\mu\text{mol Photons}}{\text{s}\cdot\text{m}^2}$  absorbed light. However, the steady decrease in carbon dioxide excretion is especially noteworthy. Most carbon dioxide is produced within catabolism, therefore the model behaviour hints at a decrease in catabolic activity in higher light intensities.

To investigate how the catabolic activity in our model simulations changes over increasing light intensities, we

further inspect representative reactions for relevant catabolic pathways in our model. As Sucrose, a disaccharide is the only carbon source in our model, we inspect representatives of the upper glycolysis, the lower glycolysis and the TCA cycle. Fig. 3 displays the fluxes of these reactions over different light intensities (shown on the x-axis as fractions of saturating light intensities) relative to their fluxes in the dark. The Sucrose Synthase and Saccharase represent upper glycolysis activity. The 6-Phosphofruktokinase, GAP Dehydrogenase and Pyruvate Kinase represent lower glycolysis activity and the Pyruvate Dehydrogenase and the Citrate Synthase represent TCA cycle activity. Furthermore, the RuBisCO rate is displayed to monitor the rate of carbon refixation. The results show that fluxes of upper glycolysis remain completely unchanged in increasing light intensities, however, the fluxes in lower glycolysis decrease in higher light conditions. An even higher impact can be observed for the TCA cycle activity. The Pyruvate Dehydrogenase activity steadily decreases, and the Citrate Synthase activity abruptly decreases in increasing light conditions. These observations show that catabolic pathways which are not responsible for energy and redox equivalent production (like upper glycolysis) are unaffected by increasing light intensities. However, the lower glycolysis and the TCA cycle, both catabolic pathways that produce energy and redox equivalents, display a strong flux decrease in higher light conditions. The increase in terpenoid synthesis flux observed in Fig 2, and the decrease in catabolic fluxes in Fig 3 strongly suggest that increasing light conditions shift the carbon partitioning from catabolic to anabolic pathways. This shift is enabled due to the energy and redox equivalent production of the photosynthetic electron transport chain in photosynthetic glandular trichomes. The metabolic network is not dependent on the energy from oxidising carbon bodies in high light conditions, and can therefore use more of those carbon bodies in terpenoid synthesis pathways. Interestingly, RuBisCO activity increases in higher light intensities, displaying that higher energy levels allow the refixation of carbon that is lost as carbon dioxide in anabolic processes (like Terpenoid synthesis). GAP, Pyruvate and Acetyl-CoA are carbon bodies which can be used to produce either energy and redox equivalents or terpenoid precursors. Acetyl-CoA is the initial substrate of the TCA cycle in which it is oxidised to gain energy and redox equivalents but is also the initial substrate of the MEV pathway. GAP and pyruvate are metabolites within the lower glycolysis pathway and also initial substrates of the MEP which is a terpenoid synthesis pathway exclusive to photosynthetic GTs.

To further analyse how the consumption of these metabolites depends on the illumination, we simulated the relative consumption rate of GAP/pyruvate and Acetyl-CoA by the aforementioned pathways over increasing light intensities. Fig 4 displays the proportions of the consumption of these compounds by the TCA, MEV and MEP pathways. In low light intensities, more than half of the substrates are consumed by the MEV pathway, and the remainder is consumed by the TCA cycle, in both cases in the form of Acetyl-CoA. In higher light intensities, the fraction of substrates consumed by the TCA cycle is decreasing until it does not consume any more substrates. At this point, the relative flux of lower glycolysis starts decreasing, and the MEP pathway is beginning to consume proportions of the substrates, gradually taking over. This is a very important observation that shows that increasing light intensities, leading to higher energy levels due to photosynthetic activity, shift the carbon partitioning from catabolic to anabolic pathways by reducing the TCA cycle and lower glycolytic flux and increasing terpenoid synthesis. Furthermore, it shows that the two terpenoid synthesis pathways, MEP and MEV, are more advantageous at different energetic levels. In lower light intensities, and therefore lower energetic levels, the MEV pathway seems to be more advantageous because the conversion of GAP and pyruvate to Acetyl-CoA produces energy and redox equivalents, and the resulting Acetyl-CoA can directly be used in the TCA cycle to generate additional energy and redox equivalents. In higher light intensities, and therefore higher energetic levels, the MEP pathway is more advantageous because the high energy levels provided by photosynthetic activity remove the necessity of providing energy and redox equivalents via lower glycolysis and the TCA cycle. Instead, GAP and pyruvate can directly be used as substrates with higher energy contents (than Acetyl-CoA) in the MEP pathway, and therefore further increase the fraction of carbon used in anabolism,

enabling more efficient terpenoid synthesis. This phenomenon can also be observed in Fig 5, in which we used model simulations to calculate the fluxes of the final MEV and MEP reactions in systematically changing light conditions and ATP maintenance costs.

In this analysis, higher ATP maintenance costs reflect increased energy requirements of cells in e.g. stress conditions. At low light conditions and low ATP maintenance costs, the MEV pathway is the main terpenoid synthesis pathway, with very little MEP pathway activity. In low light conditions and high ATP maintenance costs, the MEV pathway is the only active pathway. However, the overall terpenoid synthesis flux is relatively low due to the increased demand for catabolic flux in such conditions. At high light conditions and high ATP maintenance costs, the MEP pathway is carrying the majority of terpenoid synthesis flux. In high light conditions and low ATP maintenance costs, the MEP pathway is the only active terpenoid synthesis pathway, providing the highest terpenoid synthesis flux. It appears as the distribution of terpenoid synthesis between the MEV and the MEP pathways are highly dependent on the light conditions and resulting energy levels of the photosynthetic GTs.

The high rate of terpenoid synthesis in high light conditions is also resulting from increased rates of carbon refixation. It remains unknown how active the CBB cycle is in photosynthetic GTs. To quantify the impact of different carbon refixation fluxes, we performed a systematic analysis in which we calculated the terpenoid synthesis rate over different amounts of absorbed light and systematically changed the activities of RuBisCO. Fig. 6 displays that in higher light conditions, increased carbon refixation fluxes can increase the rate of terpenoid synthesis by almost 20%. However, interestingly, the overall rate of carbon refixation is not very high. This is because the rate of carbon refixation is purely dependent on the available carbon dioxide produced by anabolic processes, which is limited.

## 4 | DISCUSSION

In photosynthetic glandular trichomes, synthetic pathways of terpenoids and other secondary metabolites are not only found in the cytosol of the cells, but also the chloroplasts. The additional terpenoid synthesis pathways in photosynthetic glandular trichomes have been subject to many speculations. One speculation for example is that terpenoid production in chloroplasts is specialized for the production of particular secondary metabolites. Another speculation is that the pH in chloroplasts may be more beneficial for terpenoid synthesis. In our work, we introduce a simplified, yet a data-driven constraint-based model of photosynthetic glandular trichome metabolism which shows that one of the two different synthesis pathways is more advantageous for terpenoid production than the other in different energy availabilities. We show that with lower energy availability, the cytosolic MEV pathway is more advantageous for terpenoid synthesis because the catabolic pathways, producing the key initial substrate Acetyl-CoA from Sucrose, provide additional energy and redox equivalents needed for all cellular activities, including terpenoid synthesis. However, higher energy availability (coming from photosynthetic activity in higher light conditions) removes the need for the additional energy and redox equivalents gained from the conversion of Sucrose to Acetyl-CoA. Therefore substrates with higher energy levels (GAP and pyruvate) can be directly used for terpenoid synthesis. This basic shortcut of catabolic reactions reduces the loss of carbon as carbon dioxide and increases the flux of carbon through anabolic processes with the help of energy equivalents gathered through photosynthesis.

We show that in higher light conditions, energy levels of the photosynthetic GTs are so advantageous, that the energy can be spent to perform carbon refixation using the CBB cycle. In the supplementary material, we calculated that in order to maintain the same rate of Terpenoid synthesis in low light, only half of the sucrose is required in high light conditions. This displays again that the benefit of including chloroplasts in glandular trichomes is not only the ability to shift carbon partitioning from catabolic to anabolic processes but also to further maximize carbon use

efficiency. It is important to note that the increase in terpenoid synthesis from carbon refixation is not nearly as high as the increase from the shift in carbon partitioning, as seen in Fig. 6.

Interestingly our model showed that even without CBB cycle activity and without special constraints, the TCA cycle may be reversed in high energy availability and function as a reductive TCA cycle. A reductive TCA cycle could be able to replace the function of the CBB cycle, using energy to fix carbon dioxide which was produced in catabolic and anabolic reactions, increasing carbon use efficiency. This is a very interesting observation, as, from a bioenergetic point of view, such a scenario is possible. However, we decided to adjust the key reactions of the TCA cycle for this scenario as irreversible reactions to prevent this phenomenon to be included in our results for now. The reason for this decision is that the reductive TCA cycle is usually found in green sulfur bacteria and different thermophilic prokaryotes and archaea [28, 29]. This indicates that from a phylogenetic perspective, the presence of a reductive TCA cycle in photosynthetic GTs is rather unlikely. However, we think that this model suggestion is worth investigating the fluxes of the TCA cycle in light conditions in photosynthetic GTs, as it has been suggested that carbon dioxide may be recovered [30]. Generally, instead of showing that chloroplastic terpenoid synthesis pathways provide improved production of particular terpenoids, our work shows that the chloroplast in photosynthetic GTs functions as an energy battery in light conditions, which can be used to shift carbon from catabolic to anabolic fluxes and even enable carbon dioxide refixation and therefore improve carbon use efficiency. To support our findings, experiments are needed which can keep track of the rate of terpenoid synthesis in similar sucrose availability but different light absorptions.

The photosynthetic carbon refixation indicates that there may be photorespiration present in photosynthetic GTs. Although photorespiration is not included in our model, we show that in high light there is oxygen evolution in photosynthetic GTs. Therefore, further data is required to investigate putative photorespiratory activities. Furthermore, it remains unclear if and how high evolution of reactive oxygen species and photodamage is present in photosynthetic GTs. For this, quantitative metabolic data for the components of the electron transport chain is needed, as well as measurements of the photosynthetic efficiency in photosynthetic GTs. Finally, more questions regarding the dynamics, and not only bioenergetics of trichomes arise. E.g., what is the composition of terpenoids under different light intensities, or how can we improve the production of a terpenoid of interest? As these processes are heavily dependent on enzyme kinetics and saturation, constraint-based models like ours are not the optimal method for answering these questions. However, mechanistic models based on ordinary differential equations can include such information (if available) and may be helpful to give further insights into terpenoid synthesis in photosynthetic GTs. New experimental data is therefore required to advance further our understanding of the biosynthesis of these **metabolic cell factories**.

## Acknowledgements

## Conflict of interest

The authors declare that they have no conflict of interest.

## Supporting Information

The computational model presented here, together with the code to reproduce the computational analysis are openly available on the GitLab repository <https://gitlab.com/qtb-hhu/models/glandular-trichomes> or can be requested from the authors.

## Calculation of flux units and light intensity units

A suggested terpenoid production rate of glandular trichomes has been provided by Turner *et al.* [31] at  $0.017 \frac{\text{nmol}}{\text{h-gland}}$ . Assuming that this rate can be applied to the maximal terpenoid production rate of photosynthetic glandular trichomes of tomatoes, we transform our calculated fluxes to the corresponding units by  $\text{Flux} \cdot \frac{0.017 \frac{\text{nmol}}{\text{h-gland}}}{\text{max. Terp flux}}$ .

In order to convert the Fluxes of Photons into units of light intensities, we first have to calculate the surface of the glandular trichome on which light can be absorbed. Measured values of the diameters of glandular trichomes from [32] providing an estimate of 50  $\mu\text{m}$  diameter of circular surface of a glandular trichome. Assuming that the glandular trichomes surface is a circle, the surface area can be calculated as:

$$A = \pi \left( \frac{50\mu\text{m}}{2} \right)^2 = 2000\mu\text{m}^2 = 2 \cdot 10^{-9} \text{m}^2$$

To calculate the conversion factor for the photon absorption of glandular trichomes, we first calculate the units of photons absorbed by the gland at saturated light flux and maximal terpenoid production predicted by our model as:

$$0.017 * \frac{\text{Max. Light Flux}}{\text{Max. Terp. Flux}} = 0.017 * \frac{300}{7.3} = 0.7 \frac{\text{nmol/Photons}}{\text{h-gland}}$$

To convert this unit into  $\frac{\mu\text{mol}}{\text{s}\cdot\text{m}^2}$ , we first calculate the corresponding unit for  $\frac{\text{nmol/Photons}}{\text{h-gland}}$  by:

$$\frac{1\text{nmol/Photons}}{\text{h-gland}} = \frac{1}{7200} \frac{\text{mol/Photons}}{\text{s}\cdot\text{m}^2} = 14 \frac{\mu\text{mol/Photons}}{\text{s}\cdot\text{m}^2}$$

for our maximal Light flux, this corresponds to  $0.7 \cdot 14 \frac{\mu\text{mol/Photons}}{\text{s}\cdot\text{m}^2} = 9.8 \frac{\mu\text{mol/Photons}}{\text{s}\cdot\text{m}^2}$  as the saturating light intensity, providing the light flux conversion factor of  $\text{Light Flux} \cdot \frac{9.8 \frac{\mu\text{mol/Photons}}{\text{s}\cdot\text{m}^2}}{\text{sat. Light flux}}$ .

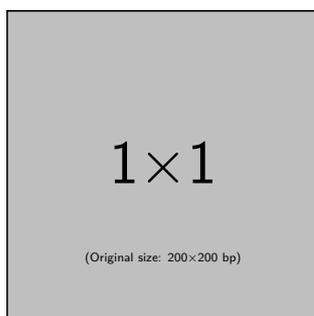
## references

- [1] E Werker. Trichome diversity and development. *Adv. Bot. Res.*, 31:1–35, jan 2000. ISSN 0065-2296. doi: 10.1016/S0065-2296(00)31005-9.
- [2] Joris J. Glas, Bernardus C.J. Schimmel, Juan M. Alba, Rocío Escobar-Bravo, Robert C. Schuurink, and Merijn R. Kant. Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores. *International Journal of Molecular Sciences*, 13(12):17077–17103, 2012. ISSN 14220067. doi: 10.3390/ijms131217077.
- [3] Gerd U. Balcke, Stefan Bennewitz, Nick Bergau, Benedikt Athmer, Anja Henning, Petra Majovsky, José M. Jiménez-Gómez, Wolfgang Hoehenwarter, and Alain Tissier. Multi-Omics of Tomato Glandular Trichomes Reveals Distinct Features of Central Carbon Metabolism Supporting High Productivity of Specialized Metabolites. *The Plant Cell*, 29(5): 960–983, 04 2017. ISSN 1040-4651. doi: 10.1105/tpc.17.00060.
- [4] R. W.J. Kortbeek, J. Xu, A. Ramirez, E. Spyropoulou, P. Diergaarde, I. Otten-Bruggeman, M. de Both, R. Nagel, A. Schmidt, R. C. Schuurink, and P. M. Bleeker. Engineering of Tomato Glandular Trichomes for the Production of Specialized Metabolites. *Methods in Enzymology*, 576:305–331, 2016. ISSN 15577988. doi: 10.1016/bs.mie.2016.02.014.
- [5] Heike Paetzolda, Stefan Garms, Stefan Bartramb, Jenny Wiczorek, Eva-Maria Uros-Gracia, Manuel Rodriguez-Concepcion, Wilhelm Boland, Dieter Strack, Bettina Hause, and Michael H. Walter. The Isogene 1-Deoxy-D-Xylulose 5-Phosphate Synthase 2 Controls Isoprenoid Profiles, Precursor Pathway Allocation, and Density of Tomato Trichomes. *Mol. Plant*, 3(5):904–916, 2010. doi: 10.1093/mp/ssp032.
- [6] Zhengzhi Xie, Jeremy Kapteyn, and David R. Gang. A systems biology investigation of the MEP/terpenoid and shikimate/phenylpropanoid pathways points to multiple levels of metabolic control in sweet basil glandular trichomes. *Plant Journal*, 54(3):349–361, 2008. ISSN 09607412. doi: 10.1111/j.1365-313X.2008.03429.x.
- [7] Nick Bergau, Stefan Bennewitz, Frank Syrowatka, Gerd Hause, and Alain Tissier. The development of type VI glandular trichomes in the cultivated tomato *Solanum lycopersicum* and a related wild species *S. habrochaites*. *BMC Plant Biology*, 15(1):1–15, 2015. ISSN 14712229. doi: 10.1186/s12870-015-0678-z.

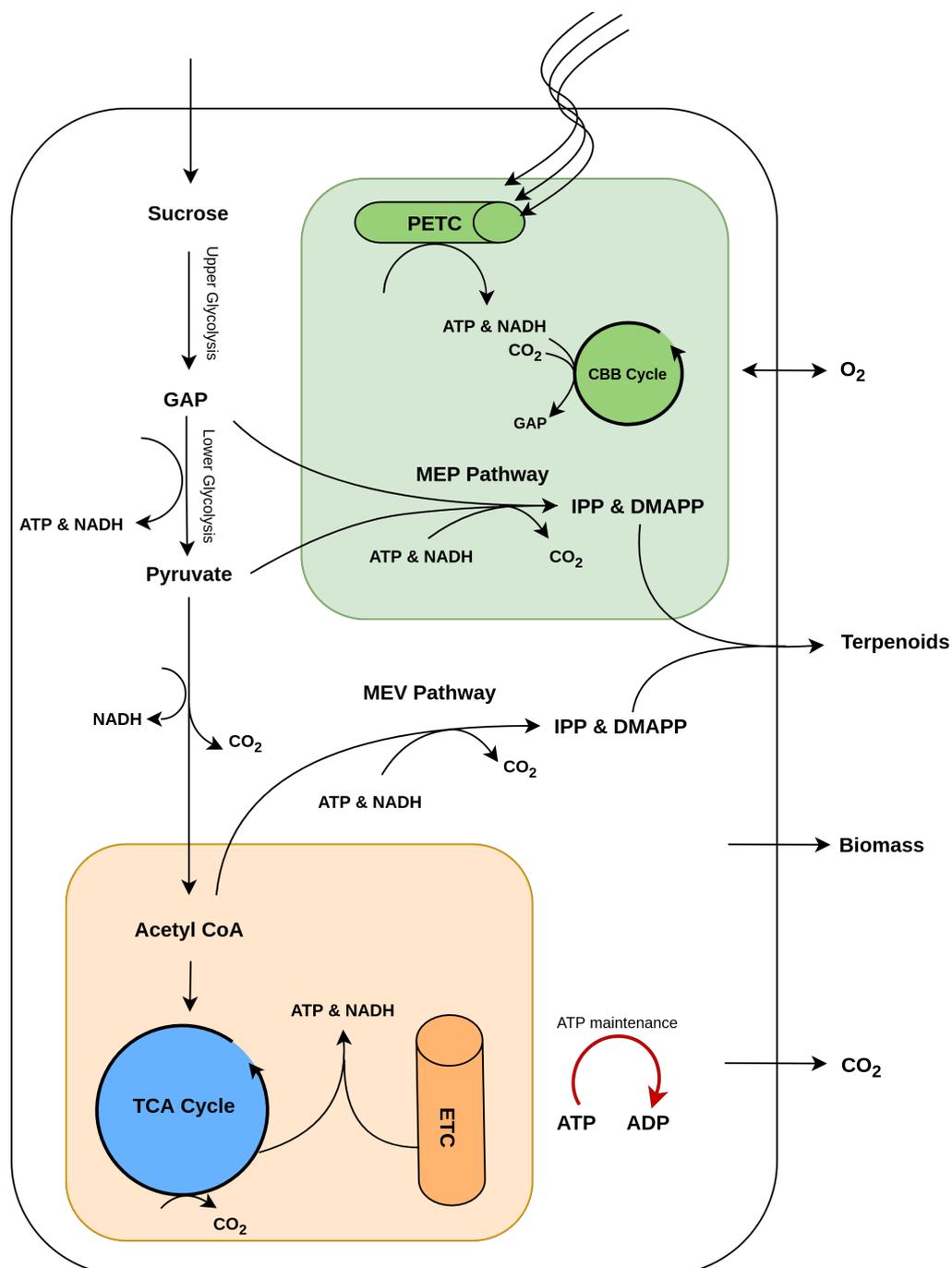
- [8] Alejandro Brand and Alain Tissier. Control of resource allocation between primary and specialized metabolism in glandular trichomes. *Current Opinion in Plant Biology*, 66:102172, 2022. ISSN 1369-5266. doi: <https://doi.org/10.1016/j.pbi.2022.102172>.
- [9] Timo R. Maarleveld, Ruchir A. Khandelwal, Brett G. Olivier, Bas Teusink, and Frank J. Bruggeman. Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnol. J.*, 8(9):997–1008, 2013. ISSN 18606768. doi: 10.1002/biot.201200291.
- [10] Reinhart Heinrich and Stefan Schuster. The regulation of cellular systems. *Book*, page 416, 1996. doi: 10.1007/978-1-4613-1161-4.
- [11] Ankita Chatterjee, Benazir Huma, Rahul Shaw, and Sudip Kundu. Reconstruction of *Oryza sativa indica* Genome Scale Metabolic Model and Its Responses to Varying RuBisCO Activity, Light Intensity, and Enzymatic Cost Conditions. *Front. Plant Sci.*, 8(November):1–18, 2017. ISSN 1664-462X. doi: 10.3389/fpls.2017.02060. URL <http://journal.frontiersin.org/article/10.3389/fpls.2017.02060/full>.
- [12] Mark G Poolman, Laurent Miguet, Lee J Sweetlove, and David A Fell. A Genome-scale Metabolic Model of *Arabidopsis thaliana* and Some of its Properties. *Plant Physiol.*, 151(November):1570–1581, 2009. ISSN 0032-0889. doi: 10.1104/pp.109.141267.
- [13] Huili Yuan, C.Y. Maurice Cheung, Mark G. Poolman, Peter A. J. Hilbers, and Natal A. W. van Riel. A genome-scale metabolic network reconstruction of tomato (*Solanum lycopersicum* L.) and its application to photorespiratory metabolism. *Plant J.*, 85(2):289–304, 2016. ISSN 09607412. doi: 10.1111/tpj.13075.
- [14] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nat. Biotechnol.*, 28(3):245–248, 2011. doi: 10.1038/nbt.1614.What.
- [15] L J Sweetlove and R G Ratcliffe. Flux-balance modeling of plant metabolism. *Front Plant Sci*, 2:38, 2011.
- [16] The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485:635–641, 2012. doi: [doi.org/10.1038/nature11119](https://doi.org/10.1038/nature11119).
- [17] Vasiliki Falara, Tariq A. Akhtar, Thuong T.H. Nguyen, Eleni A. Spyropoulou, Petra M. Bleeker, Ines Schauvinhold, Yuki Matsuba, Megan E. Bonini, Anthony L. Schillmiller, Robert L. Last, Robert C. Schuurink, and Eran Pichersky. The tomato terpene synthase gene family. *Plant Physiology*, 157(2):770–789, 2011. ISSN 15322548. doi: 10.1104/pp.111.179648.
- [18] Rigoberto Rios-Esteva, Glenn W. Turner, James M. Lee, Rodney B. Croteau, and B. Markus Lange. *A systems biology approach identifies the biochemical mechanisms regulating monoterpenoid essential oil composition in peppermint*, volume 105. 2008. ISBN 0712314105. doi: 10.1073/pnas.0712314105.
- [19] Rigoberto Rios-Esteva, Iris Lange, James M. Lee, and B. Markus Lange. Mathematical modeling-guided evaluation of biochemical, developmental, environmental, and genotypic determinants of essential oil composition and yield in peppermint leaves. *Plant Physiology*, 152(4):2105–2119, 2010. ISSN 15322548. doi: 10.1104/pp.109.152256.
- [20] Nima P. Saadat, Marvin van Aalst, and Oliver Ebenhö. Network reconstruction and modelling made reproducible with moped. *Metabolites*, 12(4), 2022. ISSN 2218-1989. doi: 10.3390/metabo12040275. URL <https://www.mdpi.com/2218-1989/12/4/275>.
- [21] Edwards JD Saha S Teclé IY Strickler SR Bombarely A Fisher-York T Pujar A Foerster H Yan A Mueller LA. Fernandez-Pozo N, Menda N. The sol genomics network (sgn) from genotype to phenotype to breeding. *Nucleic Acids Res.*, 43 (Database issue):D1036–41, 2015.
- [22] Christian Lieven, Moritz E Beber, Brett G Olivier, Frank T Bergmann, Meric Ataman, Parizad Babaei, Jennifer A Bartell, Lars M Blank, Siddharth Chauhan, Kevin Correia, et al. Memote for standardized genome-scale metabolic model testing. *Nature biotechnology*, 38(3):272–276, 2020.

- [23] Willi Gottstein, Brett G. Olivier, Frank J. Bruggeman, and Bas Teusink. Constraint-based stoichiometric modelling from single organisms to microbial communities. *Journal of The Royal Society Interface*, 13(124):20160627, 2016. doi: 10.1098/rsif.2016.0627.
- [24] CY Maurice Cheung, Thomas CR Williams, Mark G Poolman, David A Fell, R George Ratcliffe, and Lee J Sweetlove. A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. *The plant journal*, 75(6):1050–1061, 2013.
- [25] Sylvain Prigent, Clémence Frioux, Simon M Dittami, Sven Thiele, Abdelhalim Larhlimi, Guillaume Collet, Fabien Gutknecht, Jeanne Got, Damien Eveillard, Jérémie Bourdon, et al. Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. *PLoS computational biology*, 13(1):e1005276, 2017.
- [26] Katrin Besser, Andrea Harper, Nicholas Welsby, Ines Schauvinhold, Stephen Slocombe, Yi Li, Richard A Dixon, and Pierre Broun. Divergent regulation of terpenoid metabolism in the trichomes of wild and cultivated tomato species. *Plant physiology*, 149(1):499–514, 2009.
- [27] Per Gardeström and Abir U Igamberdiev. The origin of cytosolic atp in photosynthetic cells. *Physiologia Plantarum*, 157(3):367–379, 2016.
- [28] Thomas M Wahlund and F Robert Tabita. The reductive tricarboxylic acid cycle of carbon dioxide assimilation: initial studies and purification of atp-citrate lyase from the green sulfur bacterium chlorobium tepidum. *Journal of bacteriology*, 179(15):4859–4867, 1997.
- [29] Monika Beh, Gerhard Strauss, Robert Huber, Karl-Otto Stetter, and Georg Fuchs. Enzymes of the reductive citric acid cycle in the autotrophic eubacterium aquifex pyrophilus and in the archaeobacterium thermoproteus neutrophilus. *Archives of Microbiology*, 160(4):306–311, 1993.
- [30] Robert Schuurink and Alain Tissier. Glandular trichomes: micro-organs with model status? *New Phytologist*, 225(6):2251–2266, 2020.
- [31] Glenn W Turner and Rodney Croteau. Organization of monoterpene biosynthesis in mentha. immunocytochemical localizations of geranyl diphosphate synthase, limonene-6-hydroxylase, isopiperitenol dehydrogenase, and pulegone reductase. *Plant physiology*, 136(4):4215–4227, 2004.
- [32] Radosław Kowalski, Grażyna Kowalska, Monika Jankowska, Agnieszka Nawrocka, Klaudia Kałwa, Urszula Pankiewicz, and Marzena Włodarczyk-Stasiak. Secretory structures and essential oil composition of selected industrial species of lamiaceae. *Acta Scientiarum Polonorum Hortorum Cultus*, 8:2, 2019.

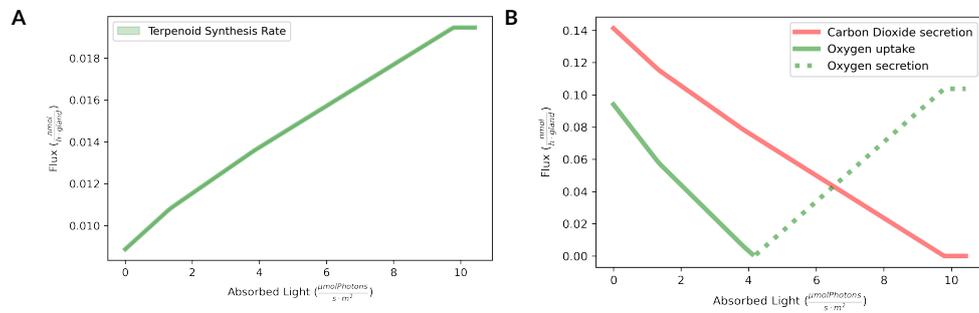
## GRAPHICAL ABSTRACT



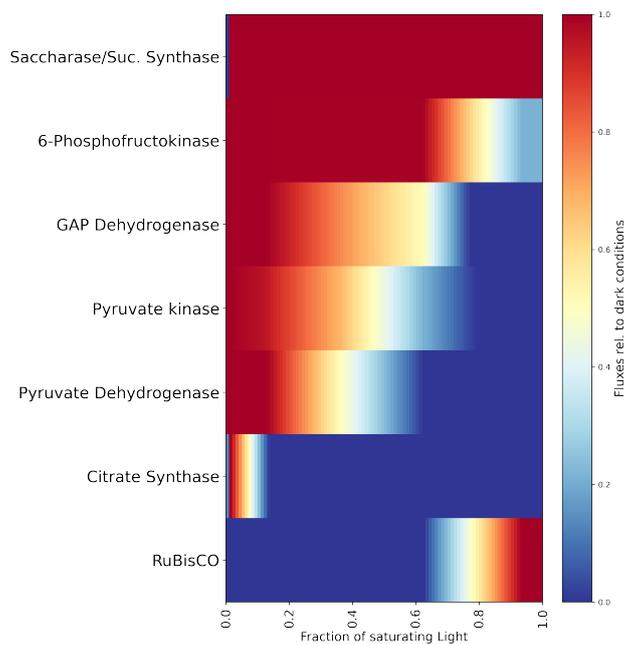
Please check the journal's author guidelines for whether a graphical abstract, key points, new findings, or other items are required for display in the Table of Contents.



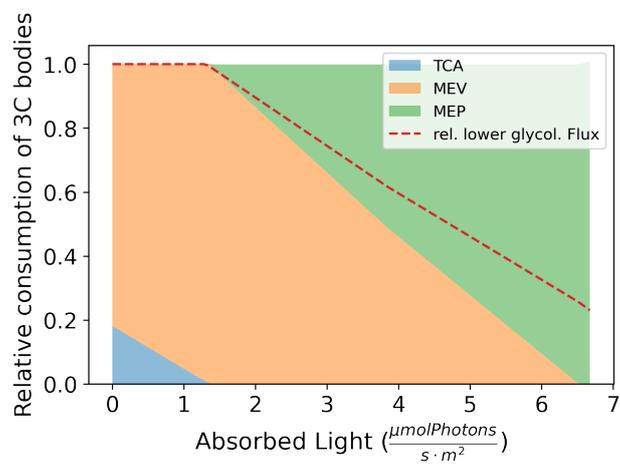
**FIGURE 1** Schematic overview of the key processes included in a constraint-based model of photosynthetic glandular trichome metabolism. While the model is built using transcriptome and metabolome data and includes a large number of reactions, only pathways and metabolites of importance to the results are highlighted in the presented model scheme.



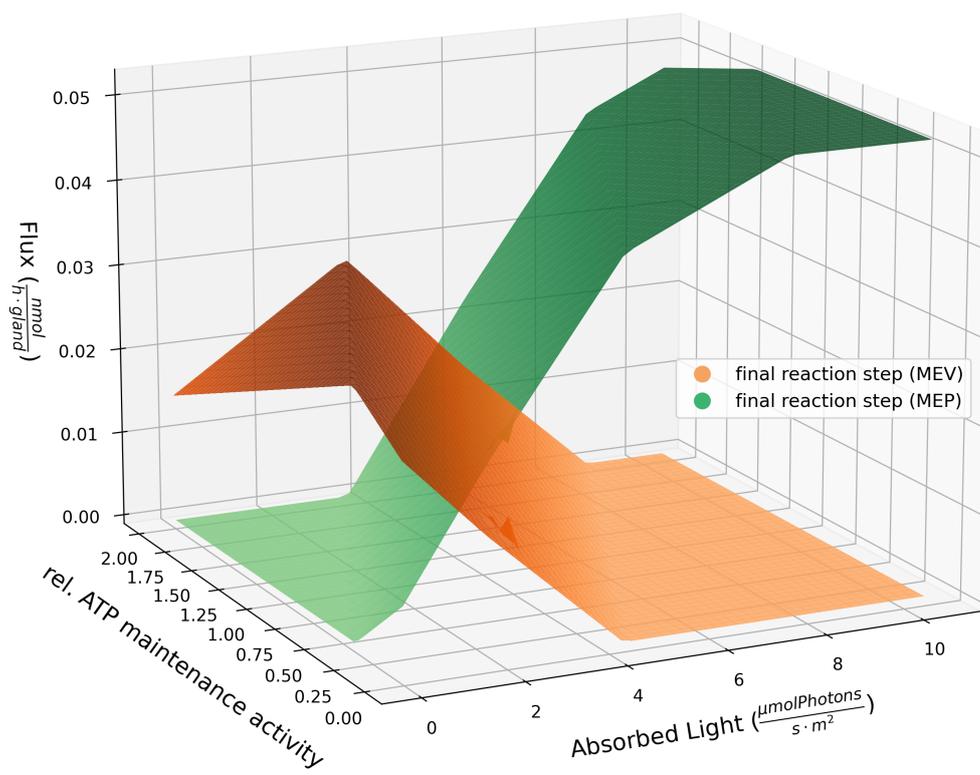
**FIGURE 2** Impact of light influx on the predicted fluxes through the photosynthetic glandular trichome. **a** Terpenoid synthesis flux over different rates of absorbed light. **b** Oxygen and carbon dioxide exchange fluxes over different rates of absorbed light.



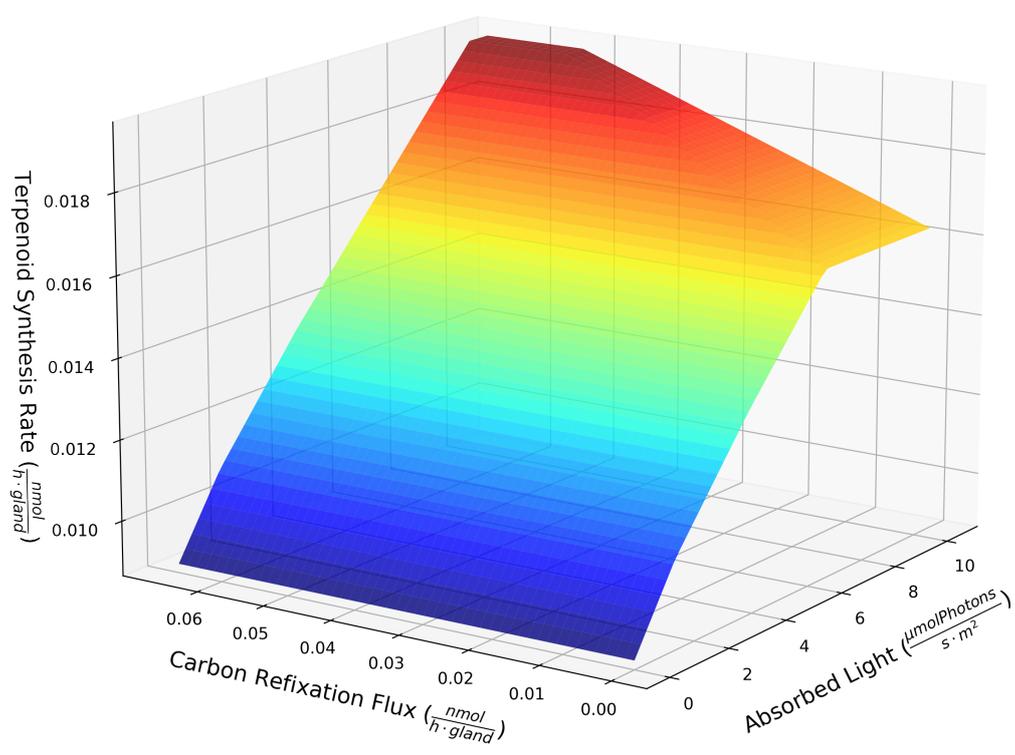
**FIGURE 3** The relative fluxes of eight selected catabolic reactions calculated for increasing fractions of saturating light. The fluxes are normalised to the respective fluxes under completely dark conditions.



**FIGURE 4** Relative consumption of GAP, pyruvate and Acetyl-CoA (here described as 3C bodies) by different pathways over increasing fractions of saturating light. The fraction of the lower glycolysis flux relative to dark conditions is displayed as a dotted line.



**FIGURE 5** Fluxes of the final reaction steps of the MEV and MEP pathway over increasing relative ATP maintenance activities, as well as increasing rates of light absorption.



**FIGURE 6** Predicted flux of the terpenoid synthesis under changing carbon refixation rates, as well as increasing rates of light absorption.

## 7 Discussion and Outlook

Photosynthesis and its light dependent and independent reactions are complex processes that are not working separately, but are connected in a tightly regulated supply-demand system. The implementation of mathematical models representing the dynamics of these systems is a powerful method to understand the underlying processes and the impact of each of the involved enzymes in different environmental conditions.

The theoretical and systematic investigation of photosynthetic processes is beneficial due to its promising use in crop design with focus on increased crop yield and plant stress resistance. With an increasing global population and consequential limitations in resources like food and energy, it has never been more important to optimise crops for improved bioeconomical usage.

In this work, the software packages `modelbase` and `moped` presented in chapters 6.1 and 6.2 have been developed and used to construct kinetic models of the photosynthetic apparatus, as well as a constraint-based genome scale metabolic model of photosynthetic glandular trichomes in tomatoes.

The work presented in chapter 6.3 focused on the analysis of a kinetic model of photosynthesis by merging models describing the photosynthetic electron transport chain and the CBB cycle. It was shown that photosynthesis can be viewed as a supply-demand system in which different light conditions and CBB cycle activities lead to changes in overall control between supply and demand processes in the system. Furthermore, the need for a "standby mode" for the CBB cycle in the form of thioredoxin regulation and pentose supply via the oxidative pentose phosphate pathway has been highlighted. The model results have been produced to represent specific experimental conditions, as well as chlorophyll contents. Depending on the cell type that is being simulated, certain parameter values may have to be substituted with measurements specific to the organism and cell of interest. However, the results that have been obtained are giving insights into the general principles of photosynthesis and have been supported by recent experimental findings [37]. Furthermore, the simulations have been performed with the assumption of carbon dioxide saturation. In conditions without carbon dioxide saturations, the control of RuBisCO would be a lot higher than in our simulation results, and would subsequently need an integration of photorespiration to the model. The results of this model analysis provided the groundwork for a functioning photosynthetic model

that opened additional research questions, especially with regards to the effect of alternative electron fluxes in the photosynthetic electron transport chain on the rest of photosynthesis.

The main focus of the work in chapter 6.4, is the behavior of linear electron flow and its alternatives (cyclic electron flow, PTOX and the Mehler reaction) in different light conditions and *in silico* mutants. For this, the model presented in 4.3 is further extended by modular addition of the ascorbate-glutathione cycle, as well as a simplified description of thioredoxin regulations. The results show that the cyclic electron flow is an important valve of the PETC and its activity prevents an overreduction of the PETC. This overreduction of PETC components otherwise leads to a strong increase in Mehler reaction rate and subsequent ROS evolution. Furthermore, an increased activity of key CBB cycle enzymes (like SBPase) systematically decrease the rate of ROS production via increased NADPH consumption and therefore providing more NADP<sup>+</sup> to function as the final electron acceptor for linear electron flow, therefore reducing overreductions in the PETC. The model results give crucial insights into the importance and impact of alternative electron flows on the photosynthetic apparatus in different conditions. However, in this work, the main focus was on the investigation of electron flows around photosystem I. In future work, adding NDH1 and AOX as modular extensions of the model could give a more detailed perspective on electron flows and valves apart from photosystem I, especially with focus on their roll in environmental stress conditions like drought-stress [53, 64].

Chapter 6.5 displays the work on the bioenergetics of photosynthetic glandular trichomes, for which a genome scale metabolic model has been developed by including transcriptomics and metabolomics data, metabolic databases and manual curation. The results show that energy and redox equivalents gained in photosynthetic activity allows the shift of carbon usage from catabolic to anabolic processes, therefore decreasing the loss of carbon in oxidative phosphorylation. This also leads to an increase in terpenoid production, as well as a shift from terpenoids produced by the MEV pathway to the MEP pathway. Furthermore, carbon dioxide produced in biosynthetic pathways can be refixed by the CBB cycle in light conditions that provide enough energy from photosynthetic activity. It is important to note that experimentally validating the simulated increase in terpenoid synthesis rate is very difficult due to the fact that the simulations are performed under the assumption that the influx of sucrose is constant in all light conditions. This may change in response to different levels of

irradiation in plants. Furthermore, the constraint based model is not capable of investigating which conditions and pathways change the composition of produced terpenoids. To answer this research question, a detailed kinetic model of the MEV and the MEP pathway is necessary.

In future work, the models that have been presented are to be improved for further research questions. The modular design of the kinetic models allow the addition and removal of reactions for specific research questions, therefore providing a scaffold for models that can be adjusted to be simplified and yet detailed instead of just expanding one existing model. For instance, the effect of dynamic oxygen and the addition of photorespiratory pathways and its effects on the control on the rest of the system. Furthermore, many research articles have displayed the regulation of photosynthetic components by phytohormone signalling to adjust changing environmental cues. The presented kinetic model can be used to describe the interdependence of photosynthetic efficiency and phytohormone crosstalk.

Adjusting the model for investigations in the aforementioned research topics requires experimental data of intermediate concentrations and photosynthetic efficiencies in plants that have been treated with phytohormones in different light intensities, as well as a good description of enzyme regulation by phytohormone activity. To ensure that the developed models can be used for such tasks, every model presented in this work is designed in a modular fashion and all models are available via open-access.

## References

- [1] Inger Andersson and Anders Backlund. Structure and function of rubisco. *Plant Physiology and Biochemistry*, 46(3):275–291, 2008.
- [2] Daniel I Arnon. The light reactions of photosynthesis. *Proceedings of the National Academy of Sciences*, 68(11):2883–2892, 1971.
- [3] Neil R Baker, Jeremy Harbinson, and David M Kramer. Determining the limitations and regulation of photosynthetic energy transduction in leaves. *Plant, Cell & Environment*, 30(9):1107–1125, 2007.
- [4] Yinon M Bar-On and Ron Milo. The global mass and average rate of rubisco. *Proceedings of the National Academy of Sciences*, 116(10):4738–4743, 2019.

- [5] Robert E Blankenship. *Molecular mechanisms of photosynthesis*. John Wiley & Sons, 2021.
- [6] Bob B Buchanan. The carbon (formerly dark) reactions of photosynthesis. *Photosynthesis research*, 128(2):215–217, 2016.
- [7] Stefano Caffarri, Tania Tibiletti, Robert C Jennings, and Stefano Santabarbara. A comparison between plant photosystem i and photosystem ii architecture and functioning. *Current Protein and Peptide Science*, 15(4):296–331, 2014.
- [8] Pascal-Antoine Christin and Colin P Osborne. The evolutionary ecology of c4 plants. *New Phytologist*, 204(4):765–781, 2014.
- [9] N D’Ambrosio, CR Guadagno, and A Santo. Is qe always the major component of non-photochemical quenching? In *Photosynthesis. Energy from the sun*, pages 1001–1004. Springer, 2008.
- [10] Oliver Ebenhöf, Geoffrey Fucile, Giovanni Finazzi, Jean-David Rochaix, and Michel Goldschmidt-Clermont. Short-term acclimation of the photosynthetic electron transfer chain to changing light: a mathematical model. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1640):20130223, 2014.
- [11] Oliver Ebenhöf, Torsten Houwaart, Heiko Lokstein, Stephanie Schlede, and Katrin Tirok. A minimal mathematical model of nonphotochemical quenching of chlorophyll fluorescence. *Biosystems*, 103(2):196–204, 2011.
- [12] Graham D Farquhar, S von von Caemmerer, and Joe A Berry. A biochemical model of photosynthetic co2 assimilation in leaves of c3 species. *planta*, 149(1):78–90, 1980.
- [13] Franziska Flügel, Stefan Timm, Stéphanie Arrivault, Alexandra Florian, Mark Stitt, Alisdair R Fernie, and Hermann Bauwe. The photorespiratory metabolite 2-phosphoglycolate regulates photosynthesis and starch accumulation in arabidopsis. *The Plant Cell*, 29(10):2537–2551, 2017.
- [14] Ulf-Ingo Flügge. Phosphate translocators in plastids. *Annual review of plant biology*, 50(1):27–45, 1999.
- [15] Laura S Green, Boihon C Yee, Bob B Buchanan, Kaeko Kamide, Yukika Sanada, and Keishiro Wada. Ferredoxin and ferredoxin-nadp reductase from photosynthetic and nonphotosynthetic tissues of tomato. *Plant Physiology*, 96(4):1207–1213, 1991.

- [16] Victor Guallar, Matt Jacobson, Ann McDermott, and Richard A Friesner. Computational modeling of the catalytic reaction in triosephosphate isomerase. *Journal of molecular biology*, 337(1):227–239, 2004.
- [17] Steinn Gudmundsson and Ines Thiele. Computationally efficient flux variability analysis. *BMC bioinformatics*, 11(1):1–3, 2010.
- [18] Alexander Hahn, Janet Vonck, Deryck J Mills, Thomas Meier, and Werner Kühlbrandt. Structure, mechanism, and regulation of the chloroplast atp synthase. *Science*, 360(6389):eaat4318, 2018.
- [19] Brian D Hahn. A mathematical model of the calvin cycle: analysis of the steady state. *Annals of botany*, 57(5):639–653, 1986.
- [20] David O Hall and Krishna Rao. *Photosynthesis*. Cambridge University Press, 1999.
- [21] Reinhart Heinrich and Stefan Schuster. *The regulation of cellular systems*. Springer Science & Business Media, 2012.
- [22] Stefan Henkes, Uwe Sonnewald, Ralf Badur, Ralf Flachmann, and Mark Stitt. A small decrease of plastid transketolase activity in antisense tobacco transformants has dramatic effects on photosynthesis and phenylpropanoid metabolism. *The Plant Cell*, 13(3):535–551, 2001.
- [23] Martin F Hohmann-Marriott, Robert E Blankenship, et al. Evolution of photosynthesis. *Annual review of plant biology*, 62(1):515–548, 2011.
- [24] Toshio Iwaki, Akira Wadano, Akiho Yokota, and Michio Himeno. Aldolase—an important enzyme in controlling the ribulose 1, 5-bisphosphate regeneration rate in photosynthesis. *Plant and cell physiology*, 32(7):1083–1091, 1991.
- [25] Pierre Joliot and Anne Joliot. Cyclic electron transfer in plant leaf. *Proceedings of the National Academy of Sciences*, 99(15):10209–10214, 2002.
- [26] Paul G Jones, Christine A Raines, and Julie C Lloyd. Nucleotide sequence of a wheat chloroplastic phosphoglycerate kinase gene. *Plant physiology*, 107(4):1483, 1995.
- [27] Zhenhui Kang, Tong Qin, and Zhiping Zhao. Thioredoxins and thioredoxin reductase in chloroplasts: A review. *Gene*, 706:32–42, 2019.

- [28] Edda Klipp, Wolfram Liebermeister, Christoph Wierling, and Axel Kowald. *Systems biology: a textbook*. John Wiley & Sons, 2016.
- [29] Jürgen Kopp, Stanislav Kopriva, Karl-Heinz Süß, and Georg E Schulz. Structure and mechanism of the amphibolic enzyme d-ribulose-5-phosphate 3-epimerase from potato chloroplasts. *Journal of molecular biology*, 287(4):761–771, 1999.
- [30] Nathan E Lewis, Kim K Hixson, Tom M Conrad, Joshua A Lerman, Pep Charusanti, Ashoka D Polpitiya, Joshua N Adkins, Gunnar Schramm, Samuel O Purvine, Daniel Lopez-Ferrer, et al. Omic data from evolved e. coli are consistent with computed optimal growth from genome-scale models. *Molecular systems biology*, 6(1):390, 2010.
- [31] Mónika Lingvay, Parveen Akhtar, Krisztina Sebők-Nagy, Tibor Páli, and Petar H Lambrev. Photobleaching of chlorophyll in light-harvesting complex ii increases in lipid environment. *Frontiers in plant science*, 11:849, 2020.
- [32] Heiko Lokstein, Gernot Renger, and Jan P Götze. Photosynthetic light-harvesting (antenna) complexes—structures and functions. *Molecules*, 26(11):3378, 2021.
- [33] John E Lunn and Elspeth MacRae. New complexities in the synthesis of sucrose. *Current opinion in plant biology*, 6(3):208–214, 2003.
- [34] Ulrich Lüttge. Ecophysiology of crassulacean acid metabolism (cam). *Annals of botany*, 93(6):629–652, 2004.
- [35] Timothy W Lyons, Christopher T Reinhard, and Noah J Planavsky. The rise of oxygen in earth’s early ocean and atmosphere. *Nature*, 506(7488):307–315, 2014.
- [36] Amane Makino, Chikahiro Miyake, and Akiho Yokota. Physiological functions of the water–water cycle (mehler reaction) and the cyclic electron flow around psi in rice leaves. *Plant and Cell physiology*, 43(9):1017–1026, 2002.
- [37] Alexander Makowka, Lars Nichelmann, Dennis Schulze, Katharina Spengler, Christoph Wittmann, Karl Forchhammer, and Kirstin Gutekunst. Glycolytic shunts replenish the calvin–benson–bassham cycle as anaplerotic reactions in cyanobacteria. *Molecular plant*, 13(3):471–482, 2020.

- [38] Anna Matuszyńska, Somayyeh Heidari, Peter Jahns, and Oliver Ebenhoeh. A mathematical model of non-photochemical quenching to study short-term light memory in plants. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1857(12):1860–1869, 2016.
- [39] Alan M McClain and Thomas D Sharkey. Triose phosphate utilization and beyond: from photosynthesis to end product synthesis. *Journal of Experimental Botany*, 70(6):1755–1766, 2019.
- [40] Henry M Mizioro. Phosphoribulokinase: current perspectives on the structure/function basis for regulation and catalysis. *Advances in enzymology and related areas of molecular biology*, 74:95–127, 2000.
- [41] Alejandro Morales, Xinyou Yin, Jeremy Harbinson, Steven M Driever, Jaap Molenaar, David M Kramer, and Paul C Struik. In silico analysis of the regulation of the photosynthetic electron transport chain in c3 plants. *Plant physiology*, 176(2):1247–1261, 2018.
- [42] Yuri Munekage, Mihoko Hashimoto, Chikahiro Miyake, Ken-Ichi Tomizawa, Tsuyoshi Endo, Masao Tasaka, and Toshiharu Shikanai. Cyclic electron flow around photosystem i is essential for photosynthesis. *Nature*, 429(6991):579–582, 2004.
- [43] Wojciech J Nawrocki, Nicolas J Tourasse, Antoine Taly, Fabrice Rappaport, Francis-André Wollman, et al. The plastid terminal oxidase: its elusive function points to multiple contributions to plastid physiology. *Annu Rev Plant Biol*, 66(1):49–74, 2015.
- [44] Dariusz M Niedzwiedzki and Robert E Blankenship. Singlet and triplet excited state properties of natural chlorophylls and bacteriochlorophylls. *Photosynthesis research*, 106(3):227–238, 2010.
- [45] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245–248, 2010.
- [46] Andrew A Pascal, Zhenfeng Liu, Koen Broess, Bart van Oort, Herbert van Amerongen, Chao Wang, Peter Horton, Bruno Robert, Wenrui Chang, and Alexander Ruban. Molecular basis of photoprotection and control of photosynthetic light-harvesting. *Nature*, 436(7047):134–137, 2005.
- [47] Gilles Peltier, Dimitri Tolleter, Emmanuelle Billon, and Laurent Cournac. Auxiliary electron transport pathways in chloroplasts of microalgae. *Photosynthesis Research*, 106(1):19–31, 2010.

- [48] Gosta Pettersson and Ulf RYDE-PETTERSSON. A mathematical model of the calvin photosynthesis cycle. *European Journal of Biochemistry*, 175(3):661–672, 1988.
- [49] Bernard E Pfeil, Benoît Schoefs, and Cornelia Spetea. Function and evolution of channels and transporters in photosynthetic membranes. *Cellular and Molecular Life Sciences*, 71(6):979–998, 2014.
- [50] Barbara Pfister and Samuel C Zeeman. Formation of starch in plant cells. *Cellular and Molecular Life Sciences*, 73(14):2781–2807, 2016.
- [51] Mark G Poolman, David A Fell, and Simon Thomas. Modelling photosynthesis and its control. *Journal of Experimental Botany*, 51(suppl\_1):319–328, 2000.
- [52] Christine A Raines. The calvin cycle revisited. *Photosynthesis research*, 75(1):1–10, 2003.
- [53] Sanna Rantala, Tapio Lempiäinen, Caterina Gerotto, Arjun Tiwari, Eva-Mari Aro, and Mikko Tikkanen. Pgr5 and ndh-1 systems do not function as protective electron acceptors but mitigate the consequences of psi inhibition. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1861(3):148154, 2020.
- [54] Alexander V Ruban. Plants in light. *Communicative & integrative biology*, 2(1):50–55, 2009.
- [55] Nima P Saadat, Tim Nies, Marvin Van Aalst, Brandon Hank, Büsra Demirtas, Oliver Ebenhöf, and Anna Matuszyńska. Computational analysis of alternative photosynthetic electron flows linked with oxidative stress. *Frontiers in plant science*, 12:750580, 2021.
- [56] Thomas D Sharkey. Discovery of the canonical calvin–benson cycle. *Photosynthesis Research*, 140(2):235–252, 2019.
- [57] Katsiaryna Skryhan, Libero Gurrieri, Francesca Sparla, Paolo Trost, and Andreas Blennow. Redox regulation of starch metabolism. *Frontiers in plant science*, 9:1344, 2018.
- [58] Mark Stitt and Tom ap Rees. Estimation of the activity of the oxidative pentose phosphate pathway in pea chloroplasts. *Phytochemistry*, 19(8):1583–1585, 1980.

- [59] Kenji Takizawa, Atsuko Kanazawa, and David M Kramer. Depletion of stromal pi induces high ‘energy-dependent’ antenna exciton quenching (qe) by decreasing proton conductivity at cfo-cf1 atp synthase. *Plant, cell & environment*, 31(2):235–243, 2008.
- [60] Masahiro Tamoi, Miki Nagaoka, Yoshiko Miyagawa, and Shigeru Shigeoka. Contribution of fructose-1, 6-bisphosphatase and sedoheptulose-1, 7-bisphosphatase to the photosynthetic rate and carbon flow in the calvin cycle in transgenic plants. *Plant and cell physiology*, 47(3):380–390, 2006.
- [61] Alexander N Tikhonov. The cytochrome b 6 f complex: biophysical aspects of its functioning in chloroplasts. In *Membrane protein complexes: structure and function*, pages 287–328. Springer, 2018.
- [62] Edelmira Valero, Hermenegilda Macià, Ildefonso M De la Fuente, José-Antonio Hernández, María-Isabel González-Sánchez, and Francisco García-Carmona. Modeling the ascorbate-glutathione cycle in chloroplasts under light/dark conditions. *BMC systems biology*, 10(1):1–20, 2015.
- [63] Floris J Van Eerden, Manuel N Melo, Pim WJM Frederix, Xavier Periole, and Siewert J Marrink. Exchange pathways of plastoquinone and plastoquinol in the photosystem ii complex. *Nature communications*, 8(1):1–8, 2017.
- [64] Greg C Vanlerberghe, Greg D Martyn, and Keshav Dahal. Alternative oxidase: a respiratory electron transport chain pathway essential for maintaining photosynthetic performance during drought stress. *Physiologia Plantarum*, 157(3):322–337, 2016.
- [65] John Whitmarsh et al. The photosynthetic process. In *Concepts in photobiology*, pages 11–51. Springer, 1999.
- [66] Thomas Wieloch, Roland Anton Werner, and Jürgen Schleucher. Carbon flux around leaf-cytosolic glyceraldehyde-3-phosphate dehydrogenase introduces a 13c signal in plant glucose. *Journal of Experimental Botany*, 72(20):7136–7144, 2021.
- [67] Wataru Yamori, Kouki Hikosaka, and Danielle A Way. Temperature response of photosynthesis in c3, c4, and cam plants: temperature acclimation and temperature adaptation. *Photosynthesis research*, 119(1):101–117, 2014.

- [68] Julia Zaks, Kapil Amarnath, David M Kramer, Krishna K Niyogi, and Graham R Fleming. A kinetic model of rapidly reversible nonphotochemical quenching. *Proceedings of the National Academy of Sciences*, 109(39):15757–15762, 2012.
- [69] Ning Zhang, Russell P Kallis, Robert G Ewy, and Archie R Portis Jr. Light modulation of rubisco in arabidopsis requires a capacity for redox regulation of the larger rubisco activase isoform. *Proceedings of the National Academy of Sciences*, 99(5):3330–3334, 2002.
- [70] Rong-guang Zhang, C Evalena Andersson, Alexei Savchenko, Tatiana Skarina, Elena Evdokimova, Steven Beasley, Cheryl H Arrowsmith, Aled M Edwards, Andrzej Joachimiak, and Sherry L Mowbray. Structure of escherichia coli ribose-5-phosphate isomerase: a ubiquitous enzyme of the pentose phosphate pathway and the calvin cycle. *Structure*, 11(1):31–42, 2003.
- [71] Xia Zhao, Tingting Chen, Baohua Feng, Caixia Zhang, Shaobing Peng, Xiufu Zhang, Guanfu Fu, and Longxing Tao. Non-photochemical quenching plays a key role in light acclimation of rice plants differing in leaf color. *Frontiers in plant science*, 7:1968, 2017.
- [72] Xin-Guang Zhu, Rafael Alba, and Eric de Sturler. A simple model of the calvin cycle has only one physiologically feasible steady state under the same external conditions. *Nonlinear Analysis: Real World Applications*, 10(3):1490–1499, 2009.
- [73] Xin-Guang Zhu, Eric De Sturler, and Stephen P Long. Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm. *Plant physiology*, 145(2):513–526, 2007.