Computational analysis and optimisation of photosynthetic carbon fixation

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Marvin Christian van Aalst
aus Düsseldorf, Deutschland

3. Juli 2024

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

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

Berichterstatter:
1. Prof. Dr. Oliver Ebenhöeh
2. Prof. Dr. Berkley Walker

Tag der mündlichen Prüfung: 10. Oktober 2024

Selbständigkeitserklärung

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, 3. Juli 2024

Marvin Christian van Aalst

Zusammenfassung

Photosynthetische Organismen verwenden Lichtenergie, um Kohlendioxid in Biomasse zu binden und bilden damit die Grundlage fast aller terrestrischen trophischen Netzwerke auf der Erde. Im Rahmen der Zivilisation werden photosynthetische Organismen auch als Nutzpflanzen und Stoffwechselfabriken, z. B. für die Produktion von Terpenen, genutzt. Die Menge an verfügbarem Licht, die für diese Zwecke genutzt wird, ist von großer Bedeutung, um die Produktion von Biomasse oder Metaboliten von Interesse zu maximieren. Dies wiederum ist für die Nachhaltigkeit und die Ernährungssicherheit von Bedeutung, da es sich auf die Effizienz der Wasser-, Nährstoff- und Landnutzung auswirkt. Um zu verstehen, wie die Effizienz der Lichtenergienutzung maximiert werden kann, muss der Prozess, wie diese Energie zur Bindung von Kohlendioxid genutzt wird, ganzheitlich verstanden werden. Dazu gehören die photosynthetische Elektronentransportkette (PETC), der Calvin-Benson-Bassham-Zyklus (CBB) und die Photorespiration. In dieser Arbeit wurden mathematische Modelle verwendet, um die Interdependenz und die Kontrolle von PETC, CBB und Photorespiration sowie die Produktion von Sekundärmetaboliten zu untersuchen.

Anhand eines Modells, das PETC und CBB kombiniert, konnte gezeigt werden, dass bei starkem Licht und hohem intrazellulären Kohlendioxid eine Erhöhung der SBPase-Aktivität die Produktion reaktiver Sauerstoffspezies (ROS) verringern kann. Ebenso erhöht eine Verringerung des zyklischen Elektronenflusses die Produktion von ROS, während gleichzeitig ATP verbraucht wird, was wiederum den CBB-Zyklus beeinträchtigt. Durch die Analyse der metabolischen Kontrolle wurde festgestellt, dass die Steuerung der Kohlenstofffixierung von der Lichtintensität und der Kohlendioxidkonzentration abhängt. Unter niedrigen Lichtbedingungen kontrollieren die Photosysteme den Fluss der Kohlenstofffixierung, während sich diese Kontrolle unter hohen Lichtbedingungen auf RuBisCO und SBPase verlagert, je nachdem, ob die intrazelluläre Kohlendioxidkonzentration niedrig oder hoch ist. Bei der Untersuchung der Interdependenz von CBB und Photorespiration wurde festgestellt, dass alternative Kohlenstofffixierungswege der vielversprechendste Weg sind, um die erhebliche Verringerung des Ernteertrags aufgrund der Photorespiration zu vermeiden. Dieser Effekt war jedoch unter Bedingungen mit niedriger intrazellulärer Kohlendioxidkonzentration am stärksten ausgeprägt, welche entweder durch eine niedrige atmosphärische Kohlendioxidkonzentration oder eine verringerte Transportrate entsteht. Schließlich hat die Untersuchung der Terpenoid-Synthese in photosynthetischen Drüsentrichomen gezeigt, dass die PETC eine Verschiebung der Kohlenstoffverteilung ermöglichen kann. Diese Verschiebung wird durch eine Verringerung der katabolen Aktivität aufgrund zusätzlicher Energiezufuhr verursacht, wodurch mehr des zugeführten Kohlenstoffs für anabole Prozesse verwendet werden kann.

Die Ergebnisse dieser Arbeit sollen sowohl zur Entwicklung effizienterer Nutz-

pflanzen für einen nachhaltigen Ansatz zur Ernährungssicherung als auch zur Entwicklung und Konzeption alternativer Wege im Allgemeinen beitragen.

Abstract

Photosynthetic organisms harvest light energy to fix carbon dioxide into biomass, creating the foundation of almost all terrestrial trophic networks on Earth. In the context of civilisation, photosynthetic organisms are also used as crops and metabolic factories, e.g. for the production of terpenes. The amount of available light used for these objectives is of great importance to maximise the production of biomass or metabolites of interest. This in turn is relevant to sustainability and food security, as it affects water, nutrient and land use efficicency. To understand how the efficiency of light energy usage can be maximised, the process of how that energy is used to fix carbon dioxide needs to be understood holistically. This includes the photosynthetic electron transport chain (PETC), the Calvin-Benson-Bassham (CBB) cycle and photorespiration. In this work mathematical models were used to study the interdependence and control of the PETC, CBB and photorespiration, as well as the production of secondary metabolites.

Using a model combining the PETC and CBB, it could be shown that under high light and high intracellular carbon dioxide an increase in SBPase activity could decrease the production of reactive oxygen species (ROS). Likewise, a reduction of cyclic electron flow increases the production of ROS, while also depleting ATP, thus affecting the CBB cycle. Expanding on this, using metabolic control analysis it was found that the control of carbon fixation is dependent on the light intensity and the carbon dioxide concentration. In low light conditions the photosystems control the flux of carbon fixation, while in high light conditions this control shifts to RuBisCO and SBPase, depending on whether the intracellular carbon dioxide concentration is low or high respectively. When studying the interdependence of the CBB and photorespiration it was found that carbon-fixing alternative pathways are the most promising way of avoiding the significant decrease in crop yield due to photorespiration. However, this effect was most pronunced in conditions with low intracellular carbon dioxide concentration, due to either by a low atmospheric carbon dioxide concentration or a reduced transport rate. Lastly, studying the terpenoid synthesis in photosynthetic glandular trichomes showed that the PETC can allow a shift in carbon partitioning. This shift is caused by a reduction in the catabolic activity due to additional energy supply which allows more of the supplied carbon to be used for anabolic processes.

It is envisaged that the findings of this work shall contribute to both the development of more effecient crops for a sustainable approach to food security and the development and design philosophy of alternative pathways in general.

Acknowledgments

I want to thank Prof. Dr. Oliver Ebenhöh for his outstanding intuition at finding just the right knob to turn when studying any system at hand. It will never cease to amaze me. Also a big thank you to Prof. Dr. Berkely Walker, who despite our geological distance has always found time for me and massively improved my scientific writing. Your cheerfulness even at the earliest hours of the day has inspired me to always try to go the extra mile when helping other people - even if I might not feel the best myself. I am always grateful for my friend and mentor Jun.-Prof Dr. Anna Matuszyńska. You are one of the very few people that manages to combine passion and commitment for your scientific research with great empathy for the people around you. You were always there for me, especially when I felt my worst, and I will always appreciate your ability to make me laugh at any given time. Dziękuję, że jesteś tak niesamowitym przyjacielem.

A big thank you to all members of the QTB team, for all the fun we had together. Thanks to my office mates Chilperic and Hettie for both all the times we laughed together and all the times we spent discussing almost every single topic imaginable. And especially thank you for staying friends even when we didn't agree! Thanks Ovidiu for your relentless cheerfulness and Nima for always making me laugh and think with your endless collection of weird stories and facts. I greatly enjoyed whenever we had the chance to work together. Thank you Tim for being amazingly hard to pinpoint. Whenever I think I understand you, there is just more to explore. I'm looking greatly forward to seeing how you will progress further in life! Thanks Yvan and Thomas for your ability to make silence enjoyable. As a naturally talkative person I have always enjoyed the calmness the two of you radiate so effortlessly. Thank you Mara for always keeping me as save as possible of the bureacratic monstrosity of German universities. I have always enjoyed your wisdom and hints and the little things that make life better. Special thanks to my dear friend Ellen, who I very much belive can singlehandedly run an entire institute while also caring for those in need and managing a couple of conferences. And all that on a Friday. It is a mystery to me how a single person can do all those things, while still being fun to hang around and genuinely kind. I wish you the very best for your future!

Last but not least, I thank my family, who are always there for me. Thank you for always supporting me and creating the core from which I draw my energy. I can be an annoying little twit sometimes, but I can always turn to you and receive nothing but love. Speaking of love, thank you to Nadine for making my life the best I could possibly imagine. You have both turned my head and life around and have done so many things to make me happy that I will forever be grateful that we have met.

x

Contents

1 Photosynthesis

Photosynthetic organisms are the basis of most modern life on earth, harnessing the sun's energy to fuse chemically inert carbon dioxide (CO_2) into usable building blocks, providing food for heterotrophic organisms and releasing oxygen as a byproduct. While the sequence of processes and biochemical reactions performing that conversion is tightly interlinked, it has been functionally separated into discrete pathways: the Photosynthetic electron transport chain (PETC), also called the light-dependent reactions and the Calvin-Benson-Bassham cycle (CBB cycle), also called carbon, light-independent or dark reactions[1] [1, 2]. The functions of these pathways are to supply energy and redox equivalents (PETC) and fix CO_2 (CBB cycle) into stable products which can be exported as triose-phosphates or hexoses [3]. In plants and algea the subcellular location for photosynthesis is the chloroplast, an organelle originally developed by endosymbiosis [4]. The chloroplast itself can be further divided into the stroma, which makes up for most of the volume, and the membrane-enclosed subcompartments called thylakoids [5]. The PETC takes place in the thylakoids, while the cbbc takes place in the chloroplast stroma.

As will become evident in the following sections, these processes are highly complex. One particular useful tool to handle such highly complex systems and coalesce large amounts of data and knowledge are mathematical models. As they allow to identify which fidelity is required to steady each system respectively, allowing simplification and systematic assessment of the hypotheses and theories at hand. In the following sections I will explain the main processes and current state of modelling for both the processes directly involved in photosynthesis (sections 1.1 and 1.2). Further, I will explain photorespiration - a process necessary to scavenge the results of a side-reaction of the CBB cycle, as well as nitrogen refixation and amino-acid production.

1.1 Photosynthetic electron transport chain

The photosynthetic apparatus is a highly complex system which converts light energy into chemical energy in the form of ATP and NADPH usable by the remainding plant metabolism [7]. The entire process of how this is achieved has been studied for hundreds of years and a complete description is far beyond the reach of this thesis. Thus, I will only introduce the necessary parts for sections 3.5 and 3.6 in the following, see Fig. 1.1 for an illustration. The reader is referred to excellent works such as by Blankenship for a full introduction into the topic [3].

The PETC mainly takes place in thylakoid membrane, with some reactions also

[1] The light-independent or "dark" reaction terminology originates from experiments being done in isolated chloroplast, which given supply of adenosine triphosphate (ATP) and Nicotinamide adenine dinucleotide phosphate (reduced) (NADPH) would perform carbon-fixation. However, *in planta* those processes are all but independent of the PETC, so this terminology is rather misleading, as their "independence" depended on an artifical supply of energy and redox equivalents.

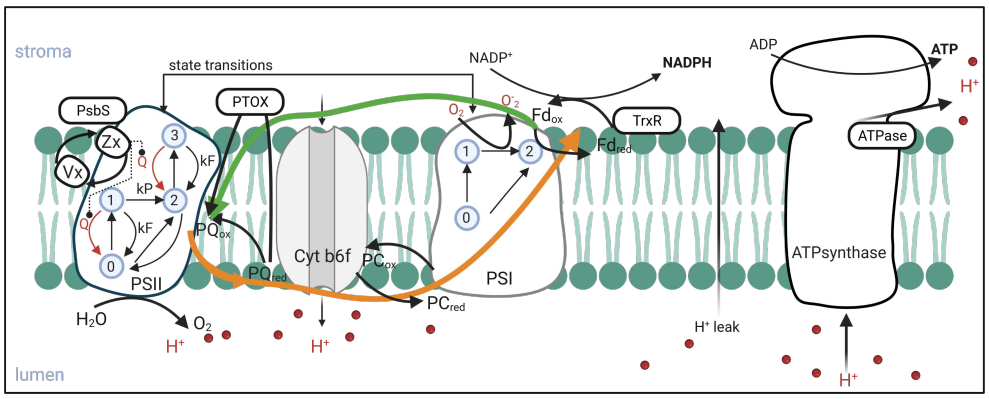


Figure 1.1: Schematic drawing of the PETC. Taken with permission from [6]

including the chloroplast stroma or the thylakoid lumen [8]. Light energy is collected by in antennae complexes, also called light harvesting complexes (LHCs), attached to photosystems I and II (PSI, PSII respectively). At PSII, two water molecules are split using the collected light energy, pushing four electrons into the PETC while producing molecular oxygen and four protons in the chloroplast lumen, causing it to acidify. Those electrons are then used to reduce plastoquinone to plastoquinol

$$2H_2O + 4hv + 2PQ + 4H^+_{stroma} \Rightarrow 2PQH_2 + O_2 + 4H^+_{lumen}$$

Subsequently, Plastoquinol is oxidised at the cytochrome b_6f complex. This complex uses the free electrons to reduce plastocyanin and pump further protons from the stroma into the lumen. As a consequence, lumen becomes even more acidic, creating a proton gradient between the stroma and the lumen [9]

$$2PC_{ox} + PQH_2 + 2H^+_{stroma} \Rightarrow 2PC_{red} + PQ + 4H^+_{lumen}$$

At PSI, light energy is used to drive electron transfer, this time from plastocyanin to ferredoxin

$$2PC_{red} + 2Fd_{ox} \Rightarrow 2PC_{ox} + 2Fd_{red}$$

Finally the ferredoxin:$NADP^+$ reductase uses the electrons from ferredoxin to reduce $NADP^+$ [10]

$$2Fd_{red} + NADP^+ + H^+_{stroma} \Rightarrow 2Fd_{ox} + NADPH$$

This sequence of reactions is called the linear electron flow and indicated with an orange arrow in figure 1.1. There is, however, a secondary route for electron transport called the cyclic electron flow, which instead of reducing $NADP^+$ reintroduces the electrons back at the cytochrome B_6f complex, reducing plastoquinone [11]

$$2Fd_{red} + PQ + 2H^+_{stroma} \Rightarrow 2Fd + 2PQH_2$$

The cyclic electron transport causes more protons to enter the thylakoid lumen and increases the proton gradient between the stroma and the lumen, while producing less NADPH. The proton gradient between the stroma and lumen creates a proton motive force, which is used by the ATP synthase (ATPSYN) to phosphorylate ADP [12].

$$3ADP + 3P_i + 14H^+_{lumen} \Rightarrow 3ATP + 3H_2O + 14H^+_{stroma}$$

The linear electron flow creates a constant ratio between the production of ATP and

NADPH. However, depending on the activity of the CBB cycle, photorespiration and other energy-dependent pathways, the ratio required by downstream metabolism differs from the fixed supply of energy and redox equivalents by the PETC. A dynamically activated cyclic electron flow allows to match the ratio of ATP and NADPH production to the demand of the downstream metabolism. While the activity of cyclic electron flow can match the ratio of supply of energy and redox-equivalents, the magnitude of that supply must also match the magnitude of the demand of the downstream metabolism.

Both CBB cycle and photorespiration are regulated by diverse molecular interactions. Particularly, the thioredoxin system plays a critical role orchestrating all parts of photosynthesis [13]. Thioredoxins can be reduced by thioredoxin reductase (TrxR) by either NADPH or ferredoxin. While their function is ubiquituous in all living organisms, in the case of photosynthesis their task is regulation of the catalytic activity of key enzymes. This is achieved as reduced thioredoxin reduces oxidised cysteine residues or cleaves disulfide bonds of enzymes, changing their catalytic activity [14]. In addition to the thioredoxin system, both the CBB cycle and photorespiration as the main consumers of energy and redox equivalents are dependent on the rubisco activity. The activity of rubisco is directly dependent on the supply of CO_2 [15]. If the supply of CO_2 is insufficient and thus the total demand of energy and redox equivalents does not match the supply by the PETC, excess energy will accumulate. This can lead to production of reactive oxygen species (ROS) by photoreduction of oxygen [16]. Those ROS cause oxidative damage to multiple cellular components, impairing growth and development [17]. Electron sinks can alleviate this problem by allowing excess energy to be dissipated.

One electron sink allowing dissipation of excess photon energy is the water-water cycle, where a plastid terminal oxidase (PTOX) oxidises plastoquinone, transferring its electrons and protons onto molecular oxygen, producing water [18, 19]:

$$2PQH_2 + O_2 + 4H^+_{stroma} \Rightarrow 2PQ + H_2O$$

This essentially is the reverse reaction that happens at PSII, in which water is split and oxygen produced, thus dissipating the excess energy in a futile way.

Another electron sink is the Mehler reaction, which occurs when PSI reduces oxygen instead of ferredoxin [20]. This causes formation of a superoxide radical, which is converted to hydrogen peroxide and oxygen by superoxide dismutase

$$2O_2^- + 2H^+ \Rightarrow H_2O_2 + O_2$$

In the peroxisome, hydrogen peroxide is disproportionated to water and oxygen by catalase, however in the chloroplast the ascorbate-glutathione cycle is used [16]. The ascorbate-glutathione cycle begins with a redox reaction of hydrogen peroxide and ascorbate (ASC) catalysed by ascorbate peroxidase (APX), creating monodehydroascorbate (MDA)

$$H_2O_2 + 2ASC + 2H^+ \Rightarrow 2MDA + H_2O$$

The monodehydroascorbate is in turn reduced by monodehydroascorbate reductase

(MDAR) recovering ascorbate

$$2\text{MDA} + \text{NADPH} + 2\text{H}^+ \Rightarrow 2\text{ASC} + \text{H}_2\text{O}$$

However, monodehydroascorbate can also spontaneously split into ascorbate and dehydroascorbate (DHA) $2\text{MDA} \Rightarrow \text{ASC} + \text{DHA} + 2\text{H}^+$ which then in turn has to be recycled by dehydroascorbate reductase (DHAR), which requires glutathione (GSG) oxidation to glutathione disulfide (GSSG)

$$\text{DHA} + 2\text{GSH} \Rightarrow \text{ASC} + \text{GSSG}$$

To replenish the glutathione, glutathione reductase uses NADPH as a reducing agent

$$\text{GSSG} + \text{NADPH} \Rightarrow \text{GSH} + \text{NADP}^+$$

While both these cycles allow dissipation of energy, the Mehler reaction produces harmful ROS. Their harm is highlighted by studies showing that catalase expression in the chloroplast enhances photo-oxidative stress tolerance [21].

Just as the demand of ATP and redox equivalents by the downstream metabolism can be smaller than the supply by the PETC, the supply of energy by the LHCs can exceed the capacity of the PETC. In that case there is an excess of excited chlorophyll molecules, which requires dissipation of energy additional to the normal photochemical quenching. This includes non-photochemical quenching (NPQ), which dissipates the excess energy as heat or fluorescence and ROS production by triplet state electrons. Thse in turn destroy the chlorophyll molecules and other structures, causing photobleaching [22, 23]. Included in NPQ are mechanisms over multiple time-spans that protect the plant against damage by excess illumination. First are the related pH-dependent mechanism qE and the zeaxanthin-dependent mechanism qZ, which are activated by a high proton gradient between the thylakoid lumen and stroma. The fast Psbs (photosystem II subunit S) component acts as a proton sensor and quickly changes the organisation of PSII and its LHC. This mechanisms dominates the total NPQ under short (<10 min) illumination times and in *Arabidopsis thaliana* is activate at roughly 900 $\frac{\mu mol}{m^2 \cdot s}$ [24]. If the proton gradient continues to be excessive, the slower Xanthophyll cycle is activated, further modifying the LHCs, which takes about 10 to 30 minutes [25]. This dual control allows quick responses to sunflecks as well as adaptation to different average illumination [26]. An additional mechanism to distribute excitation energy between PSI and PSII are state-transitions (qT) by shifting the distribution of LHCs between the photosystems, which in higher plants does not significantly contribute to NPQ (compared to green algea) [27, 28, 29]. The last component is photoinhibitory quenching (qI), which occurs at illumination of around 1800 $\frac{\mu mol}{m^2 \cdot s}$ is caused by ROS damage and degradation of reaction centers that occurs at longer illumination times over 30 minutes [30].

1.2 CBB cycle

The CBB cycle uses the energy provided by the PETC to fix CO_2 into stable products which can be exported as triose-phosphates or hexoses [3]. One turn of the cycle requires 3 ATP and 2 NADPH and thus the production of a triose 9 ATP & 6 ATP and

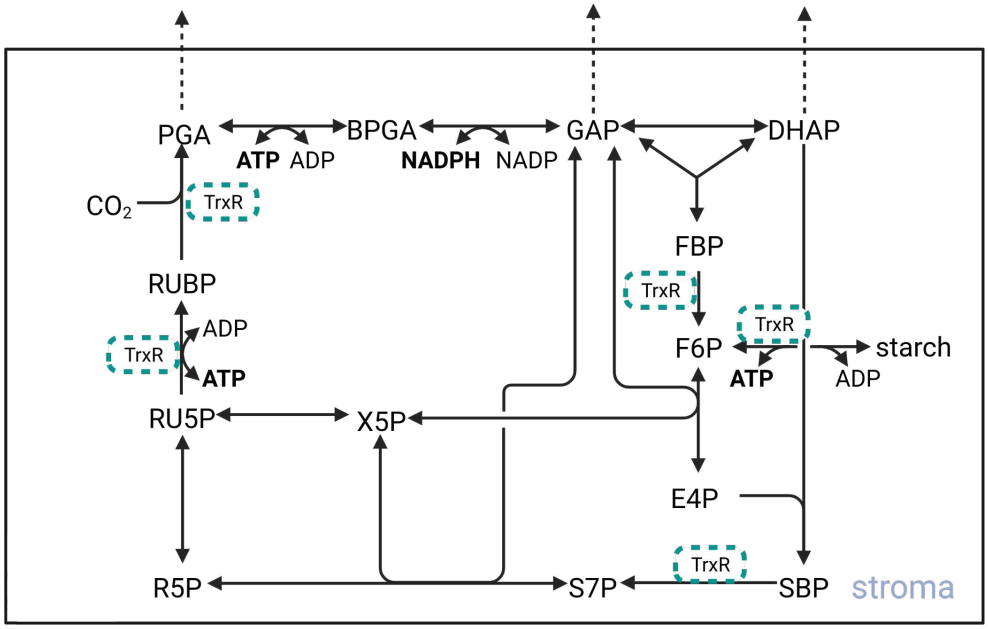


Figure 1.2: Schematic drawing of the CBB cycle. Taken with permission from [6]

the production of an hexose 18 ATP and 12 NADPH.

The cycle starts with ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), the main enzyme responsible for carbon flux into the biosphere and estimated to be the most abundant protein on Earth [31, 32, 33, 34]. It's reaction mechanism is quite complex. The enzyme tightly binds ribulose-1,5-bisphosphate (RUBP) until activation by rubisco activase and subsequent carbamylation and addition of Mg^{2+}. This in turn allows formation of the enediol enzyme complex, which can be oxygenated or carboxylated [35, 36, 37, 38, 39]. See the work by Farazdaghi for an excellent review on the reaction mechanisms [40]. In the case of the carboxylation reaction, one CO_2 is fixed and two molecules of 3-phospho-D-glycerate (PGA) are produced:

$$\mathrm{RUBP} + \mathrm{CO_2} \leftrightharpoons 2\ \mathrm{3PG}$$

while in the case of the oxygenation one PGA and one 2-phosphoglycolate (PGO) are produced, the latter of which needs to be recycled in a process called photorespiration, see section 1.3. The first few reactions in the CBB cycle are shared with gluconeogenesis and glycolysis (albeit in reverse order). This nicely illustrates the reversibility of most of the reactions, which will be notated using the symbol $\leftrightharpoons$, while irreversible reactions will be notated using the symbol $\Rightarrow$.

PGA is then phosphorylated by phosphoglycerate kinase (PGK) forming 2,3-bisphosphoglycerate (BPGA)

$$\mathrm{3PG} + \mathrm{ATP} \leftrightharpoons \mathrm{BPGA} + \mathrm{ADP}$$

which in turn is reduced by glyceraldehyde 3-phosphate dehydrogenase (GADPH) to form glyceraldehyde-3-phosphate (GAP), releasing inorganic phosphate:

$$\mathrm{BPGA} + \mathrm{NADPH} + \mathrm{H+} \leftrightharpoons \mathrm{GAP} + \mathrm{NADP^+} + \mathrm{P_i}$$

GAP is then isomerised to form dihydroxyacetone phosphate (DHAP) by triose-

phosphate isomerase (TPI)

$$\mathrm{GAP} \leftrightharpoons \mathrm{DHAP}$$

It is assumed that PGA, GAP and DHAP can all be exported into the cytosol [41]. While BPGA in principle can also be exported its concentration is usually too low and reactions producing and consuming it too fast for any significant transport process to be measurable. GAP and DHAP are then fused by aldolase (ALD), forming fructose-1,6-bisphosphate (FBP):

$$\mathrm{GAP} + \mathrm{DHAP} \leftrightharpoons \mathrm{FBP}$$

FBP is dephosphorylated by fructose-1,6-bisphosphatase (FBPase), releasing inorganic phosphate as a byproduct:

$$\mathrm{FBP} + \mathrm{H2O} \Rightarrow \mathrm{F6P} + \mathrm{P_i}$$

This is one of the few reactions in the CBB cycle that is irreversible at physiological concentrations with a $\Delta_r G'°$ of -11.8 $\frac{\mathrm{kJ}}{\mathrm{mol}}$ and was found to be regulated by the thioredoxin system [42, 15]. Thus in glycolysis it requires ATP hydrolysis and is thus replaced with phosphofructokinase.

fructose-6-phosphate (F6P) can now either be used for starch production, by using phosphoglucomutase to produce G6P, or by transketolase (TK1) to form erythrose-4-phosphate (E4P) and xylulose-5-phosphate (X5P).

$$\mathrm{F6P} + \mathrm{GAP} \leftrightharpoons \mathrm{E4P} + \mathrm{X5P}$$

E4P is then fused with DHAP to form sedoheptulose-1,7-bisphosphate (SBP)

$$\mathrm{E4P} + \mathrm{DHAP} \leftrightharpoons \mathrm{SBP}$$

which in turn is dephosphorylated by sedoheptulose-bisphosphatase (SBPase), forming sedoheptulose-7-phosphate (S7P) and releasing inorganic phosphate

$$\mathrm{SBP} + \mathrm{H2O} \Rightarrow \mathrm{S7P} + \mathrm{P_i}$$

Just like FBPase this reaction is irreversible at physiological concentrations with a $\Delta_r G'°$ of -13.1 $\frac{\mathrm{kJ}}{\mathrm{mol}}$ and regulated by the thioredoxin system [42, 15].

The next step in the CBB cycle is again a mixing-reaction by transketolase, this time producing ribose-5-phosphate (R5P) and X5P:

$$\mathrm{S7P} + \mathrm{GAP} \leftrightharpoons \mathrm{R5P} + \mathrm{X5P}$$

This repetition of the "fuse ⟶ control ⟶ mix" motif of using an aldolase followed by an irreversible dephosphorylation and transketolase causes a rapid interchange between the pools of GAP, DHAP, E4P and X5P and R5P.

The last steps of the CBB cycle repeat a similar motif of mixing followed by an irreversible reaction, by the irreversible conversion of either X5P or R5P to ribulose-5-phosphate (RU5P).

$$\mathrm{X5P} \leftrightharpoons \mathrm{RU5P}$$

$$\mathrm{R5P} \leftrightharpoons \mathrm{RU5P}$$

followed by the irreversible ($\Delta_r G'^\circ$ = -28.6 $\frac{\mathrm{kJ}}{\mathrm{mol}}$) and thioredoxin-regulated phosphoribulokinase (PRK) to regenerated RUBP:

$$\mathrm{RU5P} + \mathrm{ATP} \leftrightharpoons \mathrm{RUBP} + \mathrm{ADP}$$

A review and quantitative comparison of CBB cycle models can be found in the work of Arnold et al. [43]. In that review the models were categorised by the level of abstraction and detail they use. The models were ranked by a score consisting of the models stability (as in whether a steady-state at given conditions could be reached), compliance of metabolite concentrations to experimental data and robustness (as in whether small perturbations in kinetic parameters lead to small deviations from the steady state). Ranked first for models only concerned with carbon fixation is the work by Farquhar et al, followed by Medlyn et al, Schultz et al and Sharkey et al [44, 45, 46, 47].

For models that only describe processes within the chloroplast stroma, the highest ranking model is by Fridlyand et al, followed by Zhu et al [48, 49]. Lastly, for models describing cellular interactions as well, the highest ranked model is by Poolman et al, followed by Giersch et al, Damour et al, Hahn et al, Laisk et al (2006), Zhu et al, Woodrow et al and again Laisk et al (1989) [50, 51, 52, 53, 54, 55, 56, 57].

No matter which complexity a model has, it is important to consider where to put the boundary of description and what to assume about the processes outside of the boundary. For the CBB cycle, the main points of interchange are energy supply by the PETC, energy consumption by any downstream metabolism, export of triose phosphates, starch sucrose production and photorespiration.

None of the models above consider a dynamic PETC, although that has been achieved in the work by Matuszyńska et al, by coupling the model of the CBB cycle to models of the PETC and [58, 50, 59, 23].

Export of triose phosphates and starch production are described in a highly simplified manner in the work by Poolman et al [50], export of sucrose in the work by Laisk et al [57, 54]. None of the models above describe sucrose or starch storage in a dynamic manner, which means that questions like "where would the fixed carbon end up if plants are grown under continuous light and starch is accumulated to its maximum capacity" cannot be answered by any of these models [60]. Combining these models with models for starch metabolism is certainly an interesting open task [61, 62]. Similar considerations can be made for the kinetics of triose-phosphate export. McClain et al found that high concentration of stromal inorganic phosphate inhibits the rate of photosynthesis [63]. While the CBB cycle model by Poolman et al contains a static description of external inorganic phosphate, no dynamic description is given [50].

Of all the models studied by Arnold et al, only the one by Zhu et al contains a description of photorespiration [43, 55] Up to 20 % of rubisco activity can be oxygenation and some of the intermediates in photorespiration can inhibit both the CBB cycle and PETC [64, 65]. This means that of the 15 models studied by Arnold et al, only the model by Zhu et al should show realistic behaviour under

atmospheric conditions [43].

1.3 Photorespiration

Photorespiration is a scavenging pathway for PGO, which is produced by an oxygenation of RUBP by RuBisCO:

$$RUBP + O_2 \leftrightharpoons 3PG + PGO$$

The pathway consists of reactions in three compartments (chloroplast, peroxisome and mitochondrion), transporting metabolites via the cytosol [66]. Photorespiration gained more importance over the last 2 billion years, as photosynthesis originated in an almost anoxic, but CO_2-abundant atmosphere but over time, photosynthesis caused a significant shift in the composition of the atmosphere, resulting in a now roughly 500-fold excess of over CO_2[2]. While rubisco has a higher affinity for CO_2 than for , the excess in atmospheric causes up to 20 % of rubisco activity to be oxygenation rather than carboxylation. Whereas 3PG can directly replenish the CBB cycle, PGO inhibits multiple CBB cycle enzymes and is known to decouple the PETC [67, 68]. Thus expectedly, higher photorespiratory flux (e.g. by overexpressing glycine decarboxylase) can enhance photosynthesis by preventing the accumulation of inhibitory intermediates [64, 65, 69].

The next step is a dephosphorylation of PGO to glycolate (GLYCO) by phosphoglycolate phosphatase (PGP), releasing inorganic phosphate (P_i) in the process:

$$PGO + H2O \leftrightharpoons glycolate + P_i$$

Gycolate is then transported from the chloroplast stroma into the peroxisome and oxidised by glycolate oxidase (GLCOX), reducing oxygen and producing hydrogen peroxide in the process, which in turn is disproportionated by catalse:

$$GLYC \leftrightharpoons GLYOX + H_2O_2$$

$$H_2O_2 \leftrightharpoons H2O + 0.5\ O_2$$

The relocation of phosphoglycolate phosphatase (PGP) from the chloroplast to the peroxisome is beneficial for two reasons. Firstly, hydrogen peroxide is a ROS, which are known to damage various plant structures. In contrast to the chloroplast the peroxisome is highly specialised in the savenging of ROS, thus the potential negative effects are alleviated. Secondly, glyoxylate is known to impair rubisco activiation, so production outside the chloroplast is probably beneficial [70, 71, 72].

As the next step, glyoxylate is aminated to produce glycine. This is done by both glycine transaminase (GTA) while deaminating glutamate to 2-oxoglutarate and by serine-glyoxylate transaminase (SGTA), using serine as the amino-donor.

$$glyoxylate + glutamate \leftrightharpoons glycine + 2-oxoglutarate$$

$$glyoxylate + serine \leftrightharpoons glycine + hydroxypyruvate$$

[2] Assuming 20 % atmospheric oxygen and 400 ppm carbon dioxide

This double-usage of glyoxylate is a potentially problematic pathway motif, as either reaction can drain the other one of its substrate, creating a dead-lock in the cycle. This is especially true for Serine which is produced in a later step of the pathway and thus perturbations of the production and consumption rate of it might be problematic. In contrast, the glycine transaminase (GTA) is bound by the ratio of glutamate and 2-oxoglutarate, which are also used as a donor-acceptor pair by the GS/GOGAT[3] system. The GS/GOGAT system is essential for refixing the nitrogen lost by the glycine decarboxylase complex, the next step in the photorespiration pathway, which can be lumped together like the following

$$2\ \text{glycine} + \text{NAD} \leftrightharpoons \text{serine} + \text{CO}_2 + \text{NADH} + \text{NH4}^+$$

In this step two 2-carbon bodies are fused into one 3-carbon body, releasing one CO_2 and ammonium in the process and thus indirectly has the biggest energy cost of the entire pathway, as refixing the CO_2 costs 3 ATP and 2 reduction equivalents and refixing the nitrogen costs 1 ATP and 1 reduction equivalents.

After serine has been deaminated by serine-glyoxylate transaminase, hydroxypyruvate, the product of the reaction, is reduced by glutamate dehydrogenase (GDH) to form glycerate

$$\text{hydroxypyruvate} + \text{NADH} \leftrightharpoons \text{glycerate} + \text{NAD}$$

which in turn is phosphorylated by glycerate kinase (GK) to form 3PG

$$\text{glycerate} + \text{ATP} \leftrightharpoons \text{3PG} + \text{P}_\text{i}$$

which can again be used to replenish the CBB cycle.

Overall, photorespiration looses one CO_2, while costing 3.5 ATP and 2 reduction equivalents, compared to the 3 ATP and 2 reduction equivalents required to fix one CO_2 by the CBB cycle. This high energy cost, together with the high flux through photorespiration has been predicted to cause up to 20 % yield decrease in wheat and 36 % in soybean in the US [73]. Considering this yield loss, it is not surprising that considerable work has been done to improve on this. While there are decades on research on further optimising rubisco to either increase the catalytic rate or specificity, there seems to be growing concensus that modifications done to rubisco either make the enzyme more specific at the cost of catalytic rate or vice versa, suggesting pareto-optimality [74, 75, 76]. Notably, plants have not evolved an alternative to carbon fixation by rubisco either, but developed quite complicated mechanisms to either spatially or temporally separate carbon fixation from the formation of the first organic compounds, called C4-photosynthesis and CAM-photosynthesis respectively [77, 78, 79]. Another approach is to find alternative photorespiratory pathways that either avoid the cost of nitrogen refixation, avoid loosing the CO_2 altogether, and some very promising approaches to even fix an additional CO_2 albeit at higher energy costs. See section 3.1 for a detailed review of the currently existing pathway designs and section 3.2 for a computational comparison of all currently known alternative pathways.

[3] glutamine synthetase and glutamine oxoglutarate aminotransferase respectively

Arnold et al provide an extensive overview of modelling photorespiratory metabolism using algebraic equations, ordinary differential equations (ODEs) and genome-scale stoichiometric models [80]. As my work focusses on models based on ODEs, I will only review those models. The earliest, albeit highly simplified, models of both the CBB cycle and photorespiration have been presented by Brian Hahn [81, 53, 82], while the earliest complete description of photorespiration in isolation can be found in the work of Akiho Yokota [83], which is not discussed in the work by Arnold et al [80]. These were extended by Zhu et al and Zhao et al to contain the sucrose production pathway [55, 84]. Laisk et al further included a highly simplified description of the PETC [54]. The more recent worki of Bellasio et al resorts to highly simplifying the pathways again [85]. As Arnold et al mention, most of the models above highly simplify photorespiration and in general neglect the role of photorespiration in amino acid metabolism as well as nitrogen metabolism [80]. While these carbon-centric approaches have their merit in understanding the role of photorespiration in regard to energy consumption and central carbon metabolism, a more detailed view at their role in especially amino acid metabolism needs to be provided, especially if the design of alternative pathways is to be tested.

1.4 Goal of this thesis

The goal of this thesis is to understand how the central carbon metabolism and energy management of plants work and how they can be optimised for various purposes. For insights into plant metabolism the tight coupling of the CBB cycle to the PETC is analysed (section 3.5 and 3.6) For the work on optimising the crop yield see sections 3.1 and 3.2 and for terpene production section 3.7.

2 Software development

2.1 Motivation

In 2005, John Ioannidis published an article called "Why most published research findings are false", raising awareness to the lack of repeatability of published laboratory research findings now termed the reproducibility crisis [86]. While the paper doesn't explicitly mention theoretical work, there certainly are numerous theoretical publications which don't supply sufficient information in order to be repeated or are partly erroneous. Further, excluding the social incentives to conduct low-quality science or scientific misconduct[1], the fact remains that humans make mistakes. A particular well-known example of this is the example of excel formula errors in the work of Reinhart and Rogoff, falsely causing politicians to believe that high public debt ratios cause sharp declines in GDP growth [87, 88]. As Marc Branch writes: "if science is to regain its just position, it has to correct the problem of unrepeatable research results" [89]. While there are complex systemic problems in science that need to be addressed, there are also smaller problems that we as a community can solve much more easily. My contribution to that regard is providing software for building and analysis metabolic models.

Building and analysing computational metabolic models is a mentally taxing task, requiring constant context-switching between biological abstraction and implementation of the abstraction in the preferred programming language. While some of that complexity is inherent to the problem, we found that a lot of code for model construction and analysis is highly repetitive. This presents two problems: first, the person writing the code has to do a lot of unnecessary work to build their model and second, the person checking the code has to try to to spot a potential problem in a potentially huge amount of boilerplate code. The solution here is simple: if the way the models are build is standardised, then most of the subsequent analysis can be re-used, which at the same time dramatically increases the chances of errors to be found and corrected.

The software packages that I developed or contributed to are *a*) modelbase (section 2.2), which provides tools to construct and analyse models based on ODEs, *b*) moped (section 2.3), which provides tools to construct genome-scale models, *c*) dismo (section 2.4), which provides tools to construct and analyse discrete spatial ODE models and *d*) cobrexa (section 2.5), which provides analysis tools for models exceeding genome-scale.

[1] and misconceptions regarding p-values

2.2 modelbase

The `modelbase` Python package consists of a set of tools to standardise construction and analysis of models based on ODEs. While originally designed and implemented by Anna Matuszyńska and Oliver Ebenhoëh, I took over the development and maintenance of the project. This has led to two peer-reviewed publications, for which I am second and first author respectively.

Publication	Building Mathematical Models of Biological Systems with modelbase
Author position	Second
Status	peer-reviewed and published
doi	10.5334/jors.236

Publication	Constructing and analysing dynamic models with modelbase v1.2.3: a software update
Author position	First
Status	peer-reviewed and published
doi	10.1186/s12859-021-04122-7

Journal of
open research software

Ebenhöh, O, et al. 2018 Building Mathematical Models of Biological Systems with modelbase. *Journal of Open Research Software,* 6: 24. DOI: https://doi.org/10.5334/jors.236

SOFTWARE METAPAPER

Building Mathematical Models of Biological Systems with modelbase

Oliver Ebenhöh[1,2], Marvin van Aalst[1], Nima P. Saadat[1], Tim Nies[1] and Anna Matuszyńska[1,2]

[1] Institute for Quantitative and Theoretical Biology, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, DE

[2] Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, DE

Corresponding author: Anna Matuszyńska (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:

$$\frac{dS}{dt} = \mathrm{N}v(S, k), \tag{1}$$

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

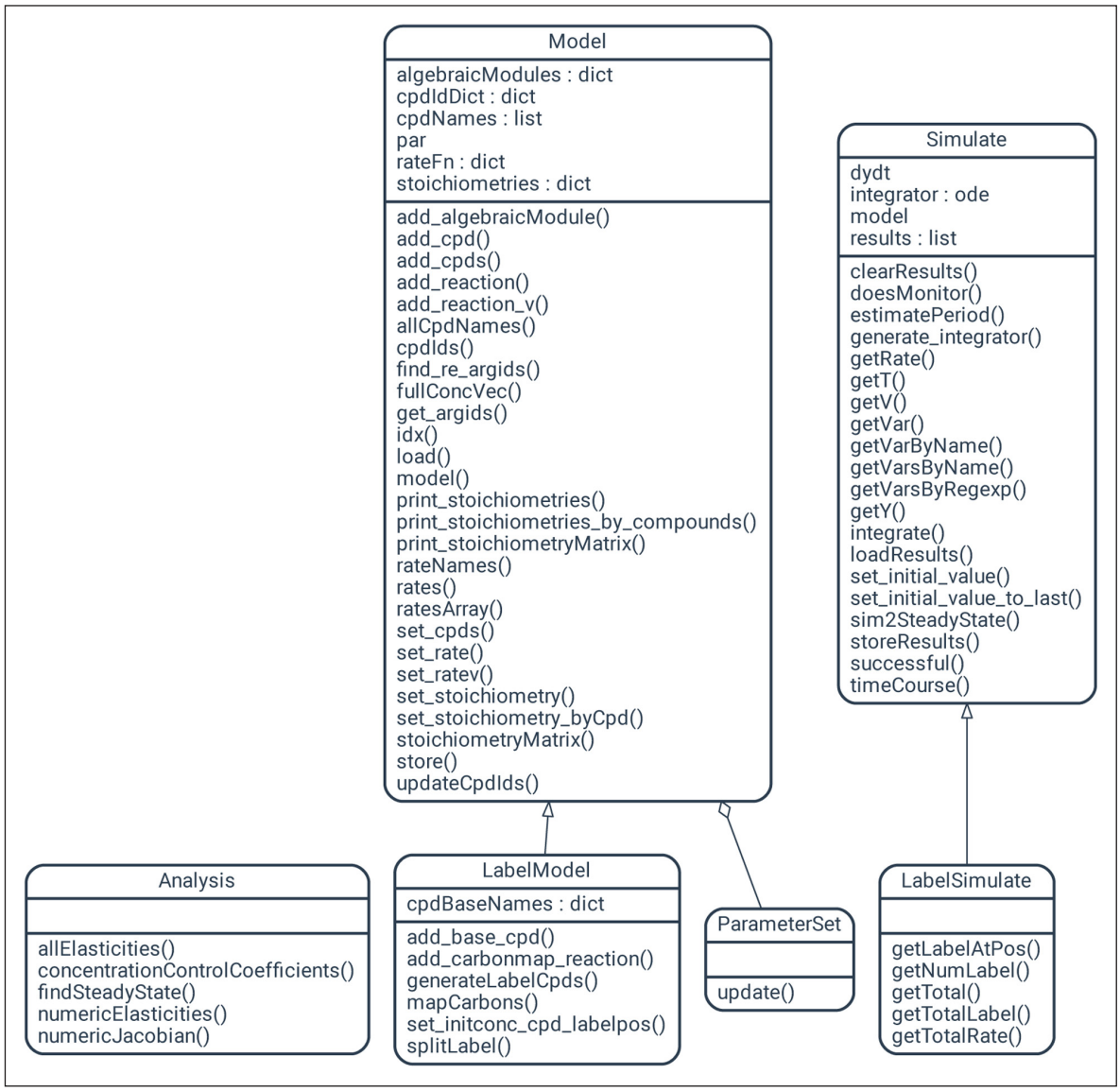


Figure 1: UML class diagram of `modelbase` software components. It consists of six classes, with *LabelModel* inheriting from *Model* and *LabelSimulate* inheriting from *Simulate*. *ParameterSet* and *Analysis* are special classes containing parameter sets and static methods for analysis respectively.

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

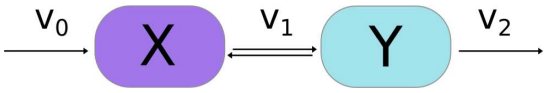
A particularly useful function of the class *Model* has been developed to facilitate the incorporation of algebraic expressions, by which dependent variables can be computed from independent ones. Examples include conserved quantities, such as the sum of adenine phosphates, which is often considered 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()`

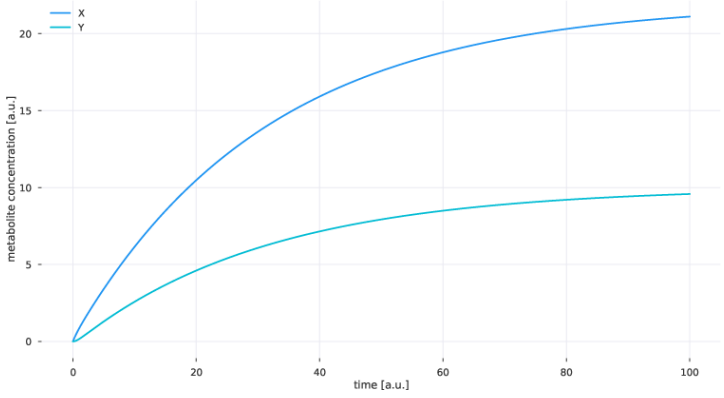
```
s = modelbase.Simulator(m)
```

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(file_name). A minimal working example can be found in our repository (https://gitlab.com/ebenhoeh/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

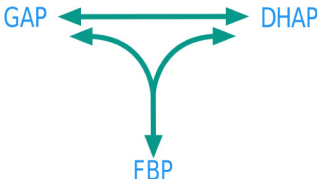
modelbase is written in the Python programming language, a general-purpose interpreted, interactive, object-oriented, and high-level programming language. It is available for every major operating system, including GNU/Linux, Mac OSX and Windows and has been tested with Python 3.6.

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:



$$\mathrm{GAP} \rightleftharpoons \mathrm{DHAP} \quad (2)$$
$$\mathrm{GAP} + \mathrm{DHAP} \rightleftharpoons \mathrm{FBP} \quad (3)$$

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 v1f(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',v1f,[2,1,0],['GAP'],['DHAP'],'GAP')
```

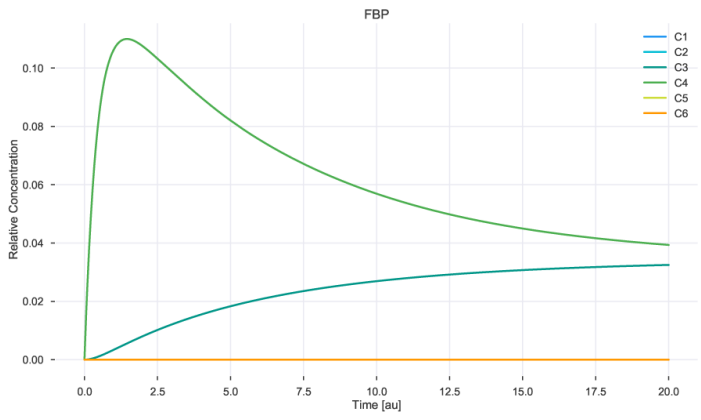
We set the initial conditions such that the total pools are in equilibrium, but carbon 1 of GAP is fully labeled

```
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-libsbml == 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, $NADP^+$, CO_2 and H^+ 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]	./petc-modelbase	A.M.
Calvin-Benson-Bassham Cycle	[16]	./cbb-modelbase	M.v.A.
Pentose Phosphate Pathway	[4, 17]	./ppp-modelbase	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 *in silico* replication of ^{13}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**).

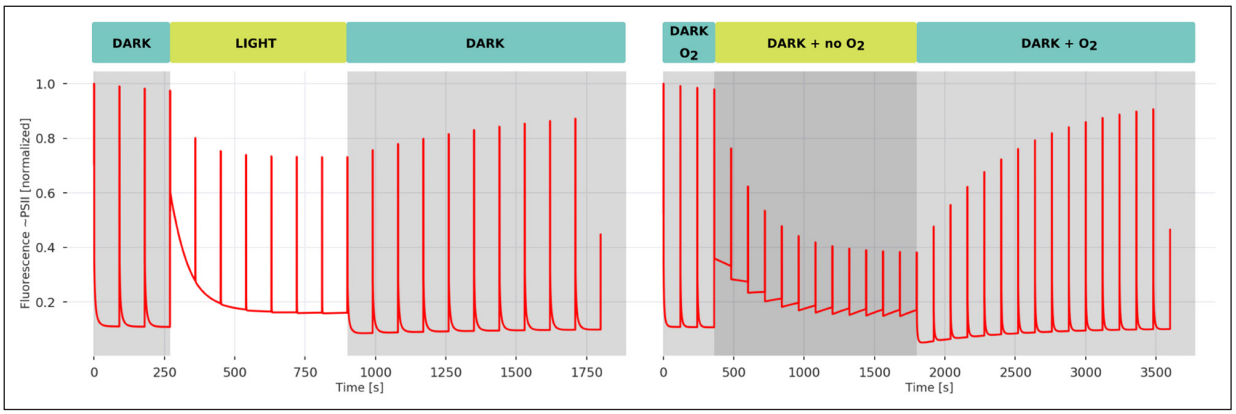


Figure 2: Reproduction of the Figures from [15]. Simulated fluorescence trace obtained through Pulse Amplitude Modulation (PAM) under light induced (left) and anoxia induced (right) conditions. The dynamics of the fluorescence decrease corresponds to the activation of a specific photoprotective mechanism 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.

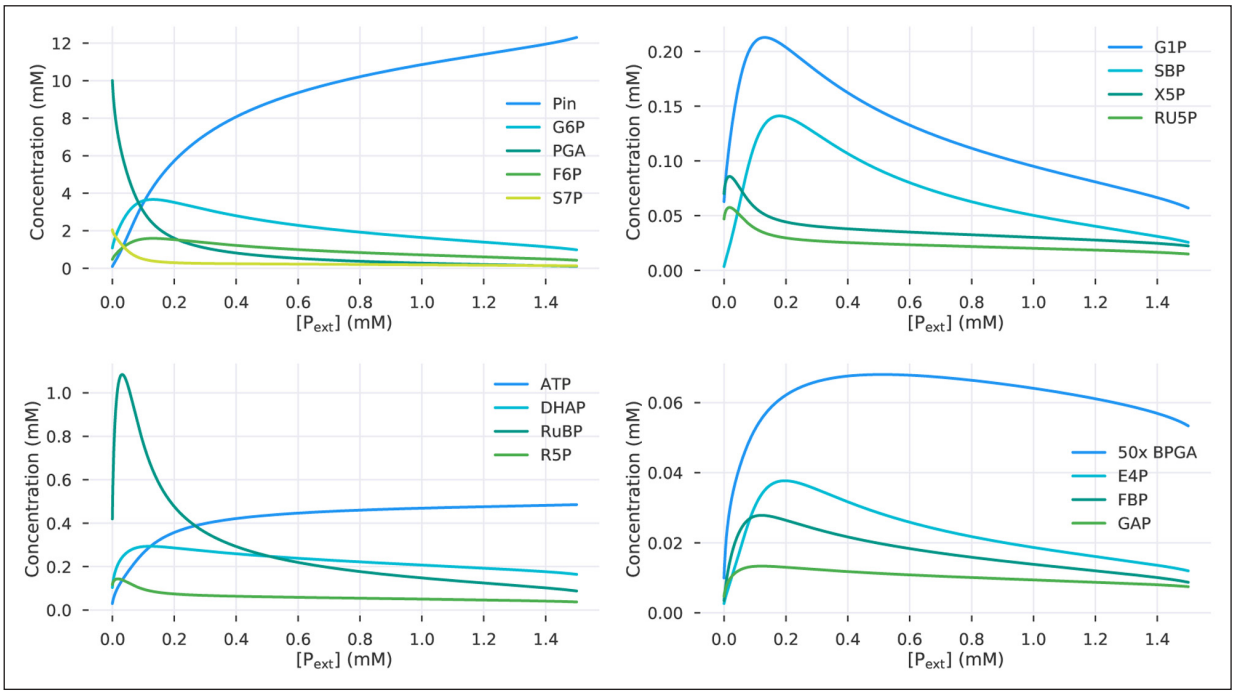


Figure 3: Metabolite steady state concentrations dependent on the extra-stromal phosphate concentration simulated with the Poolman implementation of the Pettersson and Ryde-Pettersson model of the CBB cycle [16].

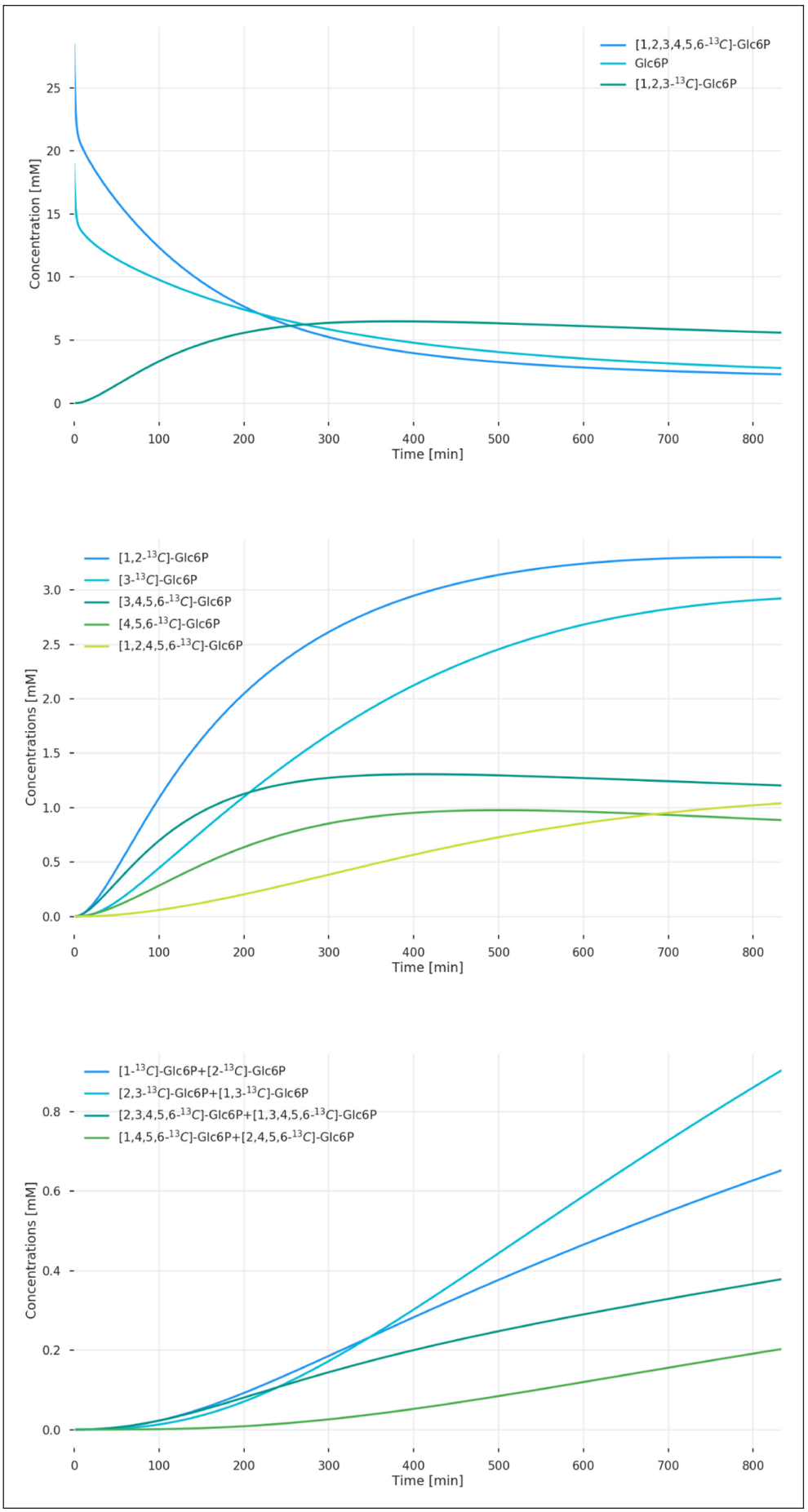


Figure 4: Formation of diverse Glc6P isotopomers in a haemolysate, obtained by solving the adapted model by Berthon *et al.* [17] reimplemented using `modelbase`.

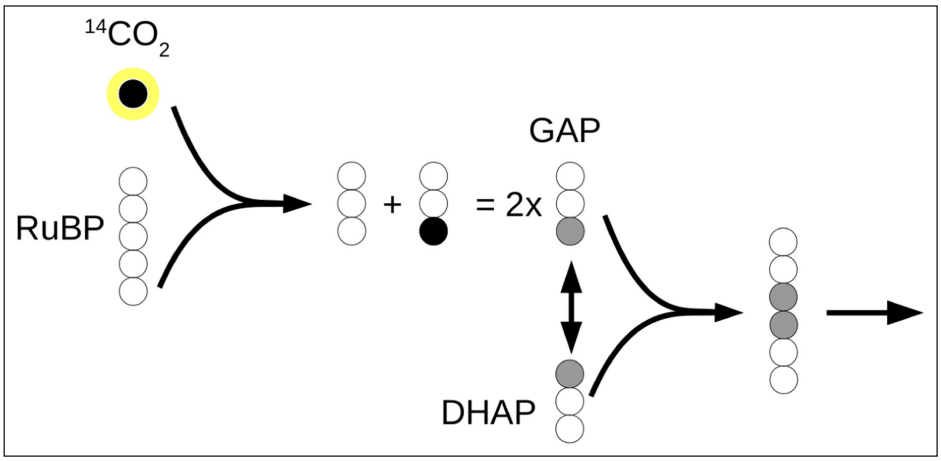


Figure 5: Schematic representation of label incorporation by the CBB cycle intermediates.

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 $^{14}CO_2$ 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/ebenhoeh/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. **Pettersson, G** and **Ryde-Pettersson, 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 31P-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. Acd. 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: COnstraints-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
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.bbabio.2016.09.003
20. **Crown, S B, Ahn, W** and **Antoniewicz, M R** 2012 Rational design of ^{13}C-labeling experiments for metabolic ux 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, Bussonnier, 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 workows. 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 modelbase. *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

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

OPEN ACCESS

van Aalst *et al. BMC Bioinformatics* (2021) 22:203
https://doi.org/10.1186/s12859-021-04122-7

BMC Bioinformatics

SOFTWARE Open Access

Constructing and analysing dynamic models with modelbase v1.2.3: a software update

Marvin van Aalst[1], Oliver Ebenhöh[1,2] and Anna Matuszyńska[1,2*]

*Correspondence:
anna.matuszynska@hhu.de
[2] CEPLAS - Cluster of Excellence on Plant Sciences, Universitätsstr. 1, 40225 Düsseldorf, Germany
Full list of author information is available at the end of the article

Abstract

Background: Computational mathematical models of biological and biomedical systems have been successfully applied to advance our understanding of various regulatory processes, metabolic fluxes, effects of drug therapies, and disease evolution and transmission. Unfortunately, despite community efforts leading to the development of SBML and the BioModels database, many published models have not been fully exploited, largely due to a lack of proper documentation or the dependence on proprietary software. To facilitate the reuse and further development of systems biology and systems medicine models, an open-source toolbox that makes the overall process of model construction more consistent, understandable, transparent, and reproducible is desired.

Results and discussion: We provide an update on the development of `modelbase`, a free, expandable Python package for constructing and analysing ordinary differential equation-based mathematical models of dynamic systems. It provides intuitive and unified methods to construct and solve these systems. Significantly expanded visualisation methods allow for convenient analysis of the structural and dynamic properties of models. After specifying reaction stoichiometries and rate equations modelbase can automatically assemble the associated system of differential equations. A newly provided library of common kinetic rate laws reduces the repetitiveness of the computer programming code. `modelbase` is also fully compatible with SBML. Previous versions provided functions for the automatic construction of networks for isotope labelling studies. Now, using user-provided label maps, `modelbase` v1.2.3 streamlines the expansion of classic models to their isotope-specific versions. Finally, the library of previously published models implemented in `modelbase` is growing continuously. Ranging from photosynthesis to tumour cell growth to viral infection evolution, all these models are now available in a transparent, reusable and unified format through `modelbase`.

Conclusion: With this new Python software package, which is written in currently one of the most popular programming languages, the user can develop new models and actively profit from the work of others. `modelbase` enables reproducing and replicating models in a consistent, tractable and expandable manner. Moreover, the expansion of models to their isotopic label-specific versions enables simulating label propagation, thus providing quantitative information regarding network topology and metabolic fluxes.

Keywords: Research software, Mathematical modelling, ODE, Metabolic networks, Systems biology, Systems medicine, Biomedical systems, Flux analysis, Labelling, Isotope tracing

Background

Mathematical models are accepted as valuable tools for advancing biological and medical research [1, 2]. In particular, models based on ordinary differential equations (ODEs) have found application in a variety of fields. Most recently, deterministic models simulating the dynamics of infectious diseases gained the interest of the general public during the Covid-19 pandemic. Consequently, a large number of ODE based mathematical models were developed and discussed, even in nonscientific journals [3–5]. Such focus on mathematical modelling is not surprising, because computational models allow for methodical investigations of complex systems under fixed, controlled and reproducible conditions. Hence, the effect of various perturbations of the systems can be inspected systematically in silico.

The model building process itself plays an important role in integrating and systematising vast amounts of available information [6]. Properly designed and verified computational models serve various purposes. They are used to develop hypotheses to guide the design of new research experiments (e.g., in immunology to study lymphoid tissue formation [7]). Models can also support metabolic engineering efforts (e.g., identification of enzymes to enhance essential oil production in peppermint [8]). More recently, models contribute to tailoring medical treatment to individual patient in the spirit of precision medicine (e.g., in oncology [2]). Finally, modelling results guide political decision making and governmental strategies (see the review on the impact of modelling for European Union Policy [9]). Considering their potential impact, models must be openly accessible so that they can be verified and corrected, if necessary.

In many publications, modelling efforts are justified by the emergence of extraordinary amounts of data provided by new experimental techniques. However, arguing for the necessity of model construction only because a certain type or amount of data exists, ignores several important aspects. Computational models are generally a result of months, if not years, of intense research, which involves gathering and sorting information, simplifying numerous details and distilling out the essentials, implementing the mathematical description in computer code, carrying out performance tests, and, finally, validating the simulation results. Our understanding of many phenomena could be deepened if, instead of constructing yet another first-generation model, we could efficiently build on the knowledge that was systematically collected in previously developed models. Moreover, the knowledge generated during the model construction process is often lost, e.g. because the main developer left the research team.

Preservation of the information collected in the form of a computational model has become an important quest in systems biology, and has, to some extent, been addressed by the community. Development of the Systems Biology Markup Language (SBML) [10] for unified communication and storage of biomedical computational models, and the existence of the BioModels repository [11] already ensures the survival of models beyond the academic life of their developers, or the lifetime of the software used to create them. However, a model in SBML format rarely reveals the logic of model construction. The

structure of `modelbase` code promotes consistent and transparent description of the model components (such as reaction rates), hence the logic of construction becomes inherently clear. Such knowledge loss can be prevented by providing simple-to-use toolboxes that enforce a universally readable model construction format.

For these reasons we developed `modelbase`[12], a Python package that allows the user to easily document the model building process. On the one hand, we defined the core of the model construction process, while on the other hand, the software does not make these definitions too strict, and fully integrates the model construction process into the Python programming language. This differentiates `modelbase` from many other Python-based modelling tools (such as ScrumPy [13], PySCeS [14], PySB [15] or tellurium [16, 17]) and other mathematical modelling languages (recently reviewed from a software engineering perspective by Schölzel and colleagues [18]). We would in particular like to stress a fundamental difference in the philosophy of `modelbase`, which distinguishes it from the other Python-based tools. In ScrumPy, PySCeS and tellurium, models are objects that are constructed by either SBML import or by a human-readable string (e.g. the Antimony representation chosen in tellurium [17]), which have methods for their numeric simulations and analysis. However, once constructed, the objects are not designed to be further modified. A modular design of different, but similar models, which all depend on sets of analogous modules, is thus difficult to represent. PySB aims at providing systematic construction methods, adding e.g. 'monomers' and 'rules' how these are converted. However, PySB deliberately ignores and overrides standard Python behaviour, making it difficult to keep multiple models in one namespace. In `modelbase`, models and simulations are two different types of objects. In analogy to experiments, a model corresponds to the biological entity, such as a cell, whereas a simulation corresponds to a particular experiment that is performed on the entity. A model object can be arbitrarily modified by numerous methods. Typically, a model is systematically constructed by instantiating an empty model object, to which components are added by dedicated methods. In this way, the model construction process remains maximally transparent, is fully integrated into the Python programming language, and is completely reproducible. Flexibility to modify and alter the model structure (incl. parameters) is ensured in this way.

Here we report new features in `modelbase` v1.2.3, developed over the last two years. We have significantly improved the interface to make model construction easier and more intuitive. The accompanying repository of replicated, published models, available from our GitLab project, has been considerably expanded, and now includes a diverse selection of biomedical models (see Additional file 1: Table 2). This diversity highlights the general applicability of our software. Essentially, every dynamic process that can be described by a system of ODEs can be implemented with `modelbase`.

Implementation

`modelbase` is a Python package to facilitate construction and analysis of ODE based mathematical models of biological systems. Version 1.2.3 introduces changes not compatible with the previous official release, version 0.2.5 [12]. All API changes are summarised in the official documentation hosted by ReadTheDocs and the differences between the versions are summarised in Table 1.

Table 1 Key changes between the first published version of `modelbase` [12] and the current update

Functionality	`modelbase` 0.2.5	`modelbase` 1.2.3
Initialization	Model takes only parameters as an argument	Model takes as arguments all other model components as dictionaries
Parameters	Hidden as a private attribute	Replaced with a vanilla dictionary
Derived parameters	No function to calculate from other model parameters	Called on initialization and prior to any numerical operations
Handling of time-dependent reactions	Time given in kwargs	Modifiers argument is introduced
Simulation	Integration via `timeCourse` that takes an array of time points for Simulation	Integration via `simulate` that takes only the endpoint of the simulation, default starting point t = 0, otherwise starts where the last simulation ended
Labelling features	Focused on carbon labelling problems	Reference of the word carbon was changed to label
Method to calculate relative label distribution in steady-state	None available	Via `LinearLabelModel`
Scan steady-state concentrations depending on parameter values	None available	Via `parameter_scan`
SBML support	Export of model stoichiometries	Import of models that match the capabilities of `modelbase` and full export of models using `ratelaw`
Metabolic Control Analysis support	None available	A full suite of methods to calculate and plot elasticities via `mca` module
Predefined kinetic laws	None available	Via `ratelaw` module
Plotting support	Time course plots	Phase-plane analysis

The model building process starts by creating a modelling object of the dedicated Python class `Model` and adding to it the chemical compounds of the system. Then, following the intuition of connecting the compounds, the reaction network is constructed by adding the reactions one by one. Each reaction requires stoichiometric coefficients and a kinetic rate law. The latter can be provided either as a custom function or by selecting one from the newly provided library of rate laws. The usage of this library (`ratelaws`) reduces the repetitiveness by avoiding boilerplate code. It requires the user to explicitly define reaction properties, such as directionality. This contributes to a systematic and understandable construction process, following the second guideline from the Zen of Python, the guiding principles for Python's design[1]: "Explicit is better than implicit".

From this, `modelbase` automatically assembles the system of ODEs. It also provides numerous methods to conveniently retrieve information about the model. In particular, the `get_*` methods can be used to inspect all the components of the model, and calculate reaction rates for given concentration values. These functions have multiple variants that return all the common data structures (array [19], dictionary, data frames [20]).

[1] python-c "import this".

After the model building process is completed, simulation and analyses of the model are performed with the `Simulator` class. Currently, we offer interfaces to two integrators to solve stiff and non-stiff ODE systems. Provided the Assimulo package [21] is available, as recommended in our installation guide, `modelbase` will use CVode, a variable-order, variable-step multi-step algorithm. The CVode class provides a direct connection to Sundials [22] which is a powerful industrial solver and robust time integrator, with a high computing performance. If Assimulo is not available, `modelbase` will automatically use the SciPy library [23]. Specifically lsoda will be used to integrate the model, which in our experience resulted in lower computing performance [24]. The whole process of assembling a model has been summarised in Fig. 1.

Metabolic control analysis

Sensitivity analysis provides a theoretical foundation to systematically quantify the effects of small parameter perturbations on global system behaviour. In particular, Metabolic Control Analysis (MCA), initially developed to study metabolic systems, is an important and widely used framework providing quantitative information about the response of the system to perturbations [25, 26]. The new version of `modelbase` has a full suite of methods to calculate elasticities. These can be plotted as a heat-map, giving a clear and intuitive colour-coded visualisation of the results. An example of such visualisation, for a re-implemented toy model of the upper part of glycolysis (Section 3.1.2 [27]), can be found in Fig. 2.

Visualisation support

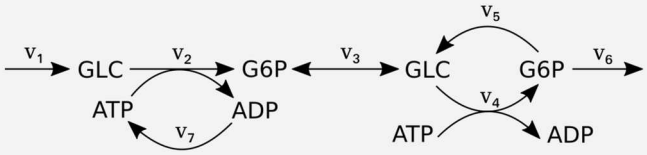
Many of the existing software packages for building computational models restrict the users by providing unmodifiable plotting routines with predefined settings that may not suit their personal preferences. In `modelbase` v1.2.3 plotting functions allow the user to pass optional keyword-arguments (often abbreviated as **kwargs), similar to Tellurium [17]. All plot elements are accessible and available for change, providing a transparent and flexible interface to the commonly used matplotlib library [28]. The easy access functions that visualise the results of simulations were expanded from the previous version. They now include funcitonality to plot selections of compounds or fluxes, phase-plane analysis and the results of MCA. An example of the latter is included in Fig. 2.

Models for isotope tracing

`modelbase` has also been developed to aid the in silico analyses of label propagation during isotopic studies. To simulate the dynamic distribution of isotopes, all possible labelling patterns for all intermediates need to be created. By providing an atom transition map in the form of a list or a tuple, all 2^N isotope-specific versions of a chemical compound are created automatically, where N denotes the number of possibly labelled atoms. Changing the name of the previous function `carbonmap` to `labelmap` in v1.2.3 acknowledges the diversity of possible labelling experiments that can be reproduced with models built using our software.

1. How to build and analyse a model with `modelbase`

We use `modelbase` to simulate a toy model of glycolysis proposed in [27], and presented on the scheme below.



First we import modules and classes from the `modelbase` , `numpy` and `matplotlib` packages

```
from modelbase.ode import Model, LabelModel, LinearLabelModel, Simulator, mca
from modelbase.ode import ratelaws as rl
from modelbase.ode import ratefunctions as rf
import numpy as np
import matplotlib.pyplot as plt
```

and instantiate the `modelbase` object

```
m = Model()
```

Next we add a dictionary of parameters and a list of state variables (compounds):

```
m.add_parameters({"k1":0.25,"k2":1,"k3":1,"k3m":1,"k4":1,"k5":1,"k6":1,"k7":2.5})
m.add_compounds(["GLC", "G6P", "F6P", "FBP", "ATP", "ADP"])
```

In the next step we define reaction rate functions. We use the build-in ratelaws function with predefined kinetic laws. As all reaction rates in the toy model are simple mass action kinetics we add 7 reactions to the model m

```
m.add_reaction(rate_name="v1", function=rf.constant, stoichiometry={"GLC": 1}, parameters=["k1"])
m.add_reaction(rate_name="v2", function=rf.mass_action_2,
    stoichiometry={"GLC": -1, "ATP": -1, "G6P": 1, "ADP": 1}, parameters=["k2"])
m.add_reaction(rate_name="v3", function=rf.reversible_mass_action_1_1,
    stoichiometry={"G6P": -1, "F6P": 1}, parameters=["k3", "k3m"],
    reversible=True)
m.add_reaction(rate_name="v4", function=rf.mass_action_2,
    stoichiometry={"F6P": -1, "ATP": -1, "ADP": 1, "FBP": 1}, parameters=["k4"])
m.add_reaction(rate_name="v5", function=rf.mass_action_1,
    stoichiometry={"FBP": -1, "F6P": 1}, parameters=["k5"])
m.add_reaction(rate_name="v6", function=rf.mass_action_1,
    stoichiometry={"FBP": -1}, parameters=["k6"])

m.add_reaction(rate_name="v7", function=rf.mass_action_1,
    stoichiometry={"ADP": -1, "ATP": 1}, parameters=["k7"])
```

To perform the computation we generate an instance of the Simulator class

```
s = Simulator(m)
```

and provide the dictionary with initial concentrations to start integration

```
y0 = {"GLC": 0, "G6P": 0,"F6P": 0, "FBP": 0,"ATP": 0.5,"ADP": 0.5}
s.initialise(y0)
t, y = s.simulate(100)
```

The results of above simulation will be displayed automatically using the plot() function of the method. Do not forget to label your axes.

```
fig, ax = s.plot()
plt.show()
```

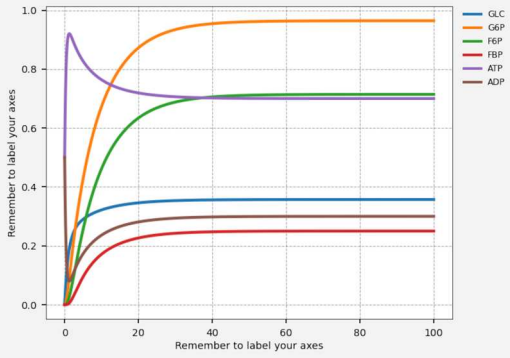


Fig. 1 An example of how to build and analyse a model with `modelbase`

Isotope tracing under stationary conditions

Sokol and Portais derived the theory of dynamic label propagation under the stationary assumption [29]. In steady-state, the space of possible solutions is reduced and the labelling dynamics can be represented by a set of linear differential equations. We have used this theory and implemented an additional class `LinearLabelModel` that

2. How to use the visualisation methods within the mca package

With the model object `m` from the previous example we can perform Metabolic Control Analysis and visualise the results in form of a heat map. To reproduce this results run Notebook upper-glycolysis.ipynb from Additional file 2.

```
parameters = m.get_parameters().keys()

crc = mca.get_concentration_response_coefficients_df(
    model=m, parameters=parameters, y=y0, normalized=True
).loc[:, ["GLC", "G6P", "F6P", "FBP", "ATP", "ADP"]]

frc = mca.get_flux_response_coefficients_df(
    model=m, parameters=parameters, y=y0, normalized=True)

fig, axs = mca.plot_multiple(
    dfs=[crc, frc],
    titles=["Concentration Response Coefficients", "Flux Response Coefficients"])
```

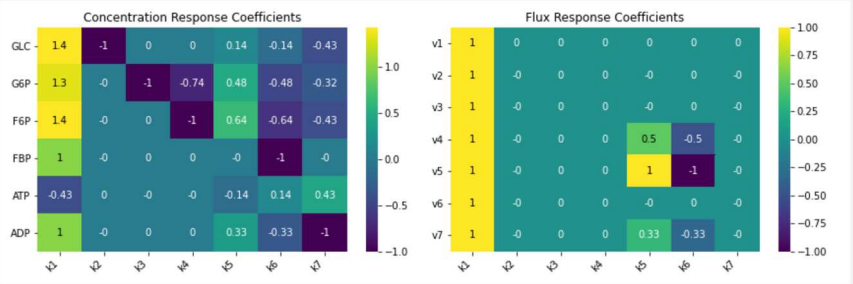


Fig. 2 An example of how to use the visualisation methods within the mca package

3. How to use meta data functionality

We construct here a model with one compound (D-ribulose-1,5-bisphosphate) and one reaction (RuBisCO). Instead of using the biochemical compound name, we call it x and provide all additional information, such as the common name, and link to existing data bases as additional keyword arguments.

```
m = Model()
m.add_compound(
    "x",
    **{
        "common_name": "RuBP",
        "formula": "C5H8O11P2",
        "database_links": {"kegg": 'C01182'},
    }
)
m.add_rate("v1", lambda *args: 0, **{"common_name": "RuBisCO"})

m.meta_info['compounds']
```

will return all provided information

```
>>> {'x': Compound(common_name='RuBP', compartment='c', formula='C5H8O11P2',
charge=None, gibbs0=None, smiles=None, database_links={'kegg': 'C01182'},
notes={})}
```

Fig. 3 An example of how to use metadata functionality

allows rapid calculation of the label propagation given the steady-state concentrations and fluxes of the metabolites [29]. `modelbase` can automatically build the linear label model from user provided label maps. An example of such a model is provided in Fig. 3, where we simulate label propagation in a linear non-reversible pathway, see Fig. 1 in [29] for comparison. The linear label models are constructed using `modelbase` rate laws, and hence can be fully exported as an SBML file.

Model metadata

Many models lose their readability due to the inconsistent, intractable or misguided naming conventions. An example is a model with reactions named v1-v10, without referencing them properly. By providing metadata fields for all `modelbase` objects, the

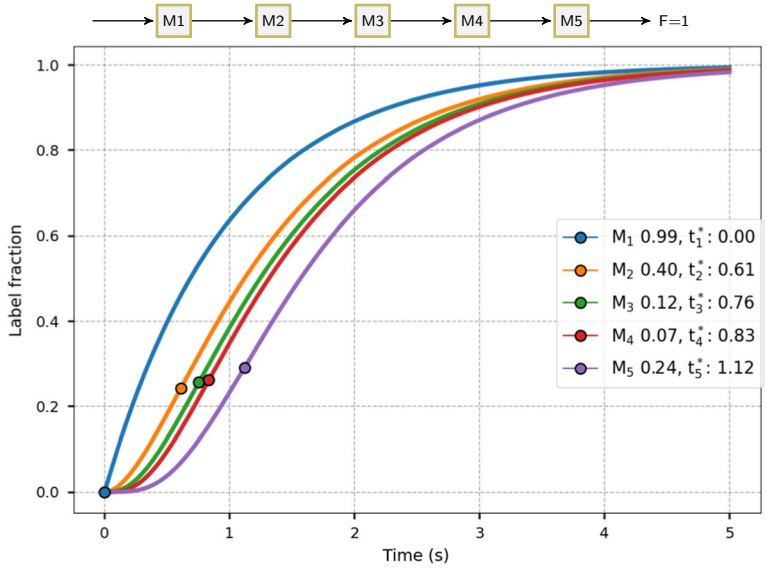


Fig. 4 Labelling curves in a linear non-reversible pathway. Example of label propagation curves for a linear non-reversible pathway of five randomly sized metabolite pools, as proposed in the paper by Sokol and Portais [29]. Circles mark the position at which the first derivative of each labelling curve reaches maximum. In the original paper, this information has been used to analyse the label shock wave (LSW) propagation. To reproduce these results run the label-propagation-2015.ipynb Jupyter Notebook from the Additional file 4: Jupyter Notebook label-propagation-2015.ipynb

user can abbreviate component names in a personally meaningful manner and supply additional annotation information in accordance with recognised standards, such as MIRIAM [30]. An example of how to use metadata functionality is provided in Fig. 4. This interface can also be used to supply additional information, such as the unit of a parameter.

SBML support

In contrast to the previous `modelbase` version, where we only supported the export of stoichiometric models to SBML format, we now support both import and export of kinetic models. The full summary of the SBML concepts supported by `modelbase` is documented in the official SBML test suite, where the output of our tests is stored. Examples where SBML models are imported and exported, using our `build_model_from_sbml` and `write_sbml_model` functions, are supplied in the `modelbase` documentation.

Results and discussion

With the newly implemented changes, `modelbase` is more versatile and user friendly. As argued before, its strength lies in its flexibility and applicability to virtually any biological system with dynamics that can be described using an ODE system. There exist countless mathematical models of biological and biomedical systems derived using ODEs. Many of these models are rarely re-used, at least not to the extent that could be reached, if models were shared in a readable, understandable and reusable way [18]. Our package can be used efficiently both for the development of new models, as well as the

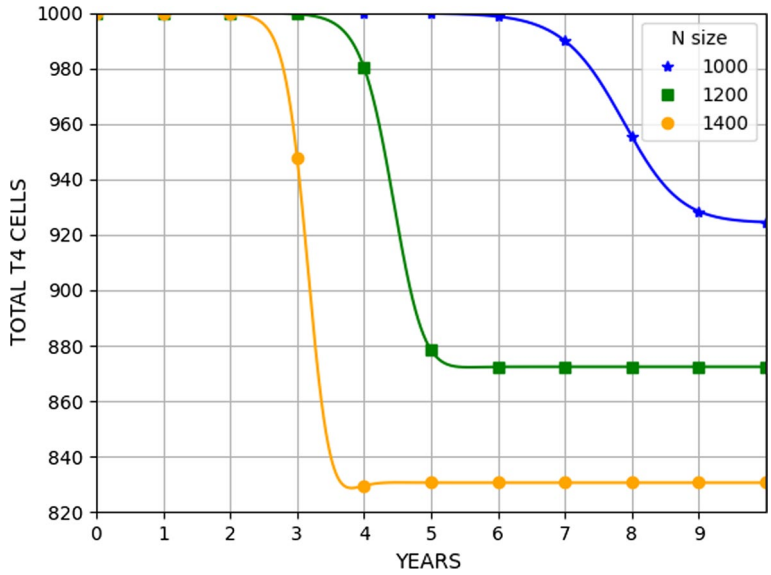


Fig. 5 The total CD4+ T-cell population versus time after the infection. We have reproduced the results from Fig. 4 from the original paper [35] showing the decrease in the overall population of CD4+ T-cell (uninfected + latently infected + actively infected CD4+) over time, depending on the number of infectious particles produced per actively infected cell (N). To reproduce these results run the hiv-t4cell.ipynb Jupyter Notebook from the Additional file 5: Jupyter Notebook hiv-t4cell.ipynb

reconstruction of existing ones. We are confident that `modelbase` will in particular support users with limited modelling experience in re-constructing already published work, serving as a starting point for their further exploration and development. We have previously demonstrated the versatility of `modelbase` by re-implementing mathematical models previously published without the source code: two models of biochemical processes in plants [31, 32], and a model of the non-oxidative pentose phosphate pathway of human erythrocytes [33, 34]. To present how the software can be applied to study medical systems, we used `modelbase` to re-implement various models, not published by our group, and reproduced key results of the original manuscripts. It was beyond our focus to verify the scientific accuracy of the corresponding model assumptions. We selected these examples to show that despite describing different processes, they all share a unified construct. This highlights that by learning how to build a dynamic model with `modelbase`, the user do not learn how to build a one-purpose model, but in fact expands the toolbox to be capable of replicating any given ODE based model. All examples are available as Jupyter notebooks and listed in the Additional file 3: Jupyter Notebook upper-glycolysis.ipynb.

Compartment model for disease evolution

For this paper, we surveyed available computational models and selected a relatively old publication of significant impact, that was published without providing the computational source code, nor details regarding the numerical integration. We chose a four-compartment model of HIV immunology that investigates the interaction of a single virus population with the immune system described only by the $CD4^+$ T cells, commonly known as T helper cells [35]. We implemented the four ODEs describing the

dynamics of uninfected (T), latently infected (L), actively infected $CD4^+$ T cells (A), and infectious HIV population (V). In Fig. 5, we reproduce the results from Fig. 4 in the original paper, whereby changing the number of infectious particles produced per actively infected cell (N) we follow the dynamics of the overall T cell population (T+L+A) over a period of 10 years. The model was also used to explore the effect of azidothymidine, an antiretroviral medication, by decreasing the value of N after 3 years by 25% or 75%, mimicking the blocking of the viral replication of HIV. A more detailed description of the time-dependent drug concentration in the body is often achieved with pharmacokinetic models. Mathematical models based on a system of differential equations that link the dosing regimen with the dynamics of a disease are called pharmacokinetic-pharmacodynamic (PK-PD) models [36]. The next example explores how `modelbase` can be used to develop such models.

PK-PD models and precision medicine

Technological advances forced a paradigm shift in many fields, including medicine, making more personalised healthcare not only a possibility but a necessity. A pivotal role in the success of precision medicine will be to correctly determine dosing regimes for drugs [37]. PK-PD models provide a quantitative tool to support this [38]. PK-PD models have proven successful in many fields, including oncology [39], here we use the classical tumour growth model developed by Simeoni and colleagues, originally implemented using the industry-standard software WinNonlin [40]. As the full pharmacokinetic model is not fully described, we reproduced only the highly simplified case, where we assume a single drug administration and investigate the effect of drug potency (k_2) on simulated tumour growth curves. In Fig. 6 we plot the simulation results of the `modelbase` implementation of the system of four ODEs over a period of 18 days, where we systematically changed the value of k_2, assuming a single drug administration on Day 9. With the MCA suite available in our software, we can calculate the response to perturbation of all other system parameters. Such a quantitative description of the system's dynamics to local parameter perturbation provides support for further studies of the rational design of combined drug therapy and the discovery of new drug targets, as described in the review by Cascante and colleagues [41].

Modelling of infectious diseases with SIR models

Finally, compartmental models based on ODE systems have a long history of application in mathematical epidemiology [42]. Many of them, including numerous recent publications studying the spread of coronavirus, are based on the classic epidemic Susceptible-Infected-Recovered (SIR) model, originating from the theories developed by Kermack and McKendrick at the beginning of the last century [43]. One of the most important insights gained from simulating the dynamics of infectious disease is the existence of disease-free or endemic equilibrium, and assessment of its stability [44]. Interestingly, periodic oscillations have been observed for several infectious diseases, including measles, influenza and smallpox [42]. To provide an overview of more `modelbase` functionalities we have implemented a relatively simple SIR model based on the recently published autonomous model for smallpox [45]. We have generated damped oscillations and visualised them using the built-in function `plot_phase_plane` (Fig. 7). In the

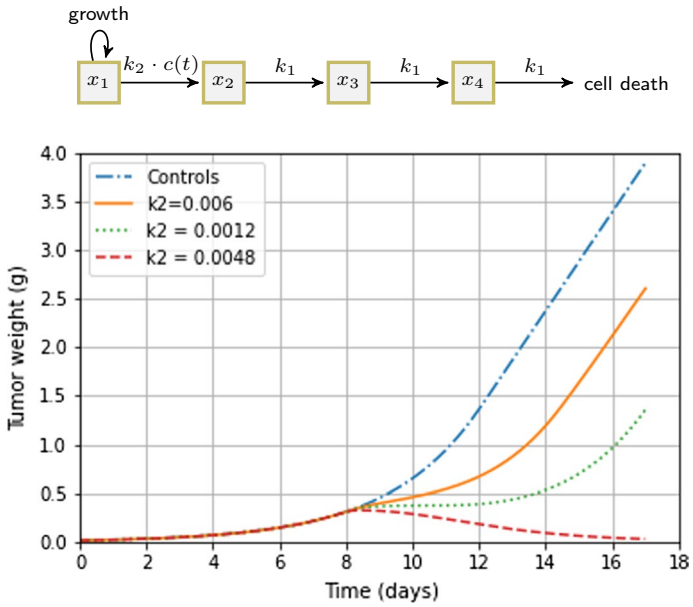


Fig. 6 Compartmental pharmacokinetic-pharmacodynamic model of tumour growth after anticancer therapy. We have reproduced the simplified version of the PK-PD model of tumour growth, where the PK part is reduced to a single input and simulated the effect of drug potency (k_2) on tumour growth curves. The system of four ODEs describing the dynamics of the system visualised on a scheme above is integrated over a period of 18 days. We systematically changed the value of k_2, assuming a single drug administration on Day 9. We have obtained the same results as in Fig. 4 in the original paper [40]. To reproduce these results run the tumour-growth-2004.ipynb Jupyter Notebook from the Additional file 6: Jupyter Notebook tumour-growth-2004.ipynb

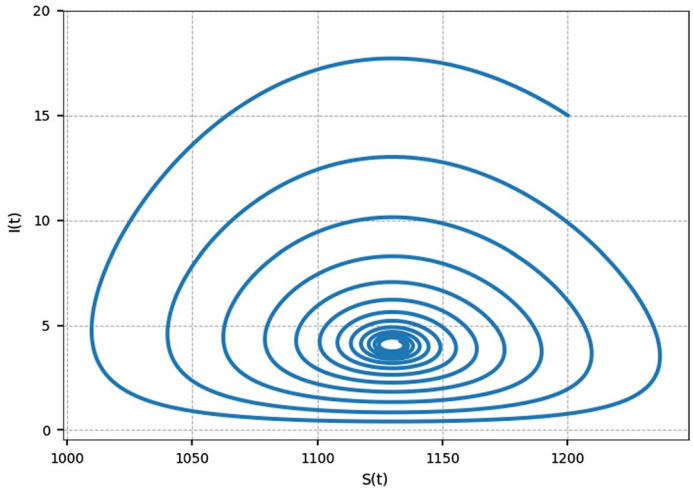


Fig. 7 Sample phase portrait obtained with SIR model with oscillations. SIR model with vital dynamics including birth rate has been adapted based on the autonomous model to simulate periodicity of chickenpox outbreak in Hida, Japan [45]. To reproduce these results run the sir-model.ipynb Jupyter Notebook from the Additional file 7: Jupyter Notebook sir-model.ipynb

accompanying Jupyter notebook we demonstrate using `modelbase`, how simply the SIR model can be built and how to modify it to construct more variants, such as the SEIR (E-exposed) or SIRD (D-deceased) models.

Conclusions

Here, we are presenting an update of our modelling software that simplifies the process of building mathematical models based on ODEs. `modelbase` is fully embedded in the Python programming language. It facilities a systematic construction of new models, and replication of models in a consistent, tractable and expandable manner. As ODEs are a core method to describe the dynamical systems, we hope that our software will serve as the base for deterministic modelling, hence its name. With the smoothed interface and clearer description of how the software can be used for medical purposes, such as simulation of possible drug regimens for precision medicine, we expect to broaden our user community. We envisage that by providing the MCA functionality, users new to mathematical modelling will adopt a working scheme where such sensitivity analyses are an integral part of model development and analysis. The value of sensitivity analyses is demonstrated by considering how the results of such analyses have given rise to new potential targets for drug discovery [41]. We anticipate that the capability of `modelbase` to automatically generate isotopic label-specific models will prove useful in predicting fluxes and label propagation dynamics through various metabolic networks. In emerging fields such as computational oncology, such models will be useful to, e.g., predict the appearance of labels in cancer cells.

Availability and requirements

Project name: `modelbase`
Project home page: https://pypi.org/project/modelbase/
Code repository: https://gitlab.com/qtb-hhu/modelbase-software
Version published: 1.2.3
Date published: 14 Jan 2021
Documentation: https://modelbase.readthedocs.io/en/latest
Operating system(s): Platform independent
Programming language: Python
Other requirements: None
Licence: GNU General Public License (GPL), version 3
Any restrictions to use by non-academics: None.

Abbreviations
MCA: Metabolic control analysis; ODE: Ordinary differential equations; PK-PD: Pharmacokinetic-pharmacodynamic; SBML: Systems biology markup language.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-021-04122-7.

Additional file 1. Table with the list of available models in the repository

Additional file 2. README.md with the Installation instructions

Additional file 3. Jupyter Notebook upper-glycolysis.ipynb: toy model of the upper path of glycolysis asintroduced in [27]

Additional file 4. Jupyter Notebook label-propagation-2015.ipynb: an example of a linear non-reversible pathwayof five randomly sized metabolites and label propagation experiments, as proposed in the paper by Sokol and Portais [29]

Additional file 5. Jupyter Notebook hiv-t4cell.ipynb: a model of the dynamics of HIV infection of CD4+ cells,considering three populations of T cells and free virus, proposed by Perelson, Kirschner and de Boer [35]

Additional file 6. Jupyter Notebook tumour-growth-2004.ipynb: minimal pharmacokinetic-pharmacodynamic(PK-PD)model linking that linking the dosing regimen of an anticancer agent to the tumour growth, proposed bySimeoni and colleagues [40]

Additional file 7. Jupyter Notebook sir-model.ipynb: classic epidemic Susceptible-Infected-Recovered (SIR) model-parameterised as the autonomous model used to simulate periodicity of chickenpox outbreak in Hida, Japan [45]

Acknowledgements
We would like to thank all members of the Institute of Quantitative and Theoretical Biology for their continuous involvement in the software testing and expansion of the repository of novel and replicated computational models.

Authors' contributions
MvA implemented, tested, bug fixed and maintained the `modelbase` software. OE conceptualised `modelbase`. AM specified requirements, designed and tested the implemented changes and provided all biomedical examples used in this paper. All authors have read and approved the manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germanys Excellence Strategy—EXC-2048/1–project ID 390686111 (OE, AM), EU's Horizon 2020 research and innovation programme under the Grant Agreement 862087 (MvA), and Deutsche Forschungsgemeinschaft Research Grant-project ID 420069095 (AM). The funding bodies did not play any role in the design of the study nor collection, analysis, and interpretation of data nor in writing the manuscript.

Availability of data and materials
All data generated or analysed during this study are included in this published article. Additionally, code documentation can be found here https://modelbase.readthedocs.io/en/latest and more Jupyter Notebooks with teaching materials and models here https://gitlab.com/qtb-hhu/models. If you have any questions regarding `modelbase`, you are very welcome to ask them on our issues page or by contacting any of the authors. It is our mission to enable reproducible science and to help to put the theory into action.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests

Author details
[1] Institute of Quantitative and Theoretical Biology, Heinrich Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany. [2] CEPLAS - Cluster of Excellence on Plant Sciences, Universitätsstr. 1, 40225 Düsseldorf, Germany.

Received: 9 September 2020 Accepted: 7 April 2021
Published online: 20 April 2021

References

1. Kitano H. Computational systems biology. Nature. 2002;420(6912):206–10. https://doi.org/10.1038/nature01254.
2. Barbolosi D, Ciccolini J, Lacarelle B, Barlési F, André N. Computational oncology-mathematical modelling of drug regimens for precision medicine. Nat Rev Clin Oncol. 2016. https://doi.org/10.1038/nrclinonc.2015.204.
3. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science. 2020;368(6492):742–6. https://doi.org/10.1126/science.abb4557.
4. Tang B, Bragazzi NL, Li Q, Tang S, Xiao Y, Wu J. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). Infect Dis Modell. 2020;5:248–55. https://doi.org/10.1016/j.idm.2020.02.001.
5. Rocklöv J, Sjödin H, Wilder-Smith A. COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures. J Travel Med. 2020;27(3):taa030.
6. Brodland GW. How computational models can help unlock biological systems. Semin Cell Dev Biol. 2015. https://doi.org/10.1016/j.semcdb.2015.07.001.
7. Butler JA, Cosgrove J, Alden K, Timmis J, Coles MC. Model-driven experimentation: a new approach to understand mechanisms of tertiary lymphoid tissue formation, function, and therapeutic resolution. Front Immunol. 2017. https://doi.org/10.3389/fimmu.2016.00658.
8. Rios-Estepa R, Lange I, Lee JM, Lange BM. Mathematical modeling-guided evaluation of biochemical, developmental, environmental, and genotypic determinants of essential oil composition and yield in peppermint leaves. Plant Physiol. 2010;152:2105–19. https://doi.org/10.1104/pp.109.152256.
9. Acs S, Ostlaender N, Listorti G, Hradec J, Hardy M, Smits P, Hordijk L. Modelling for EU Policy support: impact assessments. Technical report. Publications Office of the European Union, Luxembourg (2019). https://doi.org/10.2760/748720
10. Keating S, Waltemath D, König M, Zhang F, Dräger A, Chaouiya C, Bergmann F, Finney A, Gillespie C, Helikar T, Hoops S, Malik-Sheriff R, Moodie S, Moraru I, Myers C, Naldi A, Olivier B, Sahle S, Schaff J, Smith L, Swat M, Thieffry D, Watanabe L, Wilkinson D, Blinov M, Begley K, Faeder J, Gómez H, Hamm T, Inagaki Y, Liebermeister W, Lister A, Lucio D, Mjolsness E, Proctor C, Raman K, Rodriguez N, Shaffer C, Shapiro B, Stelling J, Swainston N, Tanimura N, Wagner J, Meier-Schellersheim M, Sauro H, Palsson B, Bolouri H, Kitano H, Funahashi A, Hermjakob H, Doyle J. SBML level 3 Community members: SBML level 3: an extensible format for the exchange and reuse of biological models. Mol Syst Biol. 2020. https://doi.org/10.15252/msb.20199110.
11. Malik-Sheriff RS, Glont M, Nguyen TVN, Tiwari K, Roberts MG, Xavier A, Vu MT, Men J, Maire M, Kananathan S, Fairbanks EL, Meyer JP, Arankalle C, Varusai TM, Knight-Schrijver V, Li L, Dueñas-Roca C, Dass G, Keating SM, Park YM, Buso N, Rodriguez N, Hucka M, Hermjakob H. BioModels-15 years of sharing computational models in life science. Nucleic Acids Res. 2020;48(D1):407–15. https://doi.org/10.1093/nar/gkz1055.
12. Ebenhöh O, Aalst MV, Saadat NP, Nies T, Matuszyńska A. Building mathematical models of biological systems with modelbase. J Open Res Softw. 2018. https://doi.org/10.5334/jors.236.
13. Poolman MG, Scrum PY. Metabolic modelling with python. IEE Proc Syst Biol. 2006;153(5):375–8. https://doi.org/10.1049/ip-syb.
14. Olivier BG, Rohwer JM, Hofmeyr JHS. Modelling cellular systems with PySCeS. Bioinformatics. 2005;21(4):560–1. https://doi.org/10.1093/bioinformatics/bti046.
15. Lopez CF, Muhlich JL, Bachman JA, Sorger PK. Programming biological models in python using PySB. Mol Syst Biol. 2013;9(1):646. https://doi.org/10.1038/msb.2013.1.
16. Choi K, Medley JK, König M, Stocking K, Smith L, Gu S, Sauro HM. Tellurium: an extensible python-based modeling environment for systems and synthetic biology. Biosystems. 2018;171:74–9. https://doi.org/10.1016/j.biosystems.2018.07.006.
17. Medley JK, Choi K, König M, Smith L, Gu S, Hellerstein J, Sealfon SC, Sauro HM. Tellurium notebooks—an environment for reproducible dynamical modeling in systems biology. PLoS Comput Biol. 2018;14(6):1006220. https://doi.org/10.1371/journal.pcbi.1006220.
18. Schölzel C, Blesius V, Ernst G, Dominik A. Characteristics of mathematical modeling languages that facilitate model reuse in systems biology: a software engineering perspective. BioRxiv. 2020. https://doi.org/10.1101/2019.12.16.875260.
19. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, Rio JF, Wiebe M, Peterson P, Gerard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. Array programming with NumPy. Nature. 2020;585(7825):357–62. https://doi.org/10.1038/s41586-020-2649-2.
20. The pandas development team: pandas-dev/pandas: Pandas (2020). https://doi.org/10.5281/zenodo.3509134.
21. Andersson C, Claus F, Akesson J. ScienceDirect assimulo: a unified framework for ODE solvers. Math Comput Simul. 2015;116:26–43. https://doi.org/10.1016/j.matcom.2015.04.007.
22. Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, Shumaker DANE, Woodward CS. Sundials?: Suite of nonlinear and differential/algebraic equation solvers. ACM Trans Math Softw. 2005;31(3):363–96.
23. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Jarrod Millman K, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey C, Polat I, Feng Y, Moore EW, Van der Plas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P. SciPy 1.0 contributors: SciPy 1.0: fundamental algorithms for scientific computing in python. Nat Methods. 2020;17:261–72. https://doi.org/10.1038/s41592-019-0686-2.
24. Malengier B, Kišon P, Tocknell J, Abert C, Bruckner F, Bisotti MA. ODES: a high level interface to ODE and DAE solvers. J Open Source Softw. 2018;3(22):165. https://doi.org/10.21105/joss.00165.
25. Kacser H, Burns JAAA. The control of flux: 21. Symp Soc Exp Biol. 1973;27:65–104.
26. Heinrich R, Rapoport TA. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. Eur J Biochem. 1974;42(1):89–95. https://doi.org/10.1111/j.1432-1033.1974.tb03318.x.
27. Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, Herwig R. Systems biology: a textbook. Weinheim: Wiley; 2013.

28. Hunter JD. Matplotlib: a 2d graphics environment. Comput Sci Eng. 2007;9(3):90–5. https://doi.org/10.1109/MCSE.2007.55.
29. Sokol S, Portais J-C. Theoretical basis for dynamic label propagation in stationary metabolic networks under step and periodic inputs. PLoS ONE. 2015;10(12):0144652. https://doi.org/10.1371/journal.pone.0144652.
30. Le Novère N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, Crampin EJ, Halstead M, Klipp E, Mendes P, Nielsen P, Sauro H, Shapiro B, Snoep JL, Spence HD, Wanner BL. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat Publ Gr. 2005. https://doi.org/10.1038/nbt1156.
31. Ebenhöh O, Fucile G, Finazzi G, Rochaix JD, Goldschmidt-Clermont M. Short-term acclimation of the photosynthetic electron transfer chain to changing light: a mathematical model. Philos Trans R Soc Lond B Biol Sci. 2014;369(1640):20130223. https://doi.org/10.1098/rstb.2013.0223.
32. Pettersson G, Ryde-Pettersson U. A mathematical model of the Calvin photosynthesis cycle. Eur J Biochem. 1988;175(3):661–72. https://doi.org/10.1111/j.1432-1033.1988.tb14242.x.
33. McIntyre LM, Thorburn DR, Bubb WA, Kuchel PW. Comparison of computer simulations of the F-type and L-type non-oxidative hexose monophosphate shunts with 31P-NMR experimental data from human erythrocytes. Eur J Biochem. 1989;180(2):399–420.
34. Berthon HA, Bubb WA, Kuchel PW. 13C n.m.r. isotopomer and computer-simulation studies of the non-oxidative pentose phosphate pathway of human erythrocytes. Biochem J. 1993;296:379–87. https://doi.org/10.1042/bj2960379.
35. Perelson AS, Kirschner DE, Boer RD. Dynamics of hiv infection of cd4+ t cells. Math Biosci. 1993;114(1):81–125. https://doi.org/10.1016/0025-5564(93)90043-A.
36. Derendorf H, Meibohm B. Modeling of pharmacokinetic/pharmacodynamic (PK/PD) relationships: concepts and perspectives. Pharm Res. 1999;16(2):176–85. https://doi.org/10.1023/A:1011907920641.
37. Lloyd KCK, Khanna C, Hendricks W, Trent J, Kotlikoff M. Precision medicine: an opportunity for a paradigm shift in veterinary medicine HHS public access. J Am Vet Med Assoc. 2016;248(1):45–8. https://doi.org/10.2460/javma.248.1.45.
38. Polasek TM, Shakib S, Rostami-Hodjegan A. Precision dosing in clinical medicine: present and future. Miton Park: Taylor and Francis Ltd.; 2018. https://doi.org/10.1080/17512433.2018.1501271.
39. Koziol JA, Falls TJ, Schnitzer JE. Different ODE models of tumor growth can deliver similar results. BMC Cancer. 2020;20(1):226. https://doi.org/10.1186/s12885-020-6703-0.
40. Simeoni M, Magni P, Cammia C, De Nicolao G, Croci V, Pesenti E, Germani M, Poggesi I, Rocchetti M. Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. Cancer Res. 2004;64(3):1094–101. https://doi.org/10.1158/0008-5472.CAN-03-2524.
41. Cascante M, Boros LG, Comin-Anduix B, de Atauri P, Centelles JJ, Lee P. Metabolic control analysis in drug discovery and disease. Nat Biotechnol. 2002;20:243–9. https://doi.org/10.1038/nbt0302-243.
42. Hethcote HW. The mathematics of infectious diseases. SIAM Rev. 2000;42(4):599–653. https://doi.org/10.1137/S0036144500371907.
43. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond Ser A Contain Pap Math Phys Charact. 1927;115(772):700–21. https://doi.org/10.1098/rspa.1927.0118.
44. Wang J, Liu S, Zheng B, Takeuchi Y. Qualitative and bifurcation analysis using an SIR model with a saturated treatment function. Math Comput Modell. 2012;55(3–4):710–22. https://doi.org/10.1016/j.mcm.2011.08.045.
45. Greer M, Saha R, Gogliettino A, Yu C, Zollo-Venecek K. Emergence of oscillations in a simple epidemic model with demographic data. R Soc Open Sci. 2020. https://doi.org/10.1098/rsos.191187.

Publisher's Note

2.3 Network Reconstruction and Modelling Made Reproducible with moped

The `moped` Python package consists of tools to automatically reconstruct genome-scale constraint-based models from combining different data sources. While easily extendable to other databases, it is primarily used with the MetaCyc database [90]. The focus of this package is the reconstruction of the metabolic models, while tools such as `cobrapy` are interfaced to facility analysis [91]. The package has been developed in close cooperation with Nima P. Saadat and we share the first authorship for it.

Publication	Network Reconstruction and Modelling Made Reproducible with moped
Author position	Shared first
Status	peer-reviewed and published
doi	10.3390/metabo12040275

metabolites

MDPI

Article

Network Reconstruction and Modelling Made Reproducible with moped

Nima P. Saadat [1,*,†], Marvin van Aalst [1,†] and Oliver Ebenhöh [1,2]

[1] Institute of Quantitative and Theoretical Biology, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany; marvin.van.aalst@hhu.de (M.v.A.); oliver.ebenhoeh@hhu.de (O.E.)
[2] Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany
* Correspondence: nima.saadat@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.; Ebenhöh, 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.

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 SMBL 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 mopeds 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.

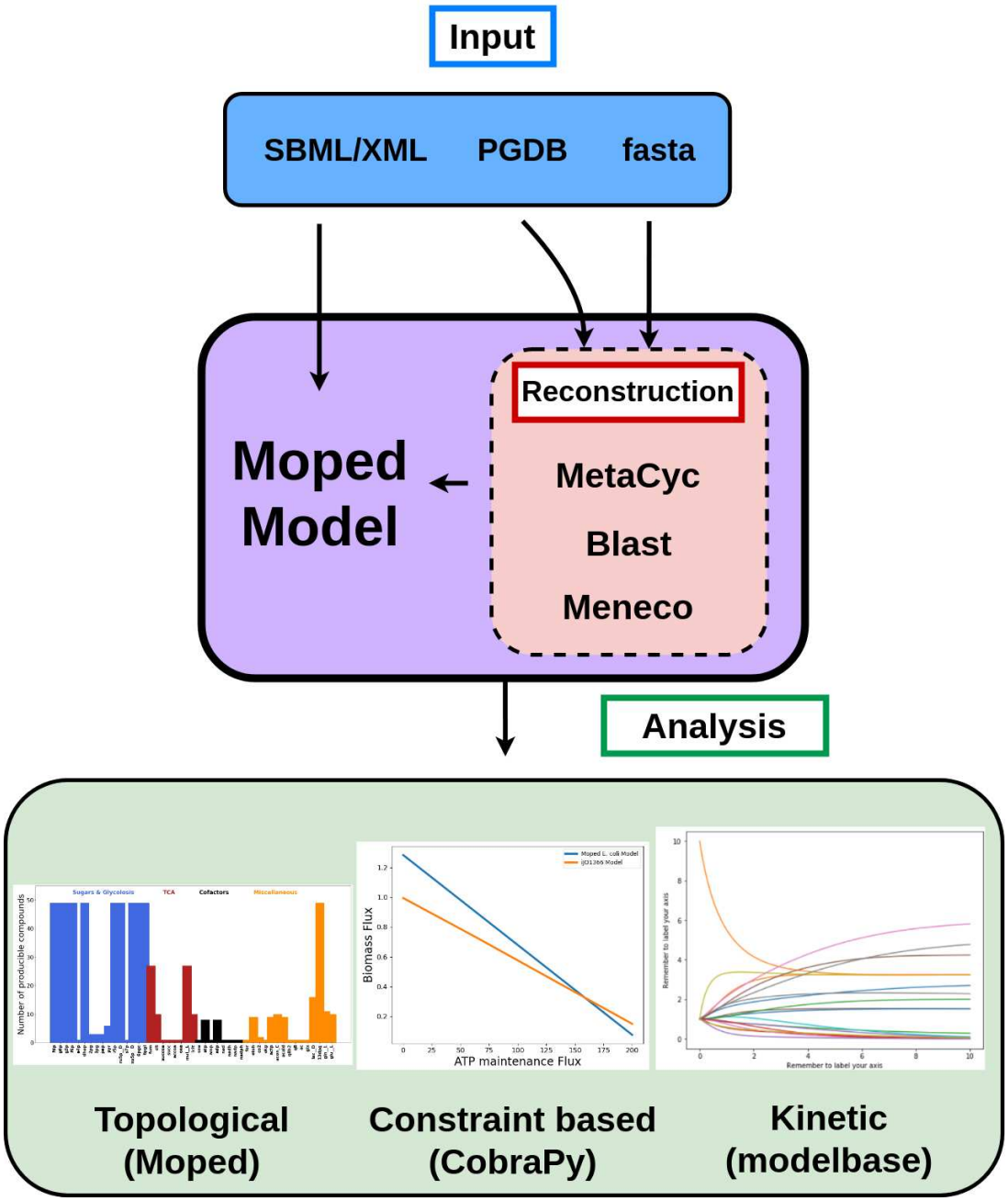


Figure 1. The modeling hub moped. moped accepts SBML, FASTA files or MetaCyc and BioCyc PGDBs as inputs. PGDBs and SBML files are directly converted into a moped object. By BLASTing genome/proteome-sequences against MetaCyc, moped models can be constructed utilizing GPR rules. Further reconstruction 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.

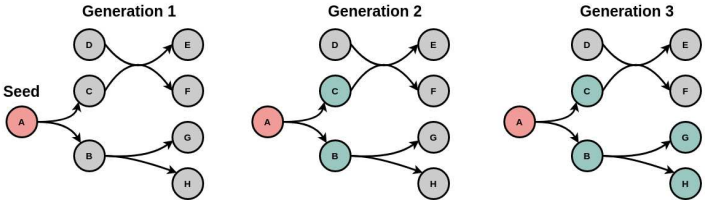


Figure 2. Metabolic network expansion. Beginning with an initial set of compounds, the seed (here panel 1), the expansion process detects all producible compounds in a network and adds them to the seed for the next generation until no additional producible compounds are found.

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.

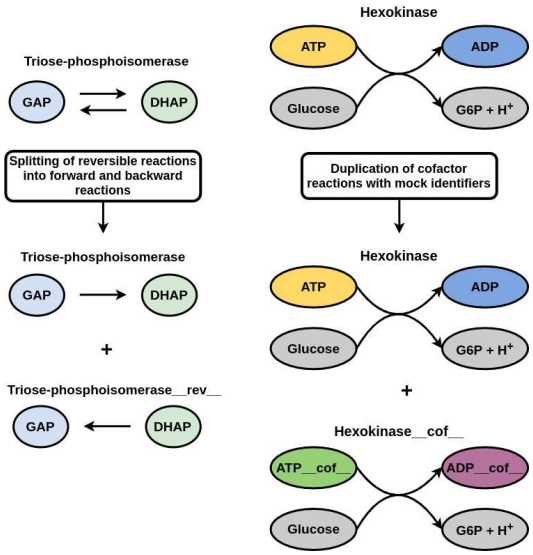


Figure 3. Topological network modifications moped offers functions for splitting reversible reactions into forward and backward reactions in a network. Adding a copy of each cofactor dependent reaction and replacing cofactors (here ATP and ADP) with mock identifiers allows unblocking cofactor dependent 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 conversation. 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.

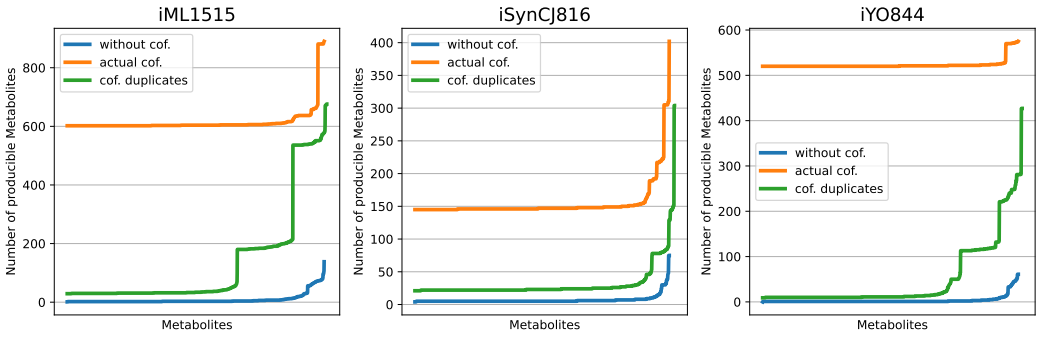


Figure 4. Metabolic scopes in established models of *E. coli* (iML1515), *B. subtilis* (iYO844) and *Synechocystis* sp. PCC 6803 (iSynCJ816). The differently colored graphs represent the same analysis but including no cofactors, actual cofactors and cofactor duplicates in the seed.

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.

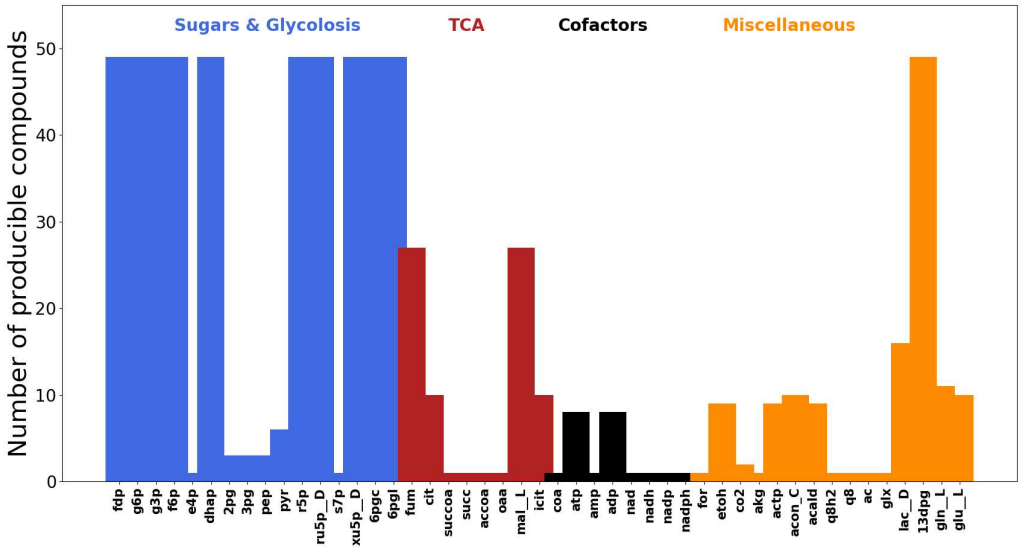


Figure 5. Metabolic scopes of all compounds in the *E. coli* core metabolic model calculated using moped. The *Y*-axis indicates the total amount of compounds producible from every compound, water and a set of acceptor mock cofactors.

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.

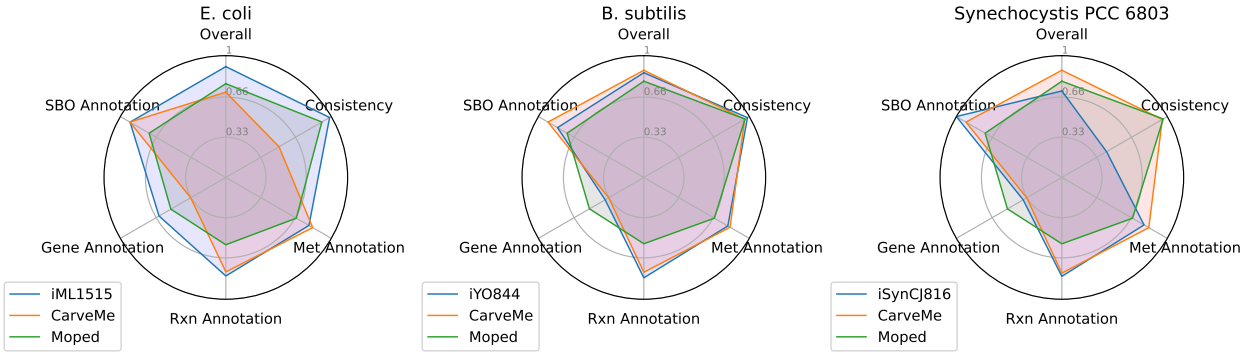


Figure 6. MEMOTE evaluations for draft model reconstructions produced by CarveMe and moped, as well as established models, for *E. coli*, *Bacillus subtilis* and *Synechocystis* sp. PCC 6803. MEMOTE evaluations include the stoichiometric consistency and the annotation level of models.

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 shows 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.

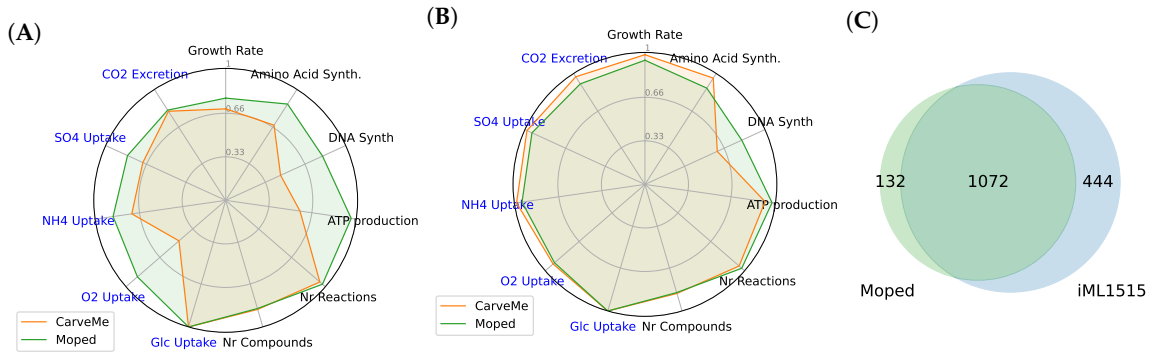


Figure 7. Functional comparison of the draft model reconstructions using moped and CarveMe with iML1515. We calculated maximal growth rates, respective ATP production rates and exchange fluxes for compounds in the medium, as well as optimal production rates for amino acids and nucleic acids for completely 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

1. 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]
2. Zomorrodi, 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]
3. 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]
4. Heinrich, R.; Schuster, S. *The Regulation of Cellular Systems*; Chapman and Hall: New York, NY, USA, 1996.
5. 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]
6. 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]
7. 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]
8. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef]
9. Raman, K.; Chandra, N. Flux balance analysis of biological systems: Applications and challenges. *Brief. Bioinform.* **2009**, *10*, 435–449. [CrossRef]
10. Orth, J.D.; Thiele, I.; Palsson, B.Ø. What is flux balance analysis? *Nat. Biotechnol.* **2010**, *28*, 245–248. [CrossRef]
11. Wunderlich, Z.; Mirny, L.A. Using the topology of metabolic networks to predict viability of mutant strains. *Biophys. J.* **2006**, *91*, 2304–2311. [CrossRef]
12. Handorf, T.; Ebenhöh, O.; Heinrich, R. Expanding metabolic networks: Scopes of compounds, robustness, and evolution. *J. Mol. Evol.* **2005**, *61*, 498–512. [CrossRef] [PubMed]
13. Romero, P.; Karp, P. Nutrient-related analysis of pathway/genome databases. In *Biocomputing 2001*; World Scientific: Singapore, 2000; pp. 471–482.
14. Ebenhöh, O.; Handorf, T. Functional classification of genome-scale metabolic networks. *EURASIP J. Bioinform. Syst. Biol.* **2009**, *2009*, 570456. [CrossRef] [PubMed]
15. Christian, N.; May, P.; Kempa, S.; Handorf, T.; Ebenhöh, O. An integrative approach towards completing genome-scale metabolic networks. *Mol. Biosyst.* **2009**, *5*, 1889–1903. [CrossRef]
16. Orth, J.D.; Palsson, B.Ø. Systematizing the generation of missing metabolic knowledge. *Biotechnol. Bioeng.* **2010**, *107*, 403–412. [CrossRef] [PubMed]
17. 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.; Ebenhöh, 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.; Herrgård, 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.; Zyoud, 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. i ML1515, 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]

2.4 dismo: building and analysing discrete spatial ODE models

The Python package dismo is a toolkit for building and analysing discrete spatial models based on ordinary differential equations. I both designed and developed the package and wrote the manuscript together with Oliver Ebenhöeh.

Publication	dismo: building and analysing discrete spatial ODE models
Author position	First
Status	pre-print
doi	10.1101/2023.10.17.562679

dismo: building and analysing discrete spatial ODE models

Marvin van Aalst [1] **and Oliver Ebenhöh** [1,2]

[1] Institute of Theoretical and Quantitative Biology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany; [2] Cluster of Excellence on Plant Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Abstract

Summary: `dismo` is a Python package for building and analysing discrete spatial models based on ordinary differential equations. Its primary purpose is to allow arbitrarily complex internal and transport processes to easily be mapped over multiple different regular grids. For this it features one, two and three-dimensional layouts, with standard and non-standard (e.g. hexagonal or triangular) grids, which can be arranged in intricate, non-regular shapes.

Availability and implementation: https://gitlab.com/qtb-hhu/dismo

Documentation: https://qtb-hhu.gitlab.io/dismo/

Contact: oliver.ebenhoeh@hhu.de

mathematical modelling | discrete | spatial

Introduction

Complex multicellular organisms depend on communication between cells. This includes exchange of information, for example by hormone signalling, and the transport of metabolites between cells and organs. Many mathematical models of biological processes, including metabolism, signalling, and ecosystem dynamics, ignore this spatial dimension and describe the system dynamics by coupled ordinary differential equations. The classical approach to model spatial dynamics is by using partial differential equations (PDEs). Famous examples include reaction-diffusion systems, which are able to describe diverse phenomena, such as the formation of patterns or the dispersion of animals in an ecosystem (Landge et al., 2020, Soh et al., 2010). There are numerous well-established tools for simulating such problems, for an excellent review see (Van Liedekerke et al., 2015). However, the numerical treatment of PDEs poses various challenges. Simulation requires a careful definition of the boundary conditions, and the spatial discretisation must fulfil certain criteria in order to guarantee convergence. Moreover, they are computationally very expensive, especially when the number of dynamic variables becomes large. However, when describing multicellular systems, each individual cell can be approximated to be well-mixed. If this assumption is fulfilled, the dynamics can be described as the interaction of discrete entities, which may correspond to the single cells, and can thus be modelled using a system of ordinary differential equations (ODEs), in which the intracellular processes and the exchange processes between cells form the basis of the coupled equations, similar to agent-based modeling (Helbing, 2012). In this way, the spatial discretisation naturally reflects the biological system under investigation, and, especially when different cell types are involved, appears more straight-forward than in a pure PDE-based approach.

There are excellent tools for solving PDEs in Python, such as `py-pde` (Zwicker, 2020), and highly sophisticated packages designed for biological cell simulation such as `biocellion` (Kang et al., 2014), `CellSys` (Abbasi et al., 2022), `Chaste` (Mirams et al., 2013), `MecaGen` (Delile et al., 2013), `PhysiCell` (Ghaffarizadeh et al., 2018) and tools based on `Morpheus` (Starruß et al., 2014, Alamoudi et al., 2023). But to the best of our knowledge there exist no tools to easily create and analyse discrete spatial ODE models aimed at beginner modellers. Here, we present `dismo`, a Python package for implementing and solving discrete spatial systems of ODEs. Our package provides methods to automatically create the complete set of coupled ODEs, based on a user-defined grid, the ODEs describing the processes within, and transport between the individual cells. All of these processes can be described using arbitrary Python functions, enabling a very flexible interface. We offer numerous regular pre-defined one-, two- and three-dimensional grids, which can be arranged in intricate, non-regular shapes. The package further supports the definition of different cell types, which can each have their specific intracellular dynamics. In this application note, we describe the functionality of `dismo` and demonstrate its capabilities for a spatially resolved model of sugar transport in a photosynthetic leaf.

Implementation and results

We build `dismo` using best object-oriented programming (OOP) practices, utilizing composition, separation of concerns and abstract base classes (Gamma et al., 1994). For example, coordinates and their respective behaviour are separated from the general grid implementation, such that it is easy to implement new grids. Furthermore, we utilized composition instead of inheritance to supply different grids for the model

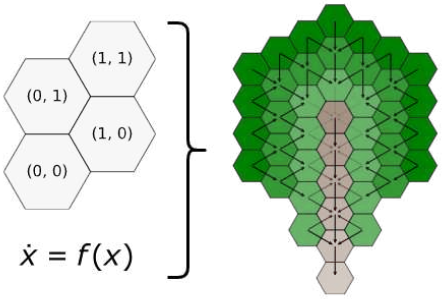


Figure 1. Schematic representation of a dismo model. The left-hand-side depicts that every model consists of a grid, which can be one-, two- or three-dimensional, and one or multiple ordinary differential equations for one or multiple state variables, that either describe the dynamics within a given cell, or transport processes between cells in that grid. The right-hand-side depicts the flux of sugars in a plant leaf model.

instances. This way it is easy to interchange parts and expand both the model and grid types. Construction and analysis of models, as well as how to subtype a model class are all described in detail in accompanying jupyter notebooks (Granger and Pérez, 2021).

Common analysis methods like time course simulation and visualization are implemented as ready-to-use functions and utilize well-known packages from the scientific Python landscape, like `NumPy`, `pandas`, `Matplotlib` as well as the `assimulo` wrapper around the `sundials` solver suite (Andersson et al., Hunter, Harris et al., McKinney, Gardner et al., 2022).

As modelling sugar transport inside of plant leaves was the motivation behind building `dismo`, the package also includes different types of plant leaf models, which build up in complexity. First we supply mesophyll models, which consist of a single variable for sucrose as well as a single mesophyll cell type and incorporate both a saturating photosynthesis function and a passive diffusion process between mesophyll cells. The mesophyll model is then expanded by supplying an additional vein cell type, which transports sucrose more rapidly and provides an outflux out the leaf. For this, passive transport processes between vein cells as well as an active (as in one-sided) transport of sucrose from the mesophyll cells into the vein cells are added to the model description. Finally, the stomata models further extend the model by both a new stomatal cell-type and a second variable for CO2. The idea here is that CO2 only enters stomatal cells, which can then transport this CO2 into mesophyll cells. Vein cells are assumed not to contain any CO2. In this model the mesophyll photosynthesis function is now dependent on the CO2 concentration in that particular cell and thus more dynamic and subject to the placement of stomatal cells in the leaf.

Conclusion

`dismo` is a new Python package developed for easily building and analysing multi-variate, discrete spatial models based on ODEs aimed at beginner modellers. It thus differs from highly sophisticated packages focused on continuous PDEs, enabling spatial discretisation that naturally reflects the biological system under investigation, making model construction and analysis more straight-forward than in a PDE-based approach.

The flexible nature of `dismo` allows arbitrary code to be run per cell, such that it can be extended to even run other types of models. As an example, it would be possible to create spatial community models, where each cell represents one or more organisms using a constrained-based model and then metabolite exchanges between the communities are modelled using diffusion processes (Orth et al., 2010).

Funding

This work was 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.).

References

Amit N. Landge, Benjamin M. Jordan, Xavier Diego, and Patrick Müller. Pattern formation mechanisms of self-organizing reaction-diffusion systems. *Developmental Biology*, 460(1):2–11, 2020. ISSN 0012-1606. doi: https://doi.org/10.1016/j.ydbio.2019.10.031. URL https://www.sciencedirect.com/science/article/pii/S001216061930377X. Systems Biology of Pattern Formation.

Siowling Soh, Marta Byrska, Kristiana Kandere-Grzybowska, and BartoszA. Grzybowski. Reaction-diffusion systems in intracellular molecular transport and control. *Angewandte Chemie International Edition*, 49(25):4170–4198, 2010. doi: https://doi.org/10.1002/anie.200905513. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.200905513.

Paul Van Liedekerke, MM Palm, N Jagiella, and Dirk Drasdo. Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results. *Computational particle mechanics*, 2:401–444, 2015.

Dirk Helbing. Agent-based modeling. In *Social self-organization: Agent-based simulations and experiments to study emergent social behavior*, pages 25–70. Springer, 2012.

David Zwicker. py-pde: A python package for solving partial differential equations. *Journal of Open Source Software*, 5(48):2158, 2020. doi: 10.21105/joss.02158. URL https://doi.org/10.21105/joss.02158.

Seunghwa Kang, Simon Kahan, Jason McDermott, Nicholas Flann, and Ilya Shmulevich. Biocellion: accelerating computer simulation of multicellular biological system models. *Bioinformatics*, 30(21):3101–3108, 2014.

Ali Abbasi, Sepideh Amjad-Iranagh, and Bahram Dabir. Cellsys: An open-source tool for building initial structures for bio-membranes and drug-delivery systems. *Journal of Computational Chemistry*, 43(5):331–339, 2022.

Gary R Mirams, Christopher J Arthurs, Miguel O Bernabeu, Rafel Bordas, Jonathan Cooper, Alberto Corrias, Yohan Davit, Sara-Jane Dunn, Alexander G Fletcher, Daniel G Harvey, et al. Chaste: an open source c++ library for computational physiology and biology. *PLoS computational biology*, 9(3):e1002970, 2013.

Julien Delile, René Doursat, and Nadine Peyriéras. Computational modeling and simulation of animal early embryogenesis with the mecagen platform. *Computational systems biology*, 2:359–405, 2013.

Ahmadreza Ghaffarizadeh, Randy Heiland, Samuel H Friedman, Shannon M Mumenthaler, and Paul Macklin. Physicell: An open source physics-based cell simulator for 3-d multicellular systems. *PLoS computational biology*, 14(2):e1005991, 2018.

Jörn Starruß, Walter De Back, Lutz Brusch, and Andreas Deutsch. Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology. *Bioinformatics*, 30(9):1331–1332, 2014.

Emad Alamoudi, Yannik Schälte, Robert Müller, Jörn Starruß, Nils Bundgaard, Frederik Graw, Lutz Brusch, and Jan Hasenauer. FitMultiCell: simulating and parameterizing computational models of multi-scale and multi-cellular processes. *Bioinformatics*, 39(11):

btad674, 11 2023. ISSN 1367-4811. doi: 10.1093/bioinformatics/btad674. URL https://doi.org/10.1093/bioinformatics/btad674.

Erich Gamma, Richard Helm, Ralph Johnson, and John M. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional, 1 edition, 1994. ISBN 0201633612. URL http://www.amazon.com/Design-Patterns-Elements-Reusable-Object-Oriented/dp/0201633612/ref=ntt_at_ep_dpi_1.

Brian E. Granger and Fernando Pérez. Jupyter: Thinking and storytelling with code and data. *Computing in Science Engineering*, 23(2):7–14, 2021. doi: 10.1109/MCSE.2021.3059263.

Christian Andersson, Claus Führer, and Johan Åkesson. Assimulo: A unified framework for ode solvers. 116:26–43. ISSN 0378-4754. doi: https://doi.org/10.1016/j.matcom.2015.04.007. URL https://www.sciencedirect.com/science/article/pii/S0378475415000701.

John D Hunter. Matplotlib: A 2d graphics environment. 9(3):90.

Charles R Harris, K Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E Oliphant. Array programming with NumPy. 585(7825):357–362. ISSN 1476-4687. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51–56.

David J Gardner, Daniel R Reynolds, Carol S Woodward, and Cody J Balos. Enabling new flexibility in the SUNDIALS suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software (TOMS)*, 2022. doi: 10.1145/3539801.

Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245–248, 2010.

2.5 COBREXA.jl: constraint-based reconstruction and exascale analysis

The COBREXA.jl Julia package consists of tools for the analysis of constraint-based models exceeding the genome-scale. Contrasting moped (section 2.3), the focus is less on the reconstruction and more on the analysis. While the majority of the work was done by Miroslav Kratochvil, I contributed functionality regarding loopless FBA and gene knockouts and am the seventh author of the publication.

Publication	COBREXA.jl: constraint-based reconstruction and exascale analysis
Author position	Seventh
Status	peer-reviewed and published
doi	10.1093/bioinformatics/btab782

Bioinformatics, 38(4), 2022, 1171–1172
https://doi.org/10.1093/bioinformatics/btab782
Advance Access Publication Date: 16 November 2021
Applications Note

Systems biology

COBREXA.jl: constraint-based reconstruction and exascale analysis

Miroslav Kratochvíl [1,†], Laurent Heirendt [1,2,†], St. Elmo Wilken[3], Taneli Pusa[1,4], Sylvain Arreckx[1], Alberto Noronha[4], Marvin van Aalst[3], Venkata P. Satagopam [1,2], Oliver Ebenhöh [3], Reinhard Schneider [1,2], Christophe Trefois [1,2,*] and Wei Gu [1,2,*]

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Campus Belval, L-4367 Belvaux, Luxembourg, [2]ELIXIR Luxembourg, Campus Belval, L-4367 Belvaux, Luxembourg, [3]Institute of Quantitative and Theoretical Biology, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany and [4]Nium S.à.r.l., L-4362 Esch-sur-Alzette, Luxembourg

*To whom correspondence should be addressed.
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Associate Editor: Alfonso Valencia

Received on June 3, 2021; revised on October 8, 2021; editorial decision on November 9, 2021; accepted on November 10, 2021

Abstract

Summary: COBREXA.jl is a Julia package for scalable, high-performance constraint-based reconstruction and analysis of very large-scale biological models. Its primary purpose is to facilitate the integration of modern high performance computing environments with the processing and analysis of large-scale metabolic models of challenging complexity. We report the architecture of the package, and demonstrate how the design promotes analysis scalability on several use-cases with multi-organism community models.
Availability and implementation: https://doi.org/10.17881/ZKCR-BT30.
Contact: christophe.trefois@uni.lu or wei.gu@uni.lu
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Understanding metabolic interactions in cells is a crucial step to investigate disease mechanisms and to discover new therapeutics (Apaolaza *et al.*, 2018; Brunk *et al.*, 2018; Cook and Nielsen, 2017). Constraint-Based Reconstruction and Analysis (COBRA) is a promising methodology for analyzing various metabolic processes at the organism- and community- levels (Fang *et al.*, 2020). The main idea behind COBRA is to represent an organism as a constrained set of interconnected reactions and metabolites based on genomic sequencing data. This leads to a straightforward interpretation of metabolism as a constrained linear system, which enables the utilization of a wide range of well-developed analysis methods (Orth *et al.*, 2010).

The increasing ubiquity of genomic sequencing has led to a rapid expansion in the number and complexity of genome-scale metabolic models, e.g. the human metabolic model that has more than 80 000 reactions (Thiele *et al.*, 2020). Recent automated reconstruction tools can generate models spanning the entire primary metabolism of both pro- and eukaryotes (Machado *et al.*, 2018). Consequently, metabolic models are becoming considerably larger in scale than their predecessors, which is further compounded by the construction of multi-member community models. This growth implies increasing analysis complexity (see Supplementary Fig. S1), which in turn drives the need to develop analysis software that can accommodate this complexity. While computing the solutions to the underlying constrained optimization problems is hard to accelerate and parallelize, many analysis types can be decomposed into individual invocations of the optimizer, which may be parallelized. However, despite continued efforts (Heirendt *et al.*, 2017), this remains challenging due to the scalability limits of existing software implementations.

Here, we present COBREXA.jl, a package for implementing and running distributed COBRA workflows. The package is implemented in the Julia programming language (Bezanson *et al.*, 2017), enabling facile extension with user-defined numeric-computing routines, and interoperability with many high-performance computing packages. It provides a 'batteries-included' solution for scaling analyses to make efficient use of high-performance computing (HPC) facilities, giving researchers a powerful toolkit for executing complicated high-volume workflows, such as the creation and exploration of digital metabolic twins in personalized medicine (Björnsson *et al.*, 2019), and analysis of extensive microbial communities in ecology and biotechnology. We report the implementation architecture, and substantiate how the

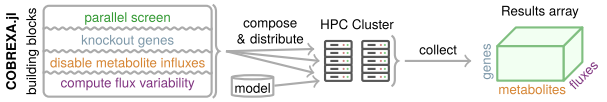


Fig. 1. Schema of an example custom analysis construction that examines flux variability in many variants of a model, its distributed execution with COBREXA.jl, and collection of many results in a multi-dimensional array

design accommodates future extensions and scaling of common analysis tasks.

2 Implementation and results

COBREXA.jl is an open architecture solution, providing interchangeable building blocks for implementing complicated COBRA workflows. Common analysis methods, such as flux balance, flux variability and gene knockout analyses (Gudmundsson and Thiele, 2010), are implemented as ready-to-use functions that may be easily composed and customized. Most importantly, the building blocks are designed so that the constructed workflows can be easily separated into parallelizable analysis steps and executed on multiple computation nodes in HPC environments (as illustrated in Fig. 1). The concurrent execution of such workflows results in significant computational speedups, without requiring user expertise in parallel programming.

The design of COBREXA.jl distinguishes it from other COBRA implementations, which typically provide parallelization support for only a few selected methods, and no current support for parallelization of custom method variants. For example, parallel single-gene deletion analysis is commonly supported, but a variant that explores the flux variability in knockouts must be reimplemented and parallelized by the user.

A variety of model exchange and representation formats are supported, including MATLAB format (Heirendt *et al.*, 2019); object-oriented JSON format (Ebrahim *et al.*, 2013), and SBML (Keating *et al.*, 2020). In addition, implementation of the workflows in Julia results in highly optimized execution of the code at the cost of minor pre-compilation overhead, which benefits large, data-heavy use cases. A detailed architecture overview is provided in Supplementary Section S1.

To evaluate the effect of the new architecture and optimizations on the performance and scalability of COBRA analyses, we benchmarked COBREXA.jl on use-cases that benefit from parallelization. We compared its performance to that obtained with COBRApy (Ebrahim *et al.*, 2013) and the COBRA Toolbox (Heirendt *et al.*, 2019), which are the widely adopted tools for running COBRA workflows. Running on a 256-CPU multi-node cluster, COBREXA.jl was able to fully utilize the available distributed computing resources and outperform the implementation of flux variability analysis in other packages by a factor of between 2× and 10×, even on relatively small models (Supplementary Table S2). We further demonstrated that COBREXA.jl is able to parallelize and distribute custom workloads by re-implementing the production envelope functionality of COBRApy; leading to speedups of over 10×, even on a single 16-core computation node (Supplementary Table S3). Consequently, we expect that the COBRA methods implemented in COBREXA.jl will enable reliable acceleration of many current and future workloads by simply adding more computing resources. The results are further discussed in Supplementary Section S3.4.

3 Conclusion

COBREXA.jl is a new package developed for large-scale distributed processing of constraint-based biological models. It differs from the other implementations of COBRA methods (Ebrahim *et al.*, 2013; Heirendt *et al.*, 2019) by focusing on computational efficiency, and simplifies high-level construction of parallelized user-defined analysis methods. This is required for performing extensive analyses of large models, future-proof extensibility and workload distribution that enables effective utilization of the common HPC infrastructure resources. The package thus enables fast analysis of datasets that may pose challenges for the currently available tools, such as the comprehensive human gut microbiome models.

Funding

This work was supported by the European Union's Horizon 2020 Programme under the PerMedCoE Project (www.permedcoe.eu) [951773]. This work was also partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC-2048/1–project ID 390686111 and EU's Horizon 2020 research and innovation program [862087]. The presented experiments were carried out using the HPC facilities of the University of Luxembourg (see https://hpc.uni.lu).

Conflict of Interest: none declared.

References

Apaolaza,I. *et al.* (2018) COBRA methods and metabolic drug targets in cancer. *Mol. Cell Oncol.*, **5**, e1389672.

Bezanson,J. *et al.* (2017) Julia: a fresh approach to numerical computing. *SIAM Rev.*, **59**, 65–98.

Björnsson,B. *et al.*; Swedish Digital Twin Consortium. (2019) Digital twins to personalize medicine. *Genome Med.*, **12**, 4–4.

Brunk,E. *et al.* (2018) Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.*, **36**, 272–281.

Cook,D.J. and Nielsen,J. (2017) Genome-scale metabolic models applied to human health and disease. *Wires Syst. Biol. Med.*, **9**, e1393.

Ebrahim,A. *et al.* (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.*, **7**, 74–76.

Fang,X. *et al.* (2020) Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat. Rev. Microbiol.*, **18**, 731–743.

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

Heirendt,L. *et al.* (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.*, **14**, 639–702.

Heirendt,L. *et al.* (2017) DistributedFBA.jl: high-level, high-performance flux balance analysis in Julia. *Bioinformatics*, **33**, 1421–1423.

Keating,S.M. *et al.*; SBML Level 3 Community members. (2020) SBML level 3: an extensible format for the exchange and reuse of biological models. *Mol. Syst. Biol.*, **16**, e9110.

Machado,D. *et al.* (2018) Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.*, **46**, 7542–7553.

Orth,J.D. *et al.* (2010) What is flux balance analysis? *Nat. Biotechnol.*, **28**, 245–248.

Thiele,I. *et al.* (2020) Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. *Mol Syst Biol*, **16**, e8982.

3 Manuscripts

3.1 Improving photosynthetic efficiency toward food security: Strategies, advances, and perspectives

The review "Improving photosynthetic efficiency toward food security: Strategies, advances, and perspectives" discusses bypassing photorespiration, enhancing light use efficiency, harnessing natural variation in photosynthetic parameters for breeding purposes, and adopting new-to-nature approaches that show promise for achieving unprecedented gains in photosynthetic efficiency. My contribution, together with Ed Smith, was writing the sections regarding the state of metabolic modelling of photorespiratory bypasses, specifically "A critical qualitative assessment of mechanistic hypotheses" and "Modeling at different scales can inform photorespiration engineering".

Publication	Improving photosynthetic efficiency toward food security: Strategies, advances, and perspectives
Author position	Second
Status	peer-reviewed and published
doi	10.1016/j.molp.2023.08.017

Molecular Plant
Review article

CellPress
Partner Journal

Improving photosynthetic efficiency toward food security: Strategies, advances, and perspectives

Edward N. Smith[1], Marvin van Aalst[2], Tiina Tosens[3], Ülo Niinemets[3], Benjamin Stich[4,10], Tomas Morosinotto[5], Alessandro Alboresi[5], Tobias J. Erb[6], Paul A. Gómez-Coronado[6], Dimitri Tolleter[7], Giovanni Finazzi[7], Gilles Curien[7], Matthias Heinemann[1], Oliver Ebenhöh[2], Julian M. Hibberd[8], Urte Schlüter[9], Tianshu Sun[8] and Andreas P.M. Weber[9,*]

[1]Faculty of Science and Engineering, Molecular Systems Biology – Groningen Biomolecular Sciences and Biotechnology, Nijenborgh 4, 9747 AG Groningen, the Netherlands

[2]Institute of Quantitative and Theoretical Biology, Cluster of Excellence on Plant Science (CEPLAS), Heinrich Heine University, Universitätsstrasse 1, 40225 Düsseldorf, Germany

[3]Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, 51006 Tartu, Estonia

[4]Institute of Quantitative Genetics and Genomics of Plants, Cluster of Excellence on Plant Science (CEPLAS), Heinrich Heine University, Universitätsstrasse 1, 40225 Düsseldorf, Germany

[5]Department of Biology, University of Padova, 35131 Padova, Italy

[6]Max Planck Institute for Terrestrial Microbiology, Department of Biochemistry & Synthetic Metabolism, 35043 Marburg, Germany

[7]Interdisciplinary Research Institute of Grenoble, IRIG-LPCV, Grenoble Alpes University, CNRS, CEA, INRAE, 38000 Grenoble, France

[8]Molecular Physiology, Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK

[9]Institute for Plant Biochemistry, Cluster of Excellence on Plant Science (CEPLAS), Heinrich Heine University, Universitätsstrasse 1, 40225 Düsseldorf, Germany

[10]Present address: Institute for Breeding Research on Agricultural Crops, Federal Research Center for Cultivated Plants, 18190 Sanitz, Germany

***Correspondence: Andreas P.M. Weber (aweber@hhu.de)**

https://doi.org/10.1016/j.molp.2023.08.017

ABSTRACT

Photosynthesis in crops and natural vegetation allows light energy to be converted into chemical energy and thus forms the foundation for almost all terrestrial trophic networks on Earth. The efficiency of photosynthetic energy conversion plays a crucial role in determining the portion of incident solar radiation that can be used to generate plant biomass throughout a growth season. Consequently, alongside the factors such as resource availability, crop management, crop selection, maintenance costs, and intrinsic yield potential, photosynthetic energy use efficiency significantly influences crop yield. Photosynthetic efficiency is relevant to sustainability and food security because it affects water use efficiency, nutrient use efficiency, and land use efficiency. This review focuses specifically on the potential for improvements in photosynthetic efficiency to drive a sustainable increase in crop yields. We discuss bypassing photorespiration, enhancing light use efficiency, harnessing natural variation in photosynthetic parameters for breeding purposes, and adopting new-to-nature approaches that show promise for achieving unprecedented gains in photosynthetic efficiency.

Key words: photosynthesis, photorespiration, photorespiratory bypass, natural variation, synthetic biology, plant metabolic engineering

Smith E.N., van Aalst M., Tosens T., Niinemets Ü., Stich B., Morosinotto T., Alboresi A., Erb T.J., Gómez-Coronado P.A., Tolleter D., Finazzi G., Curien G., Heinemann M., Ebenhöh O., Hibberd J.M., Schlüter U., Sun T., and Weber A.P.M. (2023). Improving photosynthetic efficiency toward food security: Strategies, advances, and perspectives. Mol. Plant. **16**, 1547–1563.

INTRODUCTION

Photosynthesis harnesses the energy of visible light quanta to extract electrons from water, utilizing them to convert atmospheric carbon dioxide (CO_2) into biomass. This crucial metabolic process originated over 2 billion years ago in an atmosphere abundant in CO_2 and low in oxygen (O_2). Throughout geological time, oxygenic photosynthesis caused a significant shift in the

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

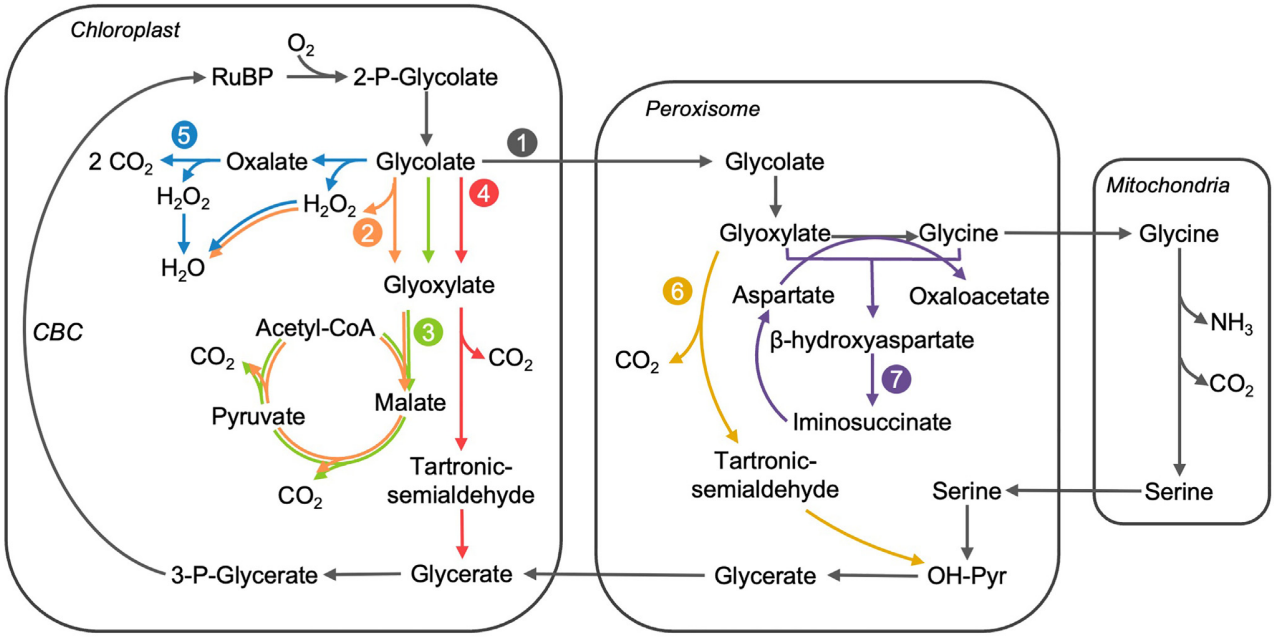


Figure 1. Bypasses to photorespiration discussed in this review.
Shown are (1) wild-type photorespiration, (2) Maier et al. (2012), (3) South et al. (2019), (4) Kebeish et al. (2007), (5) Shen et al. (2019), (6) Carvalho et al. (2011), and (7) Roell et al. (2021).

atmospheric O_2-to-CO_2 ratio, resulting in the present 500-fold excess of O_2 over CO_2. The enzyme responsible for CO_2 fixation in the Calvin–Benson cycle (CBC), known as ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco), exhibits a higher affinity for CO_2 than for O_2. However, under current atmospheric conditions, Rubisco frequently reacts with O_2, leading to production of 3-phosphoglyceric acid (3PGA) and 2-phosphoglycolic acid (2PG). Notably, 2PG acts as an inhibitor of key enzymes in the CBC; namely, triose phosphate isomerase and sedoheptulose 1,7-bisphosphate phosphatase (Flügel et al., 2017).

The two carbon atoms present in 2PG cannot be further metabolized within the CBC. Instead, conversion of 2PG to 3PGA occurs via a metabolic pathway known as photorespiration (Bauwe, 2023). During photorespiration, one of four carbon atoms contained in two molecules of 2PG is released as CO_2; i.e., previously fixed carbon is lost. Additionally, the process results in release of ammonia and consumption of ATP and redox power. Overall, photorespiration significantly diminishes the efficiency of carbon assimilation in C_3 plants, leading to yield losses of approximately 30% or higher (Walker et al., 2016a). Despite its negative impact on photosynthetic efficiency, photorespiration is an essential process that enables photosynthesis in an O_2-containing atmosphere through breakdown of 2PG (Bauwe, 2023). Mutations affecting photorespiration are typically lethal, even in the presence of carbon-concentrating mechanisms such as carboxysomes in cyanobacteria, pyrenoids in algae, or C_4 photosynthesis in land plants (Eisenhut et al., 2008; Zelitch et al., 2009; Levey et al., 2019). The topic of photorespiration has been reviewed extensively (Bauwe et al., 2010; Hodges et al., 2016; Eisenhut et al., 2019; Fernie and Bauwe, 2020; Bauwe, 2023; Broncano et al., 2023), rendering a detailed account of the pathway unnecessary in this work. Instead, here we focus on recent approaches aiming to mitigate the impact of photorespiration on photosynthetic efficiency.

To mitigate the oxygenation reaction of Rubisco, land plants have evolved carbon-concentrating mechanisms, including C_4 and crassulacean acid metabolism photosynthesis. These pathways, however, are complex and require specific leaf anatomy. Despite significant efforts, introduction of these photosynthetic subtypes into C_3 plants remains an unsolved challenge. Instead of reducing oxygenation, several research groups have focused on improving the efficiency of 2PG recovery, designing and implementing alternative pathways that bypass photorespiration, converting 2PG into CO_2, intermediates of the CBC, or metabolites of C_4 photosynthesis. In this review, we will explore these photorespiration bypasses and other strategies to enhance photosynthetic efficiency, considering their potential contributions to increased crop yields. Additionally, we will examine naturally occurring variations in photosynthetic efficiency and explore how such variations could be leveraged to enhance photosynthetic efficiency through breeding. Finally, we will discuss the potential of new-to-nature pathways to increase photosynthetic efficiency.

INCREASING PHOTOSYNTHETIC EFFICIENCY THROUGH PHOTORESPIRATORY BYPASSES

Chloroplast-localized photorespiratory bypasses

We will first describe the design and implementation of bypasses to photorespiration, focusing on *in planta* experimentally validated approaches (Figure 1). We will then discuss how such bypasses improve yield and novel strategies that have not yet

been tested in plants. We will start the discussion with the designs proposed by Kebeish et al. (2007) and Maier et al. (2012) because most of the reported photorespiratory bypasses since 2012 are variations of these schemes.

Intraplastidic conversion of 2PG into 3PGA by glyoxylate carboligase (GCL) and tartronate semialdehyde reductase

Kebeish et al. (2007) successfully introduced a bacterial glycolate metabolic pathway into the chloroplasts of *Arabidopsis thaliana*, a process that involved incorporation of five pathway enzymes fused with chloroplast-targeting peptides. The ability of certain bacteria to utilize glycolate as the sole carbon source served as the basis for this implementation. Within the bacterial glycolate pathway, glycolate dehydrogenase (GDH), composed of three subunits (D, E, and F), converts glycolate to glyoxylate. Subsequently, GCL catalyzes ligation of two glyoxylate molecules, resulting in formation of tartronic semialdehyde and release of one molecule of CO_2. Tartronic semialdehyde is further transformed into glycerate through the action of tartronic semialdehyde reductase. Native glycerate kinase converts glycerate to 3PGA.

To establish the complete bacterial glycolate pathway, Kebeish et al. (2007) introduced the corresponding pathway components using three different plasmids. Through genetic crossings of lines expressing partial pathways, they eventually combined all of the components to create the full pathway. The transgenic *Arabidopsis* plants expressing the full pathway exhibited a two-fold increase in shoot biomass and a three-fold increase in root biomass. The transgenics also demonstrated a decrease in the glycine-to-serine ratio, which is indicative of reduced photorespiratory flux. Moreover, the post-illumination burst of CO_2 release, a measure of photorespiratory glycine in the light, was reduced.

Additionally, the lines expressing the complete bacterial pathway showed a minor reduction in O_2-inhibition of photosynthetic carbon assimilation and a slight decrease in the CO_2 compensation point. These findings suggest that implementation of the pathway resulted in a modest elevation of CO_2 levels at the site of Rubisco. Intriguingly, transgenic lines expressing only GDH D, E, and F also exhibited significant increases in biomass and rosette diameter, along with decreased Gly/Ser ratios. However, the reason behind the improved plant performance solely from expression of bacterial GDH remains unexplained.

Intraplastidic glycolate oxidation by glycolate oxidase

Maier et al. (2012) developed an alternative photorespiratory bypass strategy by redirecting peroxisomal enzymes to the chloroplasts. This pathway involves complete oxidation of glycolate to CO_2 within the chloroplast stroma. To achieve this, peroxisomal glycolate oxidase is targeted to the chloroplasts, where it catalyzes oxidation of glycolate to glyoxylate, releasing hydrogen peroxide as a byproduct. The chloroplast-targeted catalase then dissipates the hydrogen peroxide. Subsequently, glyoxylate and acetyl-coenzyme A (CoA) are condensed to malate by a plastid-targeted malate synthase, an enzyme from the peroxisomal glyoxylate cycle. Malate is decarboxylated by the native chloroplast NADP-malic enzyme, generating pyruvate and Nicotinamide adenine dinucleotide phosphate (NADPH). The resulting pyruvate is further decarboxylated by the native chloroplast pyruvate dehydrogenase, yielding acetyl-CoA and NADH. Acetyl-CoA, along with another glyoxylate molecule, re-enters the cycle through malate formation via malate synthase. Overall, this cycle completely oxidizes the carbon present in glycolate and releases CO_2 within the chloroplast stroma. Additionally, it generates NADPH and NADH as reducing equivalents. Unlike the pathway described by Kebeish et al. (2007), the Maier et al. (2012) pathway requires introduction of only three transgenes.

Transgenic *Arabidopsis* plants expressing the glycolate oxidizing pathway showed significantly increased rates of CO_2 assimilation and a decreased Gly/Ser ratio. However, in contrast to the findings of Kebeish et al. (2007), the CO_2 compensation point remained unchanged. This result is surprising considering that the local CO_2/O_2 ratio within the chloroplasts of these lines is expected to be higher than in the lines of Kebeish et al. (2007). One of the two lines analyzed by Maier et al. (2012) exhibited increased leaf fresh and dry weight, along with a reduction in leaf thickness.

Expression of GDH in chloroplasts

Kebeish et al. (2007) observed that transgenic *Arabidopsis* plants expressing all three subunits of *Escherichia coli* GDH exhibited enhanced biomass accumulation and reduced flux through the conventional photorespiration pathway. Building on this discovery, Nölke et al. (2014) created transgenic potato plants that were genetically modified to express a single GDH polyprotein, wherein the D, E, and F subunits were linked by a $(Gly_4Ser)_3$ linker sequence. The polyprotein was targeted to the chloroplasts using a Rubisco small subunit-targeting peptide (rbcS1).

The transgenic potato plants demonstrated elevated rates of CO_2 assimilation at a 400 ppm CO_2 concentration, decreased repression of CO_2 assimilation by O_2, and a lowered CO_2 compensation point. Moreover, these transgenic lines exhibited increased above-ground biomass accumulation compared with the control plants, with a more than 2-fold increase in tuber yield observed in lines showing the highest GDH activity. The precise mechanism underlying the augmented biomass production and yield in the transgenic plants was not extensively investigated in this study. However, the authors put forward the hypothesis that chloroplast-produced glycolate is decarboxylated by plastidial pyruvate dehydrogenase, resulting in a localized rise in CO_2 concentration at the Rubisco site.

Combination of plastid-localized photorespiration bypasses with reduced export of glycolate from chloroplasts

South et al. (2019) conducted a comparative analysis of three alternative designs for photorespiratory bypasses (AP1–AP3) in tobacco, a model crop, and evaluated the performance of transgenic plants in field trials. AP1 corresponds to the pathway described by Kebeish et al. (2007), AP2 is based on the pathway reported by Maier et al. (2012), and AP3 is a modified version of the Maier et al. (2012) pathway. In AP3, the combination of peroxisomal glycolate oxidase and catalase was replaced by mitochondrial GDH from *Chlamydomonas reinhardtii*, which was retargeted to the chloroplasts. Unlike the

bacterial GDH with its three subunits, the algal mitochondrial enzyme consists of a single subunit and does not produce H_2O_2 (the electron acceptor remains unknown), eliminating the need for co-expression of catalase. Consequently, the AP3 design only requires two transgenes: GDH and malate synthase.

Furthermore, all three alternative pathway designs were combined with a reduction in glycolate export from chloroplasts by employing antisense repression of the chloroplastic glycolate/glycerate transporter (PLGG1; Pick et al., 2013). T2 transformants were initially screened under high-light and low-CO_2 conditions to identify lines that exhibited enhanced protection against photorespiratory stress. This pre-selection step was crucial for identifying lines with optimal expression levels and stoichiometry of the pathway components. In greenhouse trials, all pathway designs were associated with increased biomass production, with AP2 and AP3 performing better when combined with repression of PLGG1. Subsequently, the AP3 design was evaluated in replicated field trials with and without repression of PLGG1. Surprisingly, contrary to the greenhouse experiments, the AP3 design performed best in the field without repression of PLGG1. The field-grown AP3 lines exhibited significantly higher biomass productivity, increased CO_2 assimilation rates, and decreased CO_2 compensation points. Additionally, AP3 lines displayed elevated levels of glyoxylate, while serine and glycerate levels were significantly reduced.

This study is significant because it validated previously reported bypass designs through a comparative analysis in the same model system, demonstrating that growth benefits observed in a greenhouse setting can be reproduced in the field. However, it is worth noting that lines expressing only a chloroplast-targeted GDH from *C. reinhardtii* were not included in this comparison. This omission is a limitation because it could have provided insights into whether the improved performance observed for AP3 is truly dependent on the coordinated activity of GDH and malate synthase or whether, similar to Nölke et al. (2014), GDH alone can confer growth benefits.

Complete oxidation of glycolate via oxalate oxidase

The study of Shen et al. (2019) describes a modified pathway, based on the work of Maier et al. (2012), that enables complete intraplastidic decarboxylation of glycolate. This pathway involves conversion of glycolate to oxalate through the action of rice glycolate oxidase isoform 3, previously demonstrated to oxidize glycolate and glyoxylate (Zhang et al., 2012). The resulting oxalate is then fully decarboxylated to CO_2 by rice oxalate oxidase 3. The accumulation of H_2O_2, a byproduct of glycolate and oxalate oxidation, is effectively eliminated by a chloroplast-targeted catalase. Consequently, implementation of this pathway necessitates incorporation of three transgenes encoding the chloroplast-targeted enzymes: glycolate oxidase, oxalate oxidase, and catalase.

Rice plants expressing these genetic constructs exhibited enhanced photosynthetic performance, characterized by a reduced CO_2 compensation point and a higher maximum photosynthetic rate under saturating light conditions. Furthermore, the Gly/Ser ratio and glycolate levels decreased, while glyoxylate and oxalate levels increased under ambient air conditions. Notably, the yield of single plant seeds varied depending on the seeding season, with a 27% increase observed in spring seeding but a yield penalty of 13%–16% in fall seeding.

In another variation of the pathways proposed by Kebeish et al. (2007) and Maier et al. (2012), Wang et al. (2020a) substituted the bacterial GDH component of the GCL route with the glycolate oxidase/catalase system in transgenic rice plants. This approach effectively combined aspects of the two previously reported bypasses. The transgenic plants exhibited improved photosynthetic parameters and yield in replicated field trials, further confirming the potential of photorespiratory bypasses to enhance crop performance under field conditions.

Peroxisome-localized photorespiratory bypasses

In the native photorespiration process, glycolate is transported from chloroplasts to peroxisomes. Within the peroxisomes, glycolate undergoes oxidation by glycolate oxidase, resulting in production of glyoxylate. Subsequently, glyoxylate is transaminated to form glycine. The glycine molecules are then transported to the mitochondria, where glycine decarboxylase and serine hydroxymethyltransferase catalyze the conversion of two glycine molecules into serine, CO_2, and ammonia. The generated serine is subsequently transported back to the peroxisomes, where it is converted into hydroxypyruvate by the action of serine:glyoxylate aminotransferase. Hydroxypyruvate is further reduced to glycerate, which is then transported back to the chloroplasts, thereby completing the photorespiratory cycle. Given the crucial role of peroxisomes in photorespiration, researchers have endeavored to develop strategies for bypassing the ammonia- and CO_2-releasing step that occurs in the mitochondria through alternate glycolate conversion within peroxisomes. Two such attempts have been documented in the literature.

Peroxisomal conversion of glyoxylate to hydroxypyruvate

This pathway design capitalizes on conversion of glycolate to glyoxylate by the inherent peroxisomal glycolate oxidase, aiming to redirect glyoxylate toward hydroxypyruvate through involvement of GCL and hydroxypyruvate isomerase. Similar to the pathway design proposed by Kebeish et al. (2007), bacterial GCL is employed to convert glyoxylate into tartronate semialdehyde, which is subsequently transformed into hydroxypyruvate by hydroxypyruvate isomerase. Essentially, this strategy enables retrieval of 75% of the carbon content present in two glyoxylate molecules while mitigating the release of ammonia mediated by mitochondrial glycine decarboxylase. Transgenic tobacco plants carrying constructs encoding peroxisome-targeted versions of these bacterial enzymes were generated.

Under non-photorespiratory conditions at high CO_2, the transgenic plants exhibited robust growth. However, when exposed to current ambient CO_2 conditions, the leaves displayed yellow lesions, and the plants exhibited a chlorotic phenotype. Metabolic labeling experiments employing [^{14}C]-glycolate revealed that glycolate was still predominantly converted into glycine and subsequently serine, suggesting that only a minor fraction of glycolate entered the engineered pathway. Surprisingly, amino acid analysis indicated that the leaves of the transgenic plants contained higher levels of glycine and serine compared with wild-type plants, contrary to initial expectations. Notably, the researchers were unable to detect the

presence of the hydroxypyruvate isomerase protein in the transgenic plants through immunoblotting despite detectable expression of the transgene confirmed by RNA gel blots. These findings suggest that the pathway may have been incomplete and that the observed phenotypes could arise from accumulation of undesired tartronate semialdehyde in peroxisomes, exerting adverse effects.

Peroxisomal glyoxylate-to-oxaloacetate conversion via the β-hydroxyaspartate shunt

In this study (Roell et al., 2021), a recently discovered microbial pathway involved in the metabolism of glyoxylate was introduced into plant peroxisomes. The pathway, known as the β-hydroxyaspartate cycle, encompasses four enzymatic steps that convert glyoxylate and glycine into oxaloacetate. The sequential actions of β-hydroxyaspartate aldolase, β-hydroxyaspartate dehydratase, and iminosuccinate reductase transform glyoxylate and glycine into β-hydroxyaspartate, iminosuccinate, and aspartate, respectively. Finally, aspartate: glyoxylate aminotransferase converts glyoxylate into glycine and releases oxaloacetate as the end product of the pathway. Transgenic *Arabidopsis* plants were engineered to express the pathway enzymes fused with peroxisomal targeting signals. Promoters that drive gene expression specifically in photosynthetic tissues were employed to prevent undesired pathway activity in non-photosynthetic plant organs.

Roell et al. (2021) hypothesized that reducing the conversion of glyoxylate to glycine would enhance metabolic flux through the β-hydroxyaspartate cycle. Therefore, apart from wild-type *Arabidopsis* plants, the pathway was introduced into the genetic background of the *ggt1-1* mutant, which lacks peroxisomal glutamate:glyoxylate aminotransferase 1 and exhibits a photorespiratory phenotype under ambient air conditions. This enabled investigation of the β-hydroxyaspartate cycle's function by assessing its ability to complement the visual phenotype of the *ggt1-1* mutant.

Similar to the findings of Carvalho et al. (2011), wild-type plants expressing the β-hydroxyaspartate cycle exhibited reduced growth and photosynthetic rates in ambient air. However, when exposed to elevated CO_2 concentrations that suppress photorespiration, their growth was comparable with that of wild-type controls. Importantly, expression of the β-hydroxyaspartate cycle in the *ggt1-1* mutant partially rescued the photorespiratory phenotype, indicating that the introduced pathway fulfilled its expected function to some extent.

Metabolic analysis was conducted to investigate the underlying reasons for the impaired growth observed in wild-type transgenic plants expressing the β-hydroxyaspartate cycle. These plants, as well as the transformed *ggt1-1* mutants, exhibited elevated levels of aspartate and malate, while glycine levels were reduced. Furthermore, intermediates of the CBC, such as 3-phosphoglycerate and sedoheptulose 7-phosphate, were depleted, suggesting a decrease in the availability of CBC intermediates and unproductive metabolic flux into C_4 acids.

Collectively, the studies by Carvalho et al. (2011) and Roell et al. (2021) indicate that perturbing the peroxisomal steps of the canonical photorespiration pathway does not yield the anticipated improvement in photosynthetic performance.

Although not extensively discussed in this review, we note that bypasses to photorespiration often led to a multitude of pleiotropic changes, including alterations in leaf shape and anatomy, metabolic changes, and developmental effects. For example, changes in leaf anatomy can have a multitude of effects on leaf photosynthetic activities. Changes in leaf thickness alone can result in increased photosynthetic activity because of more photosynthetic biomass per unit leaf area (Onoda et al., 2017). Changes in leaf anatomy, such as modifications of leaf thickness, intercellular airspace (IAS) volume, mesophyll cell wall thickness, and chloroplast size can also affect mesophyll conductance for CO_2 (g_m; see Figure 2 for a schematic explanation of g_m) and thereby alter the CO_2 concentration in chloroplasts (Flexas et al., 2012, 2013; Knauer et al., 2022). Given that observed changes in the CO_2 compensation points have often been small in the abovementioned studies, consideration of possible changes in g_m is important. We also note that the CO_2 compensation point does not only depend on Rubisco characteristics and CO_2 and O_2 concentrations but is also affected by the rate of mitochondrial respiration in light (R_d). R_d may change as a consequence of pathway engineering; for example, generation of extra respiratory substrate, such as malate transported out of chloroplasts.

The precise relationship between modified photorespiration and its impact on plant structure and function beyond photosynthesis remains incompletely understood. It is important to acknowledge that native photorespiration does not operate as a closed cycle in which 75% of the carbon derived from glycolate is reincorporated into the CBC as glycerate and 25% is released as CO_2. Instead, considerable amounts of carbon can be diverted toward synthesis of amino acids glycine and serine (Samuilov et al., 2018; Abadie and Tcherkez, 2019; Fu et al., 2023a), which serve as building blocks for protein biosynthesis and other plant metabolites. Moreover, photorespiration indirectly transfers redox equivalents from chloroplasts to mitochondria (Heber and Krause, 1980; Heber et al., 1996), where they are oxidized by the mitochondrial electron transport chain, contributing to ATP biosynthesis. The observed effects extending beyond photosynthetic metabolism may be linked to these additional, albeit less extensively studied and recognized, functions of the photorespiratory pathway. In the following paragraphs, we will provide a critical assessment of the assumptions as to how photorespiratory bypasses function, and we highlight unresolved questions regarding these pathways.

A critical qualitative assessment of mechanistic hypotheses

Despite the fact that some of the genetic implementations of photorespiratory bypasses increased photosynthetic efficiency and even yield under some conditions, many questions remain unresolved. In particular the precise molecular mechanisms that are responsible for higher performance are not yet understood. To develop informed hypotheses that are also backed up by theoretical considerations, we give some qualitative and quantitative arguments about attempts to explain increased photosynthetic rates.

First, it should be noted that an observation of biomass increases of a certain percentage after a growth period (usually several weeks long) cannot be directly translated into increased fluxes. Even marginal differences in flux will, over time, lead to an

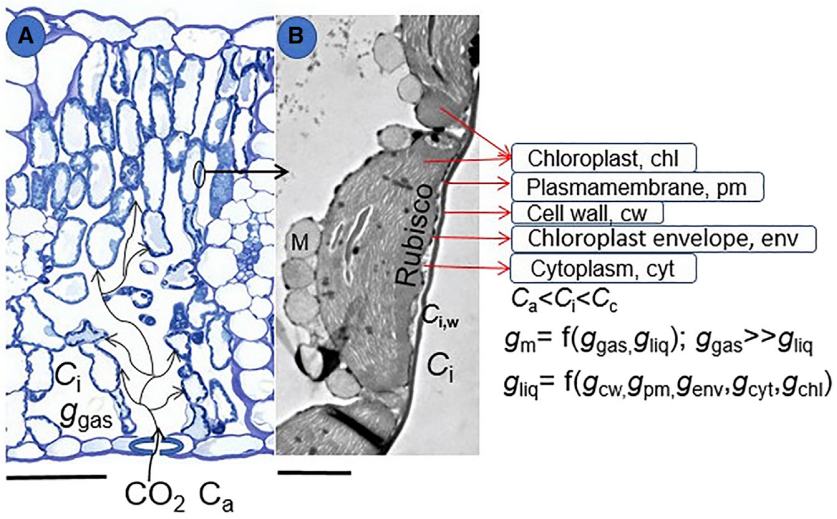


Figure 2. Semi- and ultrathin cross-sections of *Helianthus occidentalis* leaf mesophyll.

(A and B) Semi- and ultrathin cross-sections of *Helianthus occidentalis leaf mesophyll* ***(A)*** and palisade cells **(B)** to illustrate the CO_2 diffusion pathway from ambient air (C_a) to substomatal cavities (C_i) through IASs to the outer surface of the mesophyll cell wall ($C_{i,w}$) and farther into the chloroplast (C_c). The CO_2 concentration drawdown, C_a–C_i, is modulated by stomatal conductance. Mesophyll conductance (g_m) is determined by gas- and liquid-phase conductance. The C_i–$C_{i,w}$ drop is modulated by gas-phase diffusion conductance (g_{ias}), depending on mesophyll thickness and effective porosity of mesophyll airspace. The CO_2 drawdown from the outer surface of cell walls to chloroplasts, $C_{i,w}$–C_c, is determined by liquid-phase diffusion conductance (g_{liq}), which is determined by multiple liquid and lipid phase barriers: cell wall (cw), plasma membrane (pm), cytoplasm (cyt), chloroplast envelope (env), and chloroplast stroma (chl). Thus, the physical dimensions of each anatomical component of g_m determine its partial conductance, largely setting the maximum g_m in a given species (Tosens et al., 2012). In this context, the cell periphery facing the IAS is largely enveloped by chloroplasts ($S_c/S_{mes} \sim 1$, **B**). On a global scale, S_c/S_{mes} varies from 0.3–0.98. A high S_c/S_{mes} signifies the direct passage of CO_2 fluxes from the IAS into the chloroplasts; this configuration also facilitates efficient recycling of respiratory CO_2 fluxes as chloroplasts are covered by mitochondria (M) (Busch et al., 2013). Scale bars: **(A)** 0.03 mm, **(B)** 2 μm. Unpublished images by T.T.

exponential difference in overall biomass accumulation. For example, in South et al. (2019), it was shown that, under saturating CO_2, the assimilation rate is increased by approximately 10% (cf. South et al., 2019; Figure 5A). However, this increase is considerably lower than the reported 24% increase in biomass after a growth period of 6 weeks. This example illustrates that yield gains in percent stated for the diverse experimental approaches are not comparable on a quantitative level.

Other observations in South et al. (2019) are also challenging to explain. For example, it has been suggested that an increase in CO_2 locally in the chloroplast could explain a higher carbon fixation rate. However, increased CO_2 assimilation was observed even for saturating conditions, which entails that increasing the local concentration is not the primary cause for higher fixation rates. A similar result was observed when a complete glycolate decarboxylation pathway was expressed in rice, with several transgenic lines showing increased maximum rates of Rubisco carboxylation under saturating CO_2 (Shen et al., 2019). Another hypothesis was that, as a result of the glycolate oxidase activity, glycolate levels should be reduced, thus reducing its toxic effects. However, in various lines, the glycolate levels were actually increased, making this explanation unlikely. Similarly, although not explicitly measured, 2PG levels might be reduced as a result of the newly introduced pathways, but because dephosphorylation of 2PG to glycolate is highly irreversible, there is no convincing argument why 2PG levels should actually be reduced. We note, however, that it has been shown previously that more efficient removal of 2PG by overexpression of 2-phosphoglycolate phosphatase improved photosynthetic performance in *A. thaliana* under stress conditions (Timm et al., 2019).

Possibly, the growth promoting effect is of a more indirect nature. A possible explanation could be connected with the algal GDH, which transfers electrons not to H_2O_2 but to a so far unidentified electron acceptor. If this acceptor is one of the common electron carriers of the photosynthetic electron transport chain, such as plastoquinone or ferredoxin, then this new pathway would contribute to generation of redox equivalents and, thus, directly support the photosynthetic electron transport chain. The observation that the transformed plants exhibit a higher apparent quantum efficiency (φ; cf. South et al., 2019; Figure 6) is in line with this hypothesis. An experimental test would be to measure how the observed increase in growth depends on the light intensity under which plants are grown. If this speculation is correct, then the growth increase should be less pronounced the more saturating the light intensity.

In summary, these considerations illustrate the complexity of the system and the necessity to use mathematical models, which are based on clear mechanistic hypotheses and are designed to quantitatively reproduce experimental results and, thus, provide a platform to test different mechanistic hypotheses and, by making novel predictions, support the experimental design to confirm or falsify these.

Modeling at different scales can inform photorespiration engineering

The mechanisms of yield improvement caused by photorespiratory bypasses can be divided into four categories: improved stoichiometry and energy efficiency, improved kinetics (i.e., altered Rubisco carboxylation-to-oxygenation ratio), relief of inhibition by toxic intermediates, and indirect effects of altered physiology or development. Models at different scales, from simple cofactor accounting and stoichiometric models to kinetic models of photosynthesis, have provided some insight into how these mechanisms allow photorespiratory bypasses to be effective, but further work is still required if models are to explain all observed experimental results (Peterhansel et al., 2013; Xin et al., 2015; Basler et al., 2016; Trudeau et al., 2018; Khurshid et al., 2020; Osmanoglu et al., 2021).

Reference	Pathway name	CO_2 (α)	ATP (β)	Red. equiv. (γ)	Total ATP equiv.[a]	Total ATP equiv.[a] normalized to net 0 carbon lost/gained[b]
Trudeau et al., 2018	TaCo	1	7	4	17	9
Kebeish et al., 2007	chloroplast TSS	−0.5	3	1	5.5	9.5
Shih et al., 2014	3OHP[c]	1	8	4	18	10
Carvalho et al., 2011	peroxisome TSS	−0.5	3	2	8	12
Roell et al., 2021	BHAC[d]	−0.5	3	2	8	12
	photorespiration	−0.5	3.5	2	8.5	12.5
South et al., 2019	AP3	−2	2	−2	−3	13
Maier et al., 2012	GMK	−2	2	−1	−0.5	15.5
Shen et al., 2019	GOC	−2	2	1	4.5	20.5
	Calvin cycle	1	3	2	8	

Table 1. Photorespiratory bypass energy costs (adapted from Trudeau et al., 2018).
The table is based on the consumer model of photosynthesis, which describes the processes of photorespiration and the CBC as independent cycles that are able to regenerate ribulose 1,5-bisphosphate (RuBP) using ATP (β) and reducing equivalents (γ) and either consume or produce CO_2 and glyceraldehyde 3-phosphate (GAP) (α) (Trudeau et al., 2018). This description allows bypasses to be compared directly and separated from the CBC. Positive values represent consumption, and negative values represent production. See Supplemental Table 1 for detailed calculations.
3OHP, 3-hydroxypropionate; Red equiv, reducing equivalents; TaCo, tartonyl-CoA; TSS, tartronic-semialdehyde shunt; BHAC, β-hydroxyaspartate cycle; GMK, glycolate oxidase, malate synthase, catalase (KatE); AP3, alternative pathway 3; GOC, glycolate oxidase, oxalate oxidase, catalase.
[a]Assuming 2.5 ATP per reducing equivalent.
[b]Assuming CBC compensates for carbon lost by PR bypasses.
[c]3OHP bypass assuming that pyruvate is converted to GAP via pyruvate phosphate dikinase.
[d]BHAC bypass assuming that oxaloacetate is converted to GAP via phosphoenolpyruvate-carboxykinase to regenerate RuBP.

A common feature of photorespiratory bypasses is to avoid the energetic cost of ammonium refixation or to capture the reducing power from glycolate oxidation, thus decreasing the ATP and NADPH cost of photorespiration. Models for energy cofactor accounting can quantify these direct energetic benefits of bypassing photorespiration as well as indirect benefits, such as avoiding the cost of CO_2 refixation via the CBC in carbon fixing bypasses (Table 1; Peterhansel et al., 2013; Trudeau et al., 2018). Larger, genome-scale stoichiometric models can also calculate energy efficiency and have the advantage of predicting flux into biomass rather than just rates of carbon fixation, placing bypasses in the wider context of the plant metabolic network (Basler et al., 2016). For example, a curated stoichiometric model was able to predict a decrease in photorespiratory flux and a biomass output increase of ∼6.2%, qualitatively consistent with experimental data (Kebeish et al., 2007; Basler et al., 2016). However, stoichiometric models predict no benefit for bypasses that are energetically more costly than photorespiration, such as those that completely decarboxylate glycolate in the chloroplast. Experimentally, such bypasses still show increased yields when expressed in plants, suggesting either incorrect prediction of energy costs or benefits that are beyond just direct ATP and NADPH savings (Maier et al., 2012; Peterhansel et al., 2013; Xin et al., 2015; Shen et al., 2019; South et al., 2019).

Photorespiration involves transport of metabolites between three compartments as well as movement of reducing equivalents from the chloroplasts to the mitochondria. Photorespiratory bypasses can relocate reactions to different compartments, potentially avoiding the energetic costs of metabolite transport. Therefore, more complete modeling of transport reactions, including thermodynamic constraints, may improve the accuracy of energy accounting. However, even if all energetic costs were accurately modeled, stoichiometric models alone cannot account for changes in metabolite concentrations or reaction kinetics.

Several of the bypasses validated in plants aim to relocate the release of photorespiratory CO_2 from mitochondria to chloroplasts, which has two potential advantages: increasing the CO_2 concentration at the site of Rubisco and recapturing photorespiratory CO_2 that could otherwise be lost from the cell by diffusion out of mitochondria. Predicting such effects requires use of kinetic models or additional constraints on Rubisco carboxylation and oxygenation fluxes (Basler et al., 2016). A kinetic model predicted that, under high light conditions, the entire benefit of the Kebeish et al. (2007) bypass is relocation of CO_2 release, not the reduced ATP cost, which only contributes under low-light conditions (Xin et al., 2015). However, the same kinetic model failed to explain the benefit of complete glycolate decarboxylation in the chloroplast, predicting that the photosynthetic rate would be 31% lower than in the wild type, despite reported increases of more than 30% in carbon assimilation rate and biomass (Maier et al., 2012; Xin et al., 2015). Additionally, the kinetic model demonstrated that any benefit of relocating CO_2 release is dependent on the CO_2 permeability of the chloroplasts; predicting that, if more than 30% of photorespiratory CO_2 is already recaptured and refixed in wild-type plants, then there is no benefit to the bypass (Xin et al., 2015). It is therefore surprising that a glycolate decarboxylation bypass in rice, where wild-type plants already refix 38% of photorespiratory CO_2, still showed some biomass gains, suggesting additional benefits not captured by the kinetic model (Busch et al., 2013; Xin et al., 2015; Shen et al., 2019; Wang et al., 2020a; Zhang et al., 2022).

In contrast to Xin et al. (2015), a different kinetic model of a cyanobacterial complete glycolate decarboxylation bypass

expressed in the chloroplast predicted a 10% increase in photosynthetic rate, although this model did not explicitly account for the movement of CO_2 between compartments and has yet to be experimentally validated by expression of the full pathway in plants (Bilal et al., 2019; Khurshid et al., 2020; Abbasi et al., 2021). Kinetic models have been useful for estimating the potential effect of relocation of CO_2 release and distinguishing this from energy benefits, but they are still unable to explain all currently observed experimental results.

Reducing the concentration of inhibitory intermediates of photosynthetic metabolism could account for some of the benefit of photorespiratory bypasses, and this can be simulated using kinetic models. However, kinetic models of photorespiratory bypasses did not include parameters for the inhibitory effects of 2-phosphoglycolate on the CBC or the inhibitory effect of glyoxylate on Rubisco activation and Rubisco oxygenase activity (Oliver and Zelitch, 1977; Oliver, 1980; Campbell and Ogren, 1990; Xin et al., 2015; Flügel et al., 2017; Khurshid et al., 2020). Increasing the scale of kinetic models to include more inhibition terms may provide additional explanation for benefits, although this can be limited by the availability of accurate kinetic parameters.

As pointed out previously, a higher apparent φ could be responsible for yield improvements in the transformed plants. Modeling the effect of bypasses on φ would require a high-fidelity kinetic model of the photosynthetic electron transport chain and the CBC, such as in Saadat et al. (2021), to be combined with a model of photorespiration.

Additional effects beyond the cellular metabolic changes described by current models, such as altered gene expression, signaling, physiology, and pleiotropic effects, could also explain a large proportion of observed growth benefits (Maier et al., 2012; Shen et al., 2019). Using larger-scale integrated models could potentially provide more accurate prediction of the effect of engineering photorespiration in the field by accounting for interactions across scales (Wu, 2023). Additionally, attempting to model dynamic processes under non-steady-state conditions may also help identify advantages of photorespiratory bypasses not captured by current models (Fu et al., 2023b). Finally, extending current metabolic models to account for diurnal cycles could explain advantages of bypass reactions beyond altered photosynthetic metabolism, such as altered dark respiration and sucrose export during the night (Dalal et al., 2015).

No single modeling strategy can explain all of the observed phenotypes of plants expressing photorespiratory bypasses; instead, multiple models at different scales should be used as tools to help explain the underlying mechanisms of growth benefits. Questions still remain to be tested by exploring models: why does expression of GDH alone also increase photosynthetic performance and yield in *Arabidopsis*, potato, and *Camelina* (Kebeish et al., 2007; Nölke et al., 2014; Dalal et al., 2015; Abbasi et al., 2021), and how does the anatomy and physiology of different crop species affect CO_2 diffusion between cells and subcellular compartments? Future implementation of new bypass designs based on carbon-fixing rather than decarboxylating reactions will also provide valuable data for exploring the effect of a fundamentally different bypass mechanism (Trudeau et al., 2018; Scheffen et al., 2021) (see discussion of new-to-nature pathways below). With future modeling efforts and better understanding, it may be possible to further increase yields and improve the transferability of benefits between different crop species.

COMBINED OPTIMIZATION OF LIGHT USE AND CARBON ASSIMILATION EFFICIENCY TO ENHANCE PLANT PRODUCTIVITY

The intricate relationship between light conversion, carbon fixation, and their impact on plant productivity calls for a comprehensive exploration of their interdependencies. Through combining modifications of light utilization and carbon fixation, we have the potential to achieve synergistic enhancements. Here we discuss the interplay between these pathways and evaluate the possibility of maximizing overall productivity by improving the efficiency of converting light energy into ATP and NADPH while optimizing their utilization for carbon fixation.

In plants, pigments absorb light, and the excitation energy is utilized by photosystems to facilitate synthesis of ATP and NADPH, which are essential for all metabolic reactions, including carbon fixation. This initial phase of photosynthetic reactions effectively transforms light into chemical energy, and its efficiency significantly influences crop productivity (Zhu et al., 2010).

In highly dynamic environments, the amount of light absorbed by the photosynthetic apparatus can exceed the metabolic capacity of the cell, resulting in over-reduction of the photosynthetic electron transport chain and production of harmful reactive O_2 species. Photosynthetic organisms have evolved various mechanisms to regulate light utilization efficiency and photosynthetic electron transport, aiming to minimize the likelihood of over-reduction and cellular damage by safely dissipating excess excitation or electrons (Li et al., 2009). While protection of the photosynthetic apparatus plays a crucial biological role, it comes at the expense of reduced efficiency in converting sunlight into chemical energy (Alboresi et al., 2019).

Fluctuations in light intensity represent a particular challenge to regulation of photosynthesis and significantly impact primary productivity (Long et al., 2022). Abrupt increases in illumination can be detrimental because they do not allow sufficient time for activation of regulatory responses and modulation of metabolic reactions. Conversely, when light levels decrease from excessive to limiting, the photoprotective mode remains active for several minutes, leading to unnecessary energy dissipation and a subsequent reduction in carbon fixation efficiency (Wang et al., 2020b).

To address this limitation, modifications were made to the kinetics of non-photochemical quenching (NPQ), which is one of the mechanisms involved in light harvesting regulation. By overexpressing key proteins such as the Photosystem II subunit S, violaxanthine de-epoxidase, and zeaxanthin epoxidase the activation and relaxation kinetics of NPQ were accelerated. This overexpression allowed tobacco and soybean plants to effectively respond to changes in light and minimize potential damage during sudden increases in sunlight. Additionally, it

facilitated faster relaxation of NPQ when illumination intensity decreased. This approach has demonstrated positive effects on biomass productivity not only in the field, for crops like tobacco and soybean (Kromdijk et al., 2016; Souza et al., 2022), but also in high-density culture photobioreactors for the microalga *Nannochloropsis* (Perin et al., 2023). However, in *Arabidopsis*, the same approach was not as successful, suggesting that species-specific traits, including light distribution within the canopy, significantly impact plant productivity and the optimal balance between light harvesting and photoprotective responses (Garcia-Molina and Leister, 2020).

Another strategy to enhance the efficiency of sunlight utilization is modification of the light-harvesting apparatus. Leaves have evolved to efficiently capture light and outcompete other organisms for sunlight. However, in densely cultivated fields, this high light harvesting efficiency can have negative consequences. Light is primarily absorbed by the uppermost leaves in the canopy, leaving fewer photons available for the lower layers. While high light harvesting efficiency provides a competitive advantage in natural environments where individuals vie for a limited resource like light, it proves detrimental in cultivated fields. In this context, plants with paler leaves can enable a more uniform distribution of light within the canopy, which has been shown to enhance overall productivity (Rotasperti et al., 2022; Cutolo et al., 2023). Also, extra nitrogen that is no longer needed for construction of the pigment-binding machinery of photosynthesis could be invested in Rubisco and rate-limiting proteins of photosynthetic machinery, thereby increasing leaf photosynthetic capacity (Walker et al., 2017; Niinemets, 2023). It is interesting here to discuss whether these efforts in modulating regulation of light harvesting could be combined with improvements in carbon fixation efficiency. In principle, a higher-efficiency conversion of light energy into ATP and NADPH could be combined with more efficient utilization of these molecules for carbon fixation, with a potentially additive effect on productivity.

One possible implication to be considered is that photorespiration has been shown to be a major sink for photosynthetic electron transport and that this energy loss can have a beneficial effect under conditions where the photosynthetic apparatus is overexcited (Heber and Krause, 1980; Kozaki and Takeba, 1996; Hanawa et al., 2017). Light saturation occurs when electron transport is faster than the metabolic capacity of consuming ATP and NADPH produced, and, under these conditions, photorespiration can be beneficial in reducing oversaturation. If efficiency in carbon fixation is improved by introduction of photorespiratory bypasses, however, then this would also drive stronger ATP and NADPH consumption with a similar protective effect. This suggests that the critical point is that the new or modified pathways have a sufficient capacity to also compensate for ATP and NADPH consumption associated with photorespiration. If this is the case, then their introduction will increase CO_2 fixation while complementing the role of photorespiration in protection from light excess.

It is anticipated that altering crucial pathways in plant metabolism, including light conversion and carbon fixation, will inherently have interdependent effects that require additional investigation. Furthermore, there exists the possibility of enhancing the efficiency of converting light energy into ATP and NADPH while concurrently optimizing the utilization of these molecules for carbon fixation. Such combined improvements could potentially yield an additive outcome, maximizing the overall impact on productivity.

LEVERAGING NATURALLY OCCURRING VARIATION OF PHOTOSYNTHETIC PARAMETERS

Exploiting natural variation of Rubisco kinetic traits, Γ^*, and photorespiration in coordination with CO_2 diffusion

Leveraging the inherent diversity in essential photosynthetic traits presents two notable advantages: first, it provides engineering with essential knowledge about potential trade-offs, and second, it facilitates the breeding process, allowing accelerated improvements in crops. The crucial traits related to photorespiration in Rubisco include the Michaelis-Menten constant for CO_2 (K_c), the Michaelis-Menten constant for O_2 (K_o), Rubisco specificity to CO_2 over O_2 ($S_{c/o}$), and maximum turnover rates (specific activity per active center per mass or protein [s^{-1}]) for carboxylase (V_c) and oxygenase (V_o). Several authors have investigated the environmental and evolutionary trends of natural variation in Rubisco's kinetic properties and their temperature responses, including several recent meta-analyses (Galmés et al., 2019; Bouvier et al., 2021; Tcherkez and Farquhar, 2021). These studies have revealed substantial variability in major kinetic traits, which are highly sensitive to factors such as temperature, CO_2 availability, and photosynthetically active quantum flux density. Moreover, this variability is inherent among species adapted to different environments. There are several key trade-offs among Rubisco kinetic traits, most notably the reverse relationships between $S_{c/o}$ and V_c (Galmés et al., 2014, 2019), which are important to consider in Rubisco engineering.

Recent findings have demonstrated a strong co-regulation between the natural diversity of Rubisco kinetics and g_m as well as in underlying anatomical traits. For instance, plants adapted to drought exhibit reduced g_m and chloroplast CO_2 concentration (C_c) because of higher mesophyll cell wall thickness. Consequently, these plants possess Rubisco with higher $S_{c/o}$ but lower turnover rates (Galmés et al., 2019). This highlights the substantial influence of C_c associated with species adaptation on the diversity of Rubisco's kinetic properties and the trade-offs observed. Importantly, this regulation is significantly influenced by g_m, which exhibits considerable variation among species and is responsive to environmental stressors (Elferjani et al., 2021; Knauer et al., 2022).

It is noteworthy that g_m can limit photosynthesis, accounting for approximately 10%–70% of the limitation, thereby exerting an equally significant impact on photosynthetic assimilation (A) as stomatal conductance (g_s) does (Knauer et al., 2020).

The CO_2 compensation point (Γ) has traditionally served as a parameter reflecting Rubisco functionality and photosynthetic efficiency. It represents the equilibrium between the A and leaf respiration, making it a useful characteristic for categorizing crops and herbaceous species based on their inherent photosynthetic efficiency and stress resilience. In C_3 plants, Γ is typically

highest (40–100 μmol mol^{-1}), intermediate in C_3–C_4 intermediates (20–30 μmol mol^{-1}), and low in C_4 plants (3–10 μmol mol^{-1}) (Nobel, 1991; Schlüter et al., 2023).

However, Γ is a complex trait and can be estimated in a multitude of ways. To incorporate Rubisco kinetics, photorespiration, and more accurately, an alternative parameter called the photorespiratory CO_2 compensation point in absence of day respiration (Γ^*) has been proposed. Γ^* can be determined from A/C_c curves by using the common interception method measured at different light intensities and represents the C_c at which photosynthetic carbon uptake equals photorespiratory CO_2 release (Walker et al., 2016b). Γ^* and Γ differ significantly; the magnitude of difference depends on the variable g_m (Figure 2) and the fate of the mitochondrial CO_2 fluxes because both estimates affect the resulting C_c and Γ^*. An understanding of partitioning of respiratory fluxes between the proportion directed into the chloroplast and what diffuses into the cytoplasm and IAS is also crucial for correct g_m estimations. However, the fate of respiratory fluxes and g_m largely depend on similar anatomical traits, such as the proportion of mesophyll cell surface area lined with chloroplasts, and physical dimensions of the cell wall, chloroplasts, and cytoplasm (Figure 2B). All of these traits vary significantly across species, indicating substantial variability in partitioning of respiratory fluxes between chloroplasts and cytoplasm (Evans et al., 1994; Tosens et al., 2012; Ubierna et al., 2019). Thus, considering the physiological nature of CO_2 compensation points estimated through different methods, including Γ^*, offers valuable insights into how mitochondrial CO_2 effluxes curb C_c (Walker et al., 2016b; Busch, 2020; Sage, 2022).

Achieving optimal photosynthetic efficiency requires precise coordination between traits that regulate CO_2 diffusion efficiency and the functionality of Rubisco. Consequently, a major focus in improving photosynthesis, whether through breeding or engineering, is to explore and comprehend the natural diversity of key kinetic traits of Rubisco and their relationship with CO_2 availability in the chloroplast stroma (Walker et al., 2016b; Galmés et al., 2019; Flexas and Carriquí, 2020; Evans, 2021; Knauer et al., 2022; Iqbal et al., 2023). In angiosperms with elevated rates of turnover, the constraints of photosynthesis, including stomata, g_m, and photo/biochemical processes, typically coexist in a harmonious equilibrium unless stress-induced alterations disrupt this balance. Consequently, attaining optimal enhancements of photosynthesis and resource utilization efficiency requires the simultaneous manipulation of all three constraints or a shift in focus toward augmenting the g_m/g_s ratio. This approach enables a simultaneous increase in intrinsic water use efficiency (Gago et al., 2019; Flexas and Carriquí, 2020; Knauer et al., 2020; Clarke et al., 2022; Kromdijk and McCormick, 2022).

Previous studies have often addressed the kinetics components of A/C_i curves (Rubisco kinetics, RuBP turnover rate, electron transport limitations; R_d, respiration because of photorespiration [R_p], CO_2 compensation point) separately, combining *in vivo* and *in vitro* estimations, resulting in fragmented information (Bernacchi et al., 2013). However, employing state-of-the-art fast response gas-exchange and optical diagnostic systems, as described by Laisk et al. (2002), allows simultaneous measurement of all necessary parameters to comprehensively assess the photosynthetic apparatus in leaves. This approach enables a holistic understanding of how photosynthesis is optimized. A primer on measuring photorespiration is given in Supplemental File S1.

Natural genetic variation in photosynthetic parameters as a basis for breeding

One way to adapt plants to human requirements is by harnessing the natural genetic variation that has been generated through random mutations over historical time spans. This genetic variation can occur at the intraspecific and interspecific level.

The first evidence of natural genetic variation in the context of photorespiration was provided by Jordan and Ogren (1981). They observed differences in the specificity factors toward the substrates CO_2 and O_2, $S_{c/o}$, of Rubiscos purified from seven different species, ranging from 77–82. Subsequent studies, as reviewed by Hartman and Harpel (1994), revealed changes in the specificity factor because of replacement of the active-site metal, random and site-directed mutagenesis, chemical modification, and hybridization of heterologous subunits. The highest reported specificity factor for Rubisco is 238, found in the red alga *Galdieria partita* (Uemura et al., 1996), which is about three times higher than that reported for Rubisco from most crop plants (Parry et al., 1989).

Even before discovery of the oxygenase activity of Rubisco (Bowes et al., 1971), selection experiments aiming to manipulate the specificity factor of Rubisco through selection were conducted (cf. Cannell et al., 1969; Menz et al., 1969). In these experiments, plants were kept at or slightly above the compensation point, and plants with high rates of photorespiration were expected to perish.

As discussed in the previous section, over the recent years, the number of studies exploring the natural plasticity and inherent variability of Rubisco's key catalytic traits have increased significantly. Information about the interspecific variation in photorespiration has also been observed through carbon isotope fractionation (Lanigan et al., 2008). However, to the best of our knowledge, only the study by Cai et al. (2014) examined and observed natural genetic variation in CO_2 compensation points and evaluated these differences in the context of leaf anatomy variations among three *Rhododendron* species.

The genes or alleles responsible for the advantageous photorespiratory phenotype can be transferred through interspecific hybridization, protoplast fusion, transformation, or genome editing. To facilitate these processes, it is necessary to identify the underlying genes or alleles. One approach to achieve this is through large-scale comparative genomics, which incorporates phenotypic information and is referred to as phylogenetic association mapping (e.g., Collins and Didelot, 2018).

However, with the exception of Schlüter et al. (2023), no earlier study has considered natural variation in characters related to photorespiration in a densely sampled phylogenetic tree, making application of phylogenetic association mapping challenging for identifying the underlying genomic features.

An alternative approach for identifying the genomic features responsible for interspecific differences is utilization of segregating genetic material derived from interspecific hybrids.

Comparative transcriptomics is another approach that can aid in identifying the *cis* and *trans* factors responsible for interspecific differences. By comparing transcriptomes of accessions adapted to specific conditions or subjected to environmental perturbations, strategies for crop improvement can be guided. For example, in rice, RNA sequencing (RNA-seq) identified a transcription factor with the potential to enhance photosynthetic capacity and increase yield (Wei et al., 2022). However, when using bulk RNA-seq of an organ, it becomes challenging to detect differentially expressed genes specific to rare cell types because their expression is diluted. In the case of C_3–C_4 or C_4 photosynthesis, where the pathways operate in specific cell types of the leaf, obtaining cell-type-specific transcriptomes is crucial. Approaches such as mechanical separation of cell types (John et al., 2014) and laser capture microdissection followed by microarray/RNA seq have been used previously, but their scale and resolution are limited by the speed of sampling, and separating low-abundance cell types can still be challenging (Zhang et al., 2007; Aubry et al., 2014; Hua et al., 2021; Xiong et al., 2021).

The adoption of single-cell methods, particularly droplet-based technology, is gaining increasing acceptance in plant research. This approach allows capture of individual transcriptomic profiles of cells from a wide range of plant species, providing valuable insights at a remarkably low cost per cell. For example, comparative single-cell analysis of roots from three grass species provided significant insights into the evolution of cellular divergence in these crops (Guillotin et al., 2023). This approach therefore must hold potential for better understanding the compartmentation of gene expression in C_3–C_4 and C_4 species (Cuperus, 2021; Seyfferth et al., 2021). Currently the number of single-cell or single-nucleus datasets generated from leaf tissues is limited, and most studies have tended to focus on a single model species under one condition (Bezrutczyk et al., 2021; Kim et al., 2021; Lopez-Anido et al., 2021; Berrío et al., 2022; Procko et al., 2022; Sun et al., 2022). To enhance photosynthesis efficacy in crops, understanding the changes in gene expression associated with closely related C_3, C_4, or/and C_3–C_4 intermediate species is likely to provide insights into the molecular signatures of each of these traits and therefore how they might be rationally engineered. Comparative analyses of leaf anatomy, cellular ultrastructure, and photosynthetic traits between species within a genus (for example, in *Gynandropsis* [Marshall et al., 2007; Koteyeva et al., 2011], *Moricandia* [Schlüter et al., 2017], or *Flaveria* [McKown and Dengler, 2007; Kümpers et al., 2017]) have provided insights into traits that engender higher photosynthesis efficiency. Therefore, acquiring transcriptome data from individual bundle sheath and mesophyll cells of these closely related plants should provide new insights into how the patterns of transcript abundance alter in association with modifications to photosynthetic efficiency. We also anticipate that single-cell RNA-seq (scRNA-seq) will contribute to crop improvement by providing insights into underlying molecular mechanisms. Last, additional advantages can be obtained from scRNA-seq when the data are associated with a complex phenotype (single-cell transcriptome-wide association studies) and genotype (cell-type-specific expression quantitative trait loci) at the population level (Perez et al., 2022). It seems likely that such approaches will be adopted for high-resolution genotyping of bioengineered plants in synthetic biology projects. In summary, combination of scRNA-seq technology with comparative transcriptomics, population genetics, and synthetic biology is likely to serve as a powerful tool for crop improvement.

A technically simpler procedure to the above-described approach of exploiting interspecific natural genetic variation would be exploitation of the intraspecific variation within the species under consideration for photosynthetic properties because transfer or the enrichment of positive alleles is much easier to realize in comparison with interspecific variability. Within species, natural variation in leaf photosynthesis has been reported for model species (e.g., Tomeo and Rosenthal, 2018) but also for major crops (e.g., Gu et al., 2012; Driever et al., 2014). However, the improvements that can be realized in that way are up to now smaller compared with the approaches exploiting interspecific variability.

NEW-TO-NATURE APPROACHES TO IMPROVE PHOTOSYNTHETIC EFFICIENCY

Most efforts to improve carbon capture in plants have focused on engineering naturally existing enzymes and pathways (Kebeish et al., 2007; South et al., 2019). However, the emergence of synthetic biology has opened up the possibility to radically (re) draft plant metabolism to overcome the limitations of natural evolution, which is driven by co-linearity, tinkering, epistatic drift, and purifying selection, rather than by "design" (Wurtzel et al., 2019). Thus, such engineering approaches have the potential to expand the biological solution space and provide new-to-nature pathways that outcompete their natural counterparts in respect to thermodynamics and/or kinetics because they are drafted from first principles.

As an example, while nature has evolved seven different pathways for CO_2 fixation, more than 30 synthetic CO_2 fixation pathways have already been designed, which are all superior to the CBC (Bar-Even et al., 2010). Some of these new-to-nature solutions have even been realized successfully realized *in vitro* (Schwander et al., 2016; Luo et al., 2022; McLean et al., 2023) and are awaiting their transplantation *in vivo*.

The concept of designer metabolism has also been extended to photorespiration lately (Figure 3). Two studies have proposed alternative ways to use the photorespiratory metabolite glyoxylate to feed into synthetic carbon fixation cycles (Figure 3, pathways 1 and 2). The first cycle is the malyl-CoA-glycerate (MCG) pathway. In this cycle, the bacterial glyoxylate assimilation route condenses two molecules of glyoxylate to form a C3 compound, releasing CO_2 in the process. The resulting tartronate semialdehyde is then reduced and phosphorylated into 2PG. 2PG is converted to phosphoenolpyruvate and further carboxylated to oxaloacetate, subsequently reduced to malate and activated with CoA, followed by cleavage into glyoxylate and acetyl-CoA. Glyoxylate is then available to initiate the next cycle, while acetyl-CoA can be utilized for biosynthesis (Yu et al., 2018). The MGC pathway, therefore, does not result in a net loss of CO_2.

A second study proposed the 3-hydroxypropionate (3OHP) photorespiratory bypass, inspired by the naturally occurring 3OHP

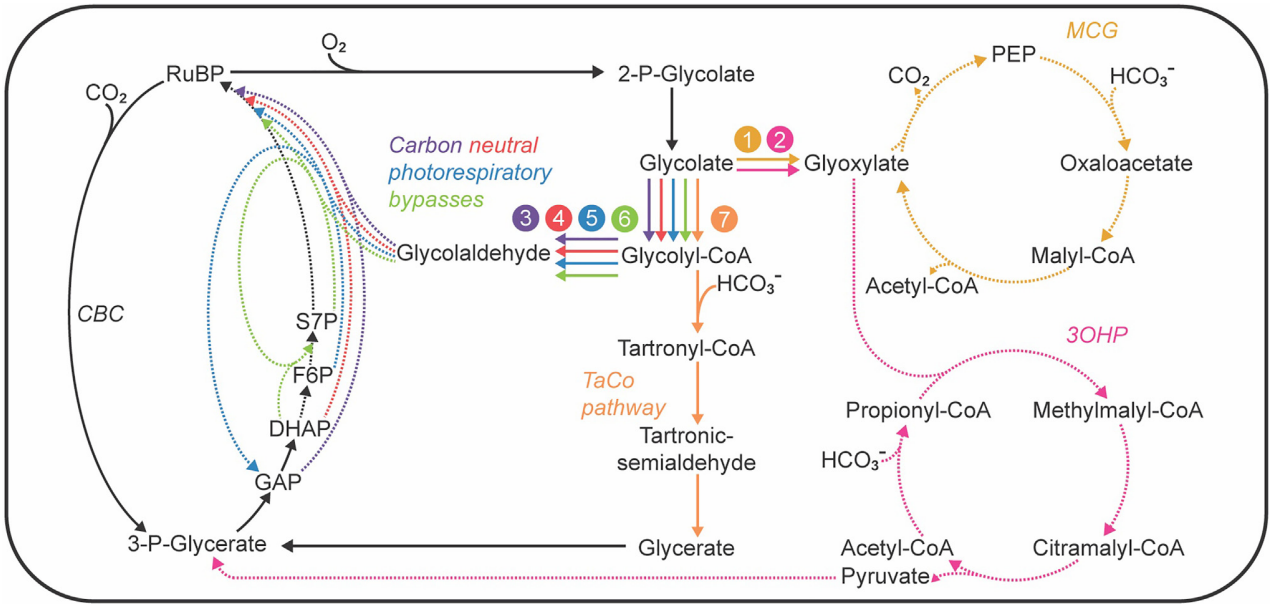


Figure 3. Designed (new-to-nature) photorespiratory bypasses.
Shown are (1) the malyl-CoA-glycerate pathway (MCG) (Yu et al., 2018), (2) 3-hydroxypropionate bypass (3OHP) (Shih et al., 2014), (3)–(6) carbon neutral bypasses (Trudeau et al., 2018), and (7) the tartronyl-CoA (TaCo) pathway (Scheffen et al., 2021).

bi-cycle (Zarzycki et al., 2009). The 3OHP bypass exploits malyl-CoA lyase to produce β-methylmalyl-CoA from glyoxylate and propionyl-CoA. Methylmalyl-CoA undergoes a series of interconversions that yield citramaly-CoA, which, in turn, serves as a substrate for malyl-CoA lyase and produces pyruvate and acetyl-CoA. Acetyl-CoA is carboxylated to malonyl-CoA, which is then reduced and further activated into the starter compound of the cycle: propionyl-CoA. Pyruvate, on the other hand, can be further converted to phosphoglycerate and re-enter the CBC (Shih et al., 2014), resulting in a net gain in carbon.

Both studies implemented their proposed cycles in cyanobacteria. Regardless of the CO_2 fixing ability of 3OHP bypass, strains carrying the cycle did not present a conclusive phenotype. However, the authors set an example for future *in vivo* implementation of carbon-capturing photorespiratory bypasses. On the other hand, strains containing the MCG pathway showed increased bicarbonate assimilation and acetyl-CoA accumulation and achieved higher optical densities than strains lacking MCG.

The aforementioned cycles represent what has been described in literature as "mix and match" synthetic pathways because they described novel pathways relying on known reactions and enzymes (Erb et al., 2017). Nevertheless, a recent study has identified several new-to-nature pathways by systematically developing reaction sequences from the pool of feasible biochemical transformations that could convert (phospho)glycolate back into a central intermediate of the CBC cycle (Trudeau et al., 2018). The design of these solutions was guided by two additional principles. First, the new reaction sequences should require as little energy as possible, and second, these pathways should be CO_2 neutral (i.e., not release CO_2) or even capture CO_2.

Through these efforts, the authors proposed four "carbon-neutral pathways" to reincorporate the C2 compound product of the oxygenation reaction of Rubisco into the CBC (Figure 3, pathways 3–6). These pathways rely on a novel reduction of glycolate to glycolaldehyde via engineered enzymes. Taking advantage of the high reactivity of glycolaldehyde, it is further combined with sugar phosphates present on the CBC, such as glyceraldehyde 3-phosphate (GAP), dihydroxyacetone phosphate, fructose 6-phosphate, or sedoheptulose 7-phosphate, via aldolase, transketolase, or transaldolase reactions. In all cases, a C5 compound is produced and further converted into the substrate of Rubisco. In the case of transketolase and transaldolase reactions, side products are formed that are also part of the CBC and, therefore, can be reused directly. Trudeau et al. (2018) proved the *in vitro* feasibility of one of their proposed carbon neutral bypasses; however, to date, no further improvement or *in vivo* implementation has been reported.

Furthermore, in the same study, another photorespiration bypass comprising novel reactions was proposed, the tartonyl-CoA (TaCo) pathway (Figure 3, pathway 7). The TaCo pathway is a five-reaction sequence that first converts photorespiratory glycolate into glycolyl-CoA, which is subsequently carboxylated into tartronyl-CoA, the namesake compound of the pathway. In two subsequent steps, tartronyl-CoA is then reduced to glycerate, which can re-enter the CBC at the level of phosphoglycerate. Compared with natural photorespiration, the TaCo pathway sequence is about 50% shorter, requires about 20% less ATP and 30% less reducing power, and does not release ammonia or, notably, CO_2 but instead captures additional CO_2 during photorespiration. In other words, through the TaCo pathway, photorespiration can be turned into a carbon-capturing process in which the oxygenation reaction of Rubisco will still lead to subsequent fixation of carbon.

Although this pathway outcompetes natural photorespiration, a challenge has been that the central TaCo pathway sequence relies on enzyme reactions that have not been described so far. This

apparent challenge was addressed through re-engineering the active sites of enzyme candidates that catalyze similar reactions to establish and/or improve the desired reactions (Scheffen et al., 2021). The full pathway was reconstituted and prototyped *in vitro*.

However, this proof of principle is only the starting point for further developments to successfully realize the TaCo pathway in the context of the living plant. This could be achieved through a combination of complementary approaches. The *in vitro* prototyping efforts could be expanded by combining high-throughput combinatorics with machine-learning-guided experimentation to optimize the enzyme stoichiometry and robustness of the pathway (Pandi et al., 2022; Vögeli et al., 2022). The optimized network could then be (partially) transplanted in a suitable (micro)organism to test its feasibility and use adaptive laboratory evolution to further improve the network and its cellular integration. The latter strategy has been recently used successfully to establish the CBC, the reverse glycine cleavage pathway, or a modified serine cycle in *E. coli* (Antonovsky et al., 2016; Yishai et al., 2017; Gleizer et al., 2019; Luo et al., 2022), in which selection strains were designed, which need to form (part) of their biomass through the new pathways. Having established a functional cycle inside of a microorganisms will guide further efforts to integrate this new reaction sequence in plants.

CONCLUDING REMARKS AND PERSPECTIVES

Sustainably increasing crop yields is crucial to meet the growing demands of a rising global population for food, feed, and other plant-derived products. Maximizing photosynthetic efficiency is hence an important factor for food security in the context of anthropogenic climate change and resource limitations. In the quest to increase crop productivity through improvement of photosynthetic efficiency, bypasses to canonical photorespiration and modifications of excess light protection have been designed, implemented in several model species and in crops, and tested in the field. Jointly, these data provide a proof of concept for yield increases through engineering of photorespiration and light harvesting.

While progress has been made, there are still unanswered questions that need to be addressed, essential to fully gather the potential improvements achievable from these approaches. For example, how do photorespiratory bypasses actually work? What are the mechanisms that underpin the observed yield increases? Resolving questions such as these will require additional experiments and comprehensive analysis by mathematical models of plant metabolism.

Available data suggest that combining photorespiratory bypasses with optimized energy dissipation holds promise to maximize the energy available for CO_2 assimilation. New-to-nature pathways show potential to exceed the yield improvements that have been achieved by current photorespiratory bypasses. To date, these pathways have been prototyped *in vitro* and in bacteria, and they are awaiting testing in plants.

Furthermore, natural variation in photosynthetic efficiency exists in domesticated crops and their wild relatives and so provides a valuable resource for breeding strategies. Techniques such as intra- and interspecific hybridization and genome editing could leverage promising alleles identified through pan-genomic association mapping and systems biology approaches to introduce beneficial genetic variants into crops. Knowledge of the natural variation in photosynthetic parameters should be combined with synthetic photorespiratory bypasses for synergistic effects. For example, alternative forms of Rubisco with higher V_c at the cost of lower V_c/V_o could be combined with more energy-efficient or CO_2-concentrating photorespiratory bypasses, resulting in additive benefits. Also, it should be possible to combine more performant Rubisco variants with better sourcing of CO_2 via optimization of g_m. The CO_2 compensation point serves as a key parameter when screening for photosynthetic efficiency; multiple methods have been employed for its determination, but its precision relies on accurate estimation of cellular CO_2 efflux and influx.

The efficacy of increasing photosynthetic efficiency in crops in the field has been clearly demonstrated, but further improvements, or reliable transfer of traits from model organisms to crop species or between crop species, will likely require additional modifications of sink tissues or harvestable biomass to reap the full benefits of increased photosynthetic efficiency. Only when sink strength can keep up with source capacity will it be possible to capitalize on gains in efficiency.

The practicality of implementing different strategies to improve photosynthetic efficiency that range from selective breeding to new-to-nature pathways depends on government regulation of genetic modification, which must be also considered when applying scientific discoveries to solve real-world problems.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

We appreciate funding by the European Union H2020 Program (project GAIN4CROPS, GA no. 862087, to B.S., G.F., G.C, D.T., T.M., T.J.E., A.P.M.W., M.H., E.N.S., O.E., J.M.H., and T.T.) and the Deutsche Forschungsgemeinschaft (Cluster of Excellence for Plant Sciences [CEPLAS] under Germany's Excellence Strategy EXC-2048/1 under project ID 390686111 to B.S., O.E., and A.P.M.W. and CRC TRR 341 "Plant Ecological Genetics" to B.S. and A.P.M.W.).

ACKNOWLEDGMENTS

No conflict of interest is declared.

Received: July 3, 2023
Revised: August 20, 2023
Accepted: August 31, 2023
Published: September 1, 2023

REFERENCES

Abadie, C., and Tcherkez, G. (2019). Plant sulphur metabolism is stimulated by photorespiration. Commun. Biol. **2**:379.

Abbasi, A.Z., Bilal, M., Khurshid, G., Yiotis, C., Zeb, I., Hussain, J., Baig, A., Shah, M.M., Chaudhary, S.U., Osborne, B., and Ahmad, R. (2021). Expression of cyanobacterial genes enhanced CO2 assimilation and biomass production in transgenic *Arabidopsis thaliana*. PeerJ **9**, e11860.

Alboresi, A., Storti, M., and Morosinotto, T. (2019). Balancing protection and efficiency in the regulation of photosynthetic electron transport across plant evolution. New Phytol. **221**:105–109.

Antonovsky, N., Gleizer, S., Noor, E., Zohar, Y., Herz, E., Barenholz, U., Zelcbuch, L., Amram, S., Wides, A., Tepper, N., et al. (2016). Sugar Synthesis from CO_2 in *Escherichia coli*. Cell **166**:115–125.

Aubry, S., Kelly, S., Kümpers, B.M.C., Smith-Unna, R.D., and Hibberd, J.M. (2014). Deep Evolutionary Comparison of Gene Expression Identifies Parallel Recruitment of Trans-Factors in Two Independent Origins of C_4 Photosynthesis. PLoS Genet. **10**, e1004365.

Bar-Even, A., Noor, E., Lewis, N.E., and Milo, R. (2010). Design and analysis of synthetic carbon fixation pathways. Proc. Natl. Acad. Sci. USA **107**:8889–8894.

Basler, G., Küken, A., Fernie, A.R., and Nikoloski, Z. (2016). Photorespiratory Bypasses Lead to Increased Growth in Arabidopsis thaliana: Are Predictions Consistent with Experimental Evidence? Front. Bioeng. Biotechnol. **4**:31.

Bauwe, H. (2023). Photorespiration – Rubisco's repair crew. J. Plant Physiol. **280**, 153899.

Bauwe, H., Hagemann, M., and Fernie, A.R. (2010). Photorespiration: players, partners and origin. Trends Plant Sci. **15**:330–336.

Bernacchi, C.J., Bagley, J.E., Serbin, S.P., Ruiz-Vera, U.M., Rosenthal, D.M., and Vanloocke, A. (2013). Modelling C_3 photosynthesis from the chloroplast to the ecosystem. Plant Cell Environ. **36**:1641–1657.

Berrío, R.T., Verstaen, K., Vandamme, N., Pevernagie, J., Achon, I., Van Duyse, J., Van Isterdael, G., Saeys, Y., De Veylder, L., Inzé, D., and Dubois, M. (2022). Single-cell transcriptomics sheds light on the identity and metabolism of developing leaf cells. Plant Physiol. **188**:898–918.

Bezrutczyk, M., Zöllner, N.R., Kruse, C.P.S., Hartwig, T., Lautwein, T., Köhrer, K., Frommer, W.B., and Kim, J.-Y. (2021). Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. Plant Cell **33**:531–547.

Bilal, M., Abbasi, A.Z., Khurshid, G., Yiotis, C., Hussain, J., Shah, M.M., Naqvi, T., Kwon, S.-Y., Park, Y.-I., Osborne, B., and Ahmad, R. (2019). The expression of cyanobacterial glycolate–decarboxylation pathway genes improves biomass accumulation in *Arabidopsis thaliana*. Plant Biotechnol. Rep. **13**:361–373.

Bouvier, J.W., Emms, D.M., Rhodes, T., Bolton, J.S., Brasnett, A., Eddershaw, A., Nielsen, J.R., Unitt, A., Whitney, S.M., and Kelly, S. (2021). Rubisco Adaptation Is More Limited by Phylogenetic Constraint Than by Catalytic Trade-off. Mol. Biol. Evol. **38**:2880–2896.

Bowes, G., Ogren, W.L., and Hageman, R.H. (1971). Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. Biochem. Biophys. Res. Commun. **45**:716–722.

Broncano, L.S., Pukacz, K.R., Reichel-Deland, V., Schlüter, U., Triesch, S., and Weber, A.P. (2023). Photorespiration is the solution, not the problem. J. Plant Physiol. **282**, 153928.

Busch, F.A. (2020). Photorespiration in the context of Rubisco biochemistry, CO_2 diffusion and metabolism. Plant J. **101**:919–939.

Busch, F.A., Sage, T.L., Cousins, A.B., and Sage, R.F. (2013). C_3 plants enhance rates of photosynthesis by reassimilating photorespired and respired CO_2. Plant Cell Environ. **36**:200–212.

Cai, Y.-F., Li, S.-F., Li, S.-F., Xie, W.-J., and Song, J. (2014). How do leaf anatomies and photosynthesis of three Rhododendron species relate to their natural environments? Bot. Stud. **55**:36.

Campbell, W.J., and Ogren, W.L. (1990). A novel role for light in the activation of ribulosebisphosphate carboxylase/oxygenase. Plant Physiol. **92**:110–115.

Cannell, R.Q., Brun, W.A., and Moss, D.N. (1969). A Search for High Net Photosynthetic Rate among Soybean Genotypes1. Crop Sci. **9**:840–841.

Carvalho, J.d.F.C., Madgwick, P.J., Powers, S.J., Keys, A.J., Lea, P.J., and Parry, M.A.J. (2011). An engineered pathway for glyoxylate metabolism in tobacco plants aimed to avoid the release of ammonia in photorespiration. BMC Biotechnol. **11**:111.

Clarke, V.C., De Rosa, A., Massey, B., George, A.M., Evans, J.R., von Caemmerer, S., and Groszmann, M. (2022). Mesophyll conductance is unaffected by expression of Arabidopsis PIP1 aquaporins in the plasmalemma of Nicotiana. J. Exp. Bot. **73**:3625–3636.

Collins, C., and Didelot, X. (2018). A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLoS Comput. Biol. **14**, e1005958.

Cuperus, J.T. (2021). Single-cell genomics in plants: current state, future directions, and hurdles to overcome. Plant Physiol. **188**:749–755.

Cutolo, E.A., Guardini, Z., Dall'Osto, L., and Bassi, R. (2023). A paler shade of green: engineering cellular chlorophyll content to enhance photosynthesis in crowded environments. New Phytol. **239**:1567–1583. https://doi.org/10.1111/nph.19064.

Dalal, J., Lopez, H., Vasani, N.B., Hu, Z., Swift, J.E., Yalamanchili, R., Dvora, M., Lin, X., Xie, D., Qu, R., and Sederoff, H.W. (2015). A photorespiratory bypass increases plant growth and seed yield in biofuel crop *Camelina sativa*. Biotechnol. Biofuels **8**:175.

Driever, S.M., Lawson, T., Andralojc, P.J., Raines, C.A., and Parry, M.A.J. (2014). Natural variation in photosynthetic capacity, growth, and yield in 64 field-grown wheat genotypes. J. Exp. Bot. **65**:4959–4973.

Eisenhut, M., Ruth, W., Haimovich, M., Bauwe, H., Kaplan, A., and Hagemann, M. (2008). The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. Proc. Natl. Acad. Sci. USA **105**:17199–17204.

Eisenhut, M., Roell, M.S., and Weber, A.P.M. (2019). Mechanistic understanding of photorespiration paves the way to a new green revolution. New Phytol. **223**:1762–1769.

Elferjani, R., Benomar, L., Momayyezi, M., Tognetti, R., Niinemets, Ü., Soolanayakanahally, R.Y., Théroux-Rancourt, G., Tosens, T., Ripullone, F., Bilodeau-Gauthier, S., et al. (2021). A meta-analysis of mesophyll conductance to CO_2 in relation to major abiotic stresses in poplar species. J. Exp. Bot **72**:erab127.

Erb, T.J., Jones, P.R., and Bar-Even, A. (2017). Synthetic metabolism: metabolic engineering meets enzyme design. Curr. Opin. Chem. Biol. **37**:56–62.

Evans, J.R. (2021). Mesophyll conductance: walls, membranes and spatial complexity. New Phytol. **229**:1864–1876.

Evans, J., Caemmerer, S., Setchell, B., and Hudson, G. (1994). The relationship between CO_2 transfer conductance and leaf anatomy in transgenic tobacco with a reduced content of rubisco. Funct. Plant Biol. **21**:475–495.

Fernie, A.R., and Bauwe, H. (2020). Wasteful, essential, evolutionary stepping stone? The multiple personalities of the photorespiratory pathway. Plant J. **102**:666–677.

Flexas, J., Barbour, M.M., Brendel, O., Cabrera, H.M., Carriquí, M., Díaz-Espejo, A., Douthe, C., Dreyer, E., Ferrio, J.P., Gago, J., et al. (2012). Mesophyll diffusion conductance to CO_2: An unappreciated central player in photosynthesis. Plant Sci. **193–194**:70–84.

Flexas, J., Niinemets, U., Gallé, A., Barbour, M.M., Centritto, M., Diaz-Espejo, A., Douthe, C., Galmés, J., Ribas-Carbo, M., Rodriguez, P.L., et al. (2013). Diffusional conductances to CO_2 as a target for increasing photosynthesis and photosynthetic water-use efficiency. Photosynth. Res. **117**:45–59.

Flexas, J., and Carriquí, M. (2020). Photosynthesis and photosynthetic efficiencies along the terrestrial plant's phylogeny: lessons for improving crop photosynthesis. Plant J. **101**:964–978.

Flügel, F., Timm, S., Arrivault, S., Florian, A., Stitt, M., Fernie, A.R., and Bauwe, H. (2017). The Photorespiratory Metabolite 2-Phosphoglycolate Regulates Photosynthesis and Starch Accumulation in Arabidopsis. Plant Cell **29**:2537–2551.

Fu, X., Gregory, L.M., Weise, S.E., and Walker, B.J. (2023a). Integrated flux and pool size analysis in plant central metabolism reveals unique roles of glycine and serine during photorespiration. Nat. Plants **9**:169–178.

Fu, X., Smith, K., Gregory, L., Roze, L., and Walker, B. (2023b). Understanding and improving crop photosynthesis. Burleigh Dodds Ser. Agric. Sci. 203–222. https://doi.org/10.19103/as.2022.0119.12.

Gago, J., Carriquí, M., Nadal, M., Clemente-Moreno, M.J., Coopman, R.E., Fernie, A.R., and Flexas, J. (2019). Photosynthesis Optimized across Land Plant Phylogeny. Trends Plant Sci. **24**:947–958.

Galmés, J., Capó-Bauçà, S., Niinemets, Ü., and Iñiguez, C. (2019). Potential improvement of photosynthetic CO_2 assimilation in crops by exploiting the natural variation in the temperature response of Rubisco catalytic traits. Curr. Opin. Plant Biol. **49**:60–67.

Galmés, J., Kapralov, M.V., Andralojc, P.J., Andralojc, P.J., Conesa, M.À., Keys, A.J., Parry, M.A., and Flexas, J. (2014). Expanding knowledge of the Rubisco kinetics variability in plant species: environmental and evolutionary trends. Plant Cell Environ **37**:1989–2001.

Garcia-Molina, A., and Leister, D. (2020). Accelerated relaxation of photoprotection impairs biomass accumulation in Arabidopsis. Nat. Plants **6**:9–12.

Gleizer, S., Ben-Nissan, R., Bar-On, Y.M., Antonovsky, N., Noor, E., Zohar, Y., Jona, G., Krieger, E., Shamshoum, M., Bar-Even, A., and Milo, R. (2019). Conversion of *Escherichia coli* to Generate All Biomass Carbon from CO_2. Cell **179**:1255–1263.e12.

Gu, J., Yin, X., Stomph, T.-J., Wang, H., and Struik, P.C. (2012). Physiological basis of genetic variation in leaf photosynthesis among rice (*Oryza sativa* L.) introgression lines under drought and well-watered conditions. J. Exp. Bot. **63**:5137–5153.

Guillotin, B., Rahni, R., Passalacqua, M., Mohammed, M.A., Xu, X., Raju, S.K., Ramírez, C.O., Jackson, D., Groen, S.C., Gillis, J., and Birnbaum, K.D. (2023). A pan-grass transcriptome reveals patterns of cellular divergence in crops. Nature **617**:785–791.

Hanawa, H., Ishizaki, K., Nohira, K., Takagi, D., Shimakawa, G., Sejima, T., Shaku, K., Makino, A., and Miyake, C. (2017). Land plants drive photorespiration as higher electron-sink: Comparative study of post-illumination transient O_2-uptake rates from liverworts to angiosperms through ferns and gymnosperms. Physiol. Plantarum **161**:138–149.

Hartman, F.C., and Harpel, M.R. (1994). Structure, Function, Regulation, and Assembly of D-Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase. Annu. Rev. Biochem. **63**:197–234.

Heber, U., and Krause, G.H. (1980). What is the physiological role of photorespiration? Trends Biochem. Sci. **5**:32–34.

Heber, U., Bligny, R., Streb, P., and Douce, R. (1996). Photorespiration is Essential for the Protection of the Photosynthetic Apparatus of C_3 Plants Against Photoinactivation Under Sunlight. Bot. Acta **109**:307–315.

Hodges, M., Dellero, Y., Keech, O., Betti, M., Raghavendra, A.S., Sage, R., Zhu, X.-G., Allen, D.K., and Weber, A.P.M. (2016). Perspectives for a better understanding of the metabolic integration of photorespiration within a complex plant primary metabolism network. J. Exp. Bot. **67**:3015–3026.

Hua, L., Stevenson, S.R., Reyna-Llorens, I., Xiong, H., Kopriva, S., and Hibberd, J.M. (2021). The bundle sheath of rice is conditioned to play an active role in water transport as well as sulfur assimilation and jasmonic acid synthesis. Plant J. **107**:268–286.

Iqbal, W.A., Lisitsa, A., and Kapralov, M.V. (2023). Predicting plant Rubisco kinetics from RbcL sequence data using machine learning. J. Exp. Bot. **74**:638–650.

John, C.R., Smith-Unna, R.D., Woodfield, H., Covshoff, S., and Hibberd, J.M. (2014). Evolutionary Convergence of Cell-Specific Gene Expression in Independent Lineages of C_4 Grasses. Plant Physiol. **165**:62–75.

Jordan, D.B., and Ogren, W.L. (1981). Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. Nature **291**:513–515.

Kebeish, R., Niessen, M., Thiruveedhi, K., Bari, R., Hirsch, H.-J., Rosenkranz, R., Stäbler, N., Schönfeld, B., Kreuzaler, F., and Peterhänsel, C. (2007). Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. Nat. Biotechnol. **25**:593–599.

Khurshid, G., Abbassi, A.Z., Khalid, M.F., Gondal, M.N., Naqvi, T.A., Shah, M.M., Chaudhary, S.U., and Ahmad, R. (2020). A cyanobacterial photorespiratory bypass model to enhance photosynthesis by rerouting photorespiratory pathway in C_3 plants. Sci. Rep. **10**, 20879.

Kim, J.-Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezrutczyk, M., Miras, M., Zöllner, N., Hartwig, T., Wudick, M.M., et al. (2021). Distinct identities of leaf phloem cells revealed by single cell transcriptomics. Plant Cell **33**:511–530.

Knauer, J., Zaehle, S., De Kauwe, M.G., Haverd, V., Reichstein, M., and Sun, Y. (2020). Mesophyll conductance in land surface models: effects on photosynthesis and transpiration. Plant J. **101**:858–873.

Knauer, J., Cuntz, M., Evans, J.R., Niinemets, Ü., Tosens, T., Veromann-Jürgenson, L.L., Werner, C., and Zaehle, S. (2022). Contrasting anatomical and biochemical controls on mesophyll conductance across plant functional types. New Phytol. **236**:357–368.

Koteyeva, N.K., Voznesenskaya, E.V., Roalson, E.H., and Edwards, G.E. (2011). Diversity in forms of *C4* in the genus Cleome (Cleomaceae). Ann. Bot. **107**:269–283.

Kozaki, A., and Takeba, G. (1996). Photorespiration protects C_3 plants from photooxidation. Nature **384**:557–560.

Kromdijk, J., and McCormick, A.J. (2022). Genetic variation in photosynthesis: many variants make light work. J. Exp. Bot. **73**:3053–3056.

Kromdijk, J., Głowacka, K., Leonelli, L., Gabilly, S.T., Iwai, M., Niyogi, K.K., and Long, S.P. (2016). Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. Science **354**:857–861.

Kümpers, B.M.C., Burgess, S.J., Reyna-Llorens, I., Smith-Unna, R., Boursnell, C., and Hibberd, J.M. (2017). Shared characteristics underpinning C_4 leaf maturation derived from analysis of multiple C_3 and C_4 species of Flaveria. J. Exp. Bot. **68**:177–189.

Laisk, A., Oja, V., Rasulov, B., Rämma, H., Eichelmann, H., Kasparova, I., Pettai, H., Padu, E., and Vapaavuori, E. (2002). A computer-operated routine of gas exchange and optical measurements to diagnose photosynthetic apparatus in leaves: Diagnosing photosynthesis. Plant Cell Environ. **25**:923–943.

Lanigan, G.J., Betson, N., Griffiths, H., and Seibt, U. (2008). Carbon Isotope Fractionation during Photorespiration and Carboxylation in Senecio. Plant Physiol. **148**:2013–2020.

Levey, M., Timm, S., Mettler-Altmann, T., Luca Borghi, G., Koczor, M., Arrivault, S., Pm Weber, A., Bauwe, H., Gowik, U., and Westhoff, P. (2019). Efficient 2-phosphoglycolate degradation is required to maintain carbon assimilation and allocation in the C_4 plant. J. Exp. Bot. **70**:575–587.

Li, Z., Wakao, S., Fischer, B.B., and Niyogi, K.K. (2009). Sensing and Responding to Excess Light. Annu. Rev. Plant Biol. **60**:239–260.

Long, S.P., Taylor, S.H., Burgess, S.J., Carmo-Silva, E., Lawson, T., De Souza, A.P., Leonelli, L., and Wang, Y. (2022). Into the Shadows and Back into Sunlight: Photosynthesis in Fluctuating Light. Annu. Rev. Plant Biol. **73**:617–648.

Lopez-Anido, C.B., Vatén, A., Smoot, N.K., Sharma, N., Guo, V., Gong, Y., Anleu Gil, M.X., Weimer, A.K., and Bergmann, D.C. (2021). Single-cell resolution of lineage trajectories in the Arabidopsis stomatal lineage and developing leaf. Dev. Cell **56**:1043–1055.e4.

Luo, S., Lin, P.P., Nieh, L.-Y., Liao, G.-B., Tang, P.-W., Chen, C., and Liao, J.C. (2022). A cell-free self-replenishing CO_2-fixing system. Nat. Catal. **5**:154–162.

Maier, A., Fahnenstich, H., von Caemmerer, S., Engqvist, M.K.M., Weber, A.P.M., Flügge, U.I., and Maurino, V.G. (2012). Transgenic Introduction of a Glycolate Oxidative Cycle into *A. thaliana* Chloroplasts Leads to Growth Improvement. Front. Plant Sci. **3**:38.

Marshall, D.M., Muhaidat, R., Brown, N.J., Liu, Z., Stanley, S., Griffiths, H., Sage, R.F., and Hibberd, J.M. (2007). Cleome, a genus closely related to Arabidopsis, contains species spanning a developmental progression from C_3 to C_4 photosynthesis. Plant J. **51**:886–896.

McKown, A.D., and Dengler, N.G. (2007). Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in Flaveria (Asteraceae). Am. J. Bot. **94**:382–399.

McLean, R., Schwander, T., Diehl, C., Cortina, N.S., Paczia, N., Zarzycki, J., and Erb, T.J. (2023). Exploring alternative pathways for the in vitro establishment of the HOPAC cycle for synthetic CO_2 fixation. Sci. Adv. **9**, eadh4299.

Menz, K.M., Moss, D.N., Cannell, R.Q., and Brun, W.A. (1969). Screening for Photosynthetic Efficiency1. Crop Sci. **9**:692–694.

Niinemets, Ü. (2023). Variation in leaf photosynthetic capacity within plant canopies: optimization, structural and physiological constraints and inefficiencies. Photosynth. Res. 1–19.

Nobel, P.S. (1991). Achievable productivities of certain CAM plants: basis for high values compared with C_3 and C_4 plants. New Phytol. **119**:183–205.

Nölke, G., Houdelet, M., Kreuzaler, F., Peterhänsel, C., and Schillberg, S. (2014). The expression of a recombinant glycolate dehydrogenase polyprotein in potato (Solanum tuberosum) plastids strongly enhances photosynthesis and tuber yield. Plant Biotechnol. J. **12**:734–742.

Oliver, D.J. (1980). The effect of glyoxylate on photosynthesis and photorespiration by isolated soybean mesophyll cells. Plant Physiol. **65**:888–892.

Oliver, D.J., and Zelitch, I. (1977). Increasing photosynthesis by inhibiting photorespiration with glyoxylate. Science **196**:1450–1451.

Onoda, Y., Wright, I.J., Evans, J.R., Hikosaka, K., Kitajima, K., Niinemets, Ü., Poorter, H., Tosens, T., and Westoby, M. (2017). Physiological and structural tradeoffs underlying the leaf economics spectrum. New Phytol. **214**:1447–1463.

Osmanoglu, Ö., Khaled AlSeiari, M., AlKhoori, H.A., Shams, S., Bencurova, E., Dandekar, T., and Naseem, M. (2021). Topological Analysis of the Carbon-Concentrating CETCH Cycle and a Photorespiratory Bypass Reveals Boosted CO_2-Sequestration by Plants. Front. Bioeng. Biotechnol. **9**, 708417.

Pandi, A., Diehl, C., Yazdizadeh Kharrazi, A., Scholz, S.A., Bobkova, E., Faure, L., Nattermann, M., Adam, D., Chapin, N., Foroughijabbari, Y., et al. (2022). A versatile active learning workflow for optimization of genetic and metabolic networks. Nat. Commun. **13**:3876.

Parry, M.A.J., Keys, A.J., and Gutteridge, S. (1989). Variation in the Specificity Factor of C_3 Higher Plant Rubiscos Determined by the Total Consumption of Ribulose-P 2. J. Exp. Bot. **40**:317–320.

Perez, R.K., Gordon, M.G., Subramaniam, M., Kim, M.C., Hartoularos, G.C., Targ, S., Sun, Y., Ogorodnikov, A., Bueno, R., Lu, A., et al. (2022). Single-cell RNA-seq reveals cell type–specific molecular and genetic associations to lupus. Science **376**, eabf1970.

Perin, G., Bellan, A., Michelberger, T., Lyska, D., Wakao, S., Niyogi, K.K., and Morosinotto, T. (2023). Modulation of xanthophyll cycle impacts biomass productivity in the marine microalga Nannochloropsis. Proc. Natl. Acad. Sci. USA **120**, e2214119120.

Peterhansel, C., Blume, C., and Offermann, S. (2013). Photorespiratory bypasses: how can they work? J. Exp. Bot. **64**:709–715.

Pick, T.R., Bräutigam, A., Schulz, M.A., Obata, T., Fernie, A.R., and Weber, A.P.M. (2013). PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. Proc. Natl. Acad. Sci. USA **110**:3185–3190.

Procko, C., Lee, T., Borsuk, A., Bargmann, B.O.R., Dabi, T., Nery, J.R., Estelle, M., Baird, L., O'Connor, C., Brodersen, C., et al. (2022). Leaf cell-specific and single-cell transcriptional profiling reveals a role for the palisade layer in UV light protection. Plant Cell **34**:3261–3279.

Roell, M.-S., Schada von Borzykowski, L., Westhoff, P., Plett, A., Paczia, N., Claus, P., Urte, S., Erb, T.J., and Weber, A.P.M. (2021). A synthetic C_4 shuttle via the β-hydroxyaspartate cycle in C_3 plants. Proc. Natl. Acad. Sci. USA **118**, e2022307118.

Rotasperti, L., Tadini, L., Chiara, M., Crosatti, C., Guerra, D., Tagliani, A., Forlani, S., Ezquer, I., Horner, D.S., Jahns, P., et al. (2022). The barley mutant *happy under the sun 1* (*hus1*): An additional contribution to pale green crops. Environ. Exp. Bot. **196**, 104795.

Saadat, N.P., Nies, T., van Aalst, M., Hank, B., Demirtas, B., Ebenhöh, O., and Matuszyńska, A. (2021). Computational Analysis of Alternative Photosynthetic Electron Flows Linked With Oxidative Stress. Front. Plant Sci. **12**, 750580.

Sage, R.F. (2002). Variation in the k_{cat} of Rubisco in C_3 and C_4 plants and some implications for photosynthetic performance at high and low temperature. J. Exp. Bot. **53**:609–620.

Samuilov, S., Brilhaus, D., Rademacher, N., Flachbart, S., Arab, L., Alfarraj, S., Kuhnert, F., Kopriva, S., Weber, A.P.M., Mettler-Altmann, T., and Rennenberg, H. (2018). The Photorespiratory BOU Gene Mutation Alters Sulfur Assimilation and Its Crosstalk With Carbon and Nitrogen Metabolism in *Arabidopsis thaliana*. Front. Plant Sci. **9**:1709.

Scheffen, M., Marchal, D.G., Beneyton, T., Schuller, S.K., Klose, M., Diehl, C., Lehmann, J., Pfister, P., Carrillo, M., He, H., et al. (2021). A new-to-nature carboxylation module to improve natural and synthetic CO_2 fixation. Nat. Catal. **4**:105–115.

Schlüter, U., Bräutigam, A., Gowik, U., Melzer, M., Christin, P.-A., Kurz, S., Mettler-Altmann, T., and Weber, A.P. (2017). Photosynthesis in C_3–C_4 intermediate Moricandia species. J. Exp. Bot. **68**:191–206.

Schlüter, U., Bouvier, J.W., Guerreiro, R., Malisic, M., Kontny, C., Westhoff, P., Stich, B., and Weber, A.P.M. (2023). Brassicaceae display diverse photorespiratory carbon recapturing mechanisms. J. Exp. Bot. erad250. https://doi.org/10.1093/jxb/erad250.

Schwander, T., Schada von Borzyskowski, L., Burgener, S., Cortina, N.S., and Erb, T.J. (2016). A synthetic pathway for the fixation of carbon dioxide in vitro. Science **354**:900–904.

Seyfferth, C., Renema, J., Wendrich, J.R., Eekhout, T., Seurinck, R., Vandamme, N., Blob, B., Saeys, Y., Helariutta, Y., Birnbaum, K.D., et al. (2021). Advances and Opportunities of Single-Cell Transcriptomics for Plant Research. Annu. Rev. Plant Biol. **72**:1–20.

Shen, B.-R., Wang, L.-M., Lin, X.-L., Yao, Z., Xu, H.-W., Zhu, C.-H., Teng, H.-Y., Cui, L.-L., Liu, E.E., Zhang, J.-J., et al. (2019). Engineering a New Chloroplastic Photorespiratory Bypass to Increase Photosynthetic Efficiency and Productivity in Rice. Mol. Plant **12**:199–214.

Shih, P.M., Zarzycki, J., Niyogi, K.K., and Kerfeld, C.A. (2014). Introduction of a synthetic CO_2-fixing photorespiratory bypass into a cyanobacterium. J. Biol. Chem. **289**:9493–9500.

South, P.F., Cavanagh, A.P., Liu, H.W., and Ort, D.R. (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. Science **363**, eaat9077.

De Souza, A.P., Burgess, S.J., Doran, L., Hansen, J., Manukyan, L., Maryn, N., Gotarkar, D., Leonelli, L., Niyogi, K.K., and Long, S.P. (2022). Soybean photosynthesis and crop yield are improved by accelerating recovery from photoprotection. Science **377**:851–854.

Sun, G., Xia, M., Li, J., Ma, W., Li, Q., Xie, J., Bai, S., Fang, S., Sun, T., Feng, X., et al. (2022). The maize single-nucleus transcriptome comprehensively describes signaling networks governing movement and development of grass stomata. Plant Cell **34**:1890–1911.

Tcherkez, G., and Farquhar, G.D. (2021). Rubisco catalytic adaptation is mostly driven by photosynthetic conditions – Not by phylogenetic constraints. J. Plant Physiol. **267**, 153554.

Timm, S., Woitschach, F., Heise, C., Hagemann, M., and Bauwe, H. (2019). Faster Removal of 2-Phosphoglycolate through Photorespiration Improves Abiotic Stress Tolerance of Arabidopsis. Plants **8**:563.

Tomeo, N.J., and Rosenthal, D.M. (2018). Photorespiration differs among *Arabidopsis thaliana* ecotypes and is correlated with photosynthesis. J. Exp. Bot. **69**:5191–5204.

Tosens, T., Niinemets, Ü., Westoby, M., and Wright, I.J. (2012). Anatomical basis of variation in mesophyll resistance in eastern Australian sclerophylls: news of a long and winding path. J. Exp. Bot. **63**:5105–5119.

Trudeau, D.L., Edlich-Muth, C., Zarzycki, J., Scheffen, M., Goldsmith, M., Khersonsky, O., Avizemer, Z., Fleishman, S.J., Cotton, C.A.R., Erb, T.J., et al. (2018). Design and in vitro realization of carbon-conserving photorespiration. Proc. Natl. Acad. Sci. USA **115**:E11455–E11464.

Ubierna, N., Cernusak, L.A., Holloway-Phillips, M., Busch, F.A., Cousins, A.B., and Farquhar, G.D. (2019). Critical review: incorporating the arrangement of mitochondria and chloroplasts into models of photosynthesis and carbon isotope discrimination. Photosynth. Res. **141**:5–31.

Uemura, K., Suzuki, Y., Shikanai, T., Wadano, A., Jensen, R.G., Chmara, W., and Yokota, A. (1996). A Rapid and Sensitive Method for Determination of Relative Specificity of RuBisCO from Various Species by Anion-Exchange Chromatography. Plant Cell Physiol. **37**:325–331.

Vögeli, B., Schulz, L., Garg, S., Tarasava, K., Clomburg, J.M., Lee, S.H., Gonnot, A., Moully, E.H., Kimmel, B.R., Tran, L., et al. (2022). Cell-free prototyping enables implementation of optimized reverse β-oxidation pathways in heterotrophic and autotrophic bacteria. Nat. Commun. **13**:3058.

Walker, B.J., Vanloocke, A., Bernacchi, C.J., and Ort, D.R. (2016a). The Costs of Photorespiration to Food Production Now and in the Future. Annu. Rev. Plant Biol. **67**:107–129.

Walker, B.J., Skabelund, D.C., Busch, F.A., and Ort, D.R. (2016b). An improved approach for measuring the impact of multiple CO_2 conductances on the apparent photorespiratory CO_2 compensation point through slope–intercept regression. Plant Cell Environ. **39**:1198–1203.

Wang, L.-M., Shen, B.-R., Li, B.-D., Zhang, C.-L., Lin, M., Tong, P.-P., Cui, L.-L., Zhang, Z.-S., and Peng, X.-X. (2020a). A Synthetic Photorespiratory Shortcut Enhances Photosynthesis to Boost Biomass and Grain Yield in Rice. Mol. Plant **13**:1802–1815.

Walker, B.J., Drewry, D.T., Slattery, R.A., VanLoocke, A., Cho, Y.B., and Ort, D.R. (2017). Chlorophyll Can Be Reduced in Crop Canopies with Little Penalty to Photosynthesis. Plant Physiol. **176**:1215–1232.

Wang, X., Wang, Y., Ling, A., Guo, Z., Asim, M., Song, F., Wang, Q., Sun, Y., Khan, R., Yan, H., and Shi, Y. (2020b). Rationale: Photosynthesis of Vascular Plants in Dim Light. Front. Plant Sci. **11**, 573881.

Wei, S., Li, X., Lu, Z., Zhang, H., Ye, X., Zhou, Y., Li, J., Yan, Y., Pei, H., Duan, F., et al. (2022). A transcriptional regulator that boosts grain yields and shortens the growth duration of rice. Science **377**, eabi8455.

Wu, A. (2023). Modelling plants across scales of biological organisation for guiding crop improvement. Funct. Plant Biol. **50**:435–454.

Wurtzel, E.T., Vickers, C.E., Hanson, A.D., Millar, A.H., Cooper, M., Voss-Fels, K.P., Nikel, P.I., and Erb, T.J. (2019). Revolutionizing agriculture with synthetic biology. Nat. Plants **5**:1207–1210.

Xin, C.-P., Tholen, D., Devloo, V., and Zhu, X.-G. (2015). The benefits of photorespiratory bypasses: how can they work? Plant Physiol. **167**:574–585.

Xiong, H., Hua, L., Reyna-Llorens, I., Shi, Y., Chen, K.-M., Smirnoff, N., Kromdijk, J., and Hibberd, J.M. (2021). Photosynthesis-independent production of reactive oxygen species in the rice bundle sheath during high light is mediated by NADPH oxidase. Proc. Natl. Acad. Sci. USA **118**, e2022702118.

Yishai, O., Goldbach, L., Tenenboim, H., Lindner, S.N., and Bar-Even, A. (2017). Engineered Assimilation of Exogenous and Endogenous Formate in *Escherichia coli*. ACS Synth. Biol. **6**:1722–1731.

Yu, H., Li, X., Duchoud, F., Chuang, D.S., and Liao, J.C. (2018). Augmenting the Calvin-Benson-Bassham cycle by a synthetic malyl-CoA-glycerate carbon fixation pathway. Nat. Commun. **9**:2008.

Zarzycki, J., Brecht, V., Müller, M., and Fuchs, G. (2009). Identifying the missing steps of the autotrophic 3-hydroxypropionate CO2 fixation cycle in *Chloroflexus aurantiacus*. Proc. Natl. Acad. Sci. USA **106**:21317–21322.

Zelitch, I., Schultes, N.P., Peterson, R.B., Brown, P., and Brutnell, T.P. (2009). High glycolate oxidase activity is required for survival of maize in normal air. Plant Physiol. **149**:195–204.

Zhang, X., Madi, S., Borsuk, L., Nettleton, D., Elshire, R.J., Buckner, B., Janick-Buckner, D., Beck, J., Timmermans, M., Schnable, P.S., and Scanlon, M.J. (2007). Laser Microdissection of Narrow Sheath Mutant Maize Uncovers Novel Gene Expression in the Shoot Apical Meristem. PLoS Genet. **3**, e101.

Zhang, Z., Lu, Y., Zhai, L., Deng, R., Jiang, J., Li, Y., He, Z., and Peng, X. (2012). Glycolate Oxidase Isozymes Are Coordinately Controlled by GLO1 and GLO4 in Rice. PLoS One **7**, e39658.

Zhang, C., Zhong, X., Lin, D., Wu, K., Wu, Z., Zhang, Z., and Peng, X. (2022). Grain Quality Affected by Introducing Photorespiratory Bypasses into Rice. Agronomy **12**:566.

Zhu, X.-G., Long, S.P., and Ort, D.R. (2010). Improving photosynthetic efficiency for greater yield. Annu. Rev. Plant Biol. **61**:235–261.

3.2 Secondary carbon-fixation improves photorespiration

The paper "Secondary carbon-fxation improves photorespiration" evaluates all currently known photorespiratory bypasses using both a genome-scale stoichiometric model, as well as a smaller, mechanistic model based on ordinary differential equations. My contribution to this was building and analysing the ODE model, as well as writing the manuscript together with Ed Smith.

Publication	Evaluating photorespiratory bypasses
Author position	Shared first
Status	pre-print
doi	...

Alternatives to photorespiration: A systems-level analysis reveals mechanisms of enhanced plant productivity

Edward N. Smith [1] †, Marvin van Aalst [2], Andreas P.M. Weber [3], Oliver Ebenhöh [2,3], and Matthias Heinemann [1]

[1] Molecular Systems Biology, Groningen Biomolecular Sciences and Biotechnology, 9747 AG Groningen, the Netherlands
[2] Institute of Theoretical and Quantitative Biology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[3] Institute of Plant Biochemistry, Cluster of Excellence on Plant Science (CEPLAS), Heinrich Heine University Düsseldorf, Düsseldorf, Germany
† Present address: Department of Biology, University of Oxford, Oxford, UK

Photorespiration causes a significant decrease in crop yield due to mitochondrial decarboxylation. Alternative pathways (APs) have been designed to relocate the decarboxylating step or even fix additional carbon. To improve the success of transferring those engineered APs from model species to crops we must understand how they will interact with metabolism and how the plant physiology affects their performance. Here we used multiple mathematical modelling techniques to analyse and compare existing AP designs. We show that carbon-fixing APs are the most promising candidates to replace native photorespiration in major crop species. Our results demonstrate the different metabolic routes that APs employ to increase yield and which plant physiology can profit most from them. We anticipate our results to guide the design of new APs and to help improve existing ones.

Correspondence: Edward N. Smith, edward.smith@biology.ox.ac.uk
Code: https://gitlab.com/gain4crops/2024-paper

Introduction

Rubisco is the primary site of carbon fixation in plants. Alongside CO_2, ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) can also react with oxygen to produce one molecule of 3-phosphoglycerate (3PGA) and one molecule of 2-phosphoglycolate (2PG). Through a process called photorespiration, the 2-carbon product of the rubisco oxygenase reaction is converted back to an intermediate that can replenish the Calvin-Benson-Bassham cycle (CBB cycle). In this process, CO_2 and ammonia are released in the mitochondria and reassimilated in the chloroplasts. Photorespiration consumes ATP and reducing power and causes losses of approximately 26 % of fixed CO_2 and up to 36 % of the yield of certain crops [1, 2]. Thus, next to efforts in reducing the oxygenase activity of rubisco, the carbon and energy efficiency of photorespiration can be improved [2, 3, 4, 5]. This is highly desired to feed an increasing human population and to mitigate climate change.

Alternative pathways (APs) can increase the carbon and energy efficiency of photorespiration while still serving the primary function of photorespiration: detoxifying 2PG and replenishing the intermediates of the CBB cycle following the rubisco oxygenase reaction. There are four ways this can be achieved, ranging from fixing an additional CO_2 to two CO_2 per cycle being released (Figure 1). For fixing additional carbon, the two carbon 2PG can be converted into a 3-carbon compound such as 3PGA by addition of one carbon from CO_2. A carbon-neutral photorespiratory pathway, with no net loss or fixation of CO_2, is accomplished by combining 2PG with other compounds to generate a 4- or 5-carbon compound that can replenish the CBB cycle directly. In partial decarboxylation pathways, such as native photorespiration, a 2PG molecule is decarboxylated to form a 1-carbon compound that can be combined with another 2PG to make a 3-carbon compound that can enter the CBB cycle. Finally, 2PG can be completely decarboxylated, releasing two CO_2 that can be refixed by the CBB cycle.

From a carbon perspective an AP that fixes additional CO_2 would be preferred. However, additional carbon fixation comes along with extra energy and redox requirements, decreasing the supply for other cellular processes. In contrast, decarboxylating APs generate reducing power that can be captured as redox equivalents as in the Kebeish and South AP3 pathways [6, 7]. In addition to that, decarboxylating APs can shift the rubisco activity towards carboxylation by releasing CO_2 in the chloroplast and thereby increasing the local concentration of CO_2. Thus, next to mere carbon efficiency, several other cellular parameters might determine the overall benefit of an AP.

Native photorespiration has cellular functions other than just clearing 2PG and replenishing the CBB cycle. Further functions include helping to balance ATP:NADPH supply and demand [8], increasing nitrogen assimilation into amino acids [9] and acting as an important source of 1C units [10] as well as serine and glycine [11]. An ideal APs must therefore still support these other roles of native photorespiration.

All these aspects together indicate that implementing

photorespiratory APs to accomplish increased crop yield is far from trivial. In fact, simultaneously interfering with carbon, nitrogen, energy, and redox balances, means interfering with a highly complex, intertwined system. Thus, whether any change increases or decreases crop yield is dependent on metabolism, physiology and conditions, i.e. gas exchange differs with physiology and conditions and metabolism differs with conditions and growth stage/time of day etc. Towards rational engineering and full understanding of APs, global experimental analyses and mathematical modelling will likely be required to evaluate and understand the effect of variants. Exemplary for the challenge and for our yet limited understanding is the fact we still do not fully understand the true cause for the increased yields observed in complete decarboxylation pathways that have been experimentally tested in plants [12].

In this work, we quantitatively examine 12 APs using different mathematical approaches (Figure 1), building on previous modelling efforts [4]. These APs include those that have experimentally been tested *in planta*, but also include newly designed pathways that are either carbon-neutral or fix additional CO_2 [13]. Our assessment includes carbon efficiency, energy and redox effects and effects emerging from altered subcellular CO_2 levels in order to make judgments on the yield increases that one can expect on the system level. In addition, we make predictions on optimal variants of AP designs (e.g. optimal enzyme location) and which types of plants might benefit most from certain pathways and under which environmental conditions. By developing a mechanistic understanding of how APs can be effective, we identify the most promising pathways. Thereby, we hope to contribute to the challenge of increasing photosynthetic efficiency and thereby crop yield, to meet pressing societal issues.

Results

Description of APs

Several APs have been proposed or implemented to date, which can be classified with regards to their CO_2 stoichiometry, ranging from fixing an additional CO_2, via being CO_2 neutral down to two molecules of CO_2 being released (see Figure 2 and [4] for a review).

A CO_2 fixing AP is the tartronyl-CoA pathway (TaCo), which activates glycolate with CoA, carboxylates the glycolyl-CoA to tartronyl-CoA and then reduces the tartronyl-CoA to generate glycerate (blue line in Figure 2) [13, 14]. The enzymes in the pathway and the metabolite tartronyl-CoA are not known to occur in nature, and therefore novel enzyme activities had to be engineered starting from related promiscuous enzymes. In particular, the glycol-CoA carboxylase, catalysing the fixation of bicarbonate to form tartronyl-CoA, has been subject to multiple rounds of engineering to alter the kinetic properties and to reduce the futile hydrolysis of ATP [14, 15]. While the TaCo pathway has been validated *in-vivo* [14], it has yet to be tested in photosynthetic organisms.

Carbon-neutral pathways have been proposed based on a glycolate reduction pathway, which converts glycolate to glycolaldehyde with glycolyl-CoA synthetase and glycolyl-CoA reductase, consuming ATP and NADPH (yellow lines in Figure 2) [13]. The glycolaldehyde is subsequently condensed with a sugar phosphate to generate longer chain sugars or sugar phosphates that can re-enter the CBB cycle via further conversion steps. Four variants have been proposed depending on the intermediates generated, namely arabinose-5P, ribulose-1P, erythrulose or xylulose. These carbon-neutral pathways required the engineering of novel enzyme activities for the two steps of the glycolate reduction. These novel pathways have been validated *in-vivo* [13], but are yet to be tested in photosynthetic organisms.

Partial decarboxylating pathways rely on combining two glycolate molecules and releasing one CO_2 to generate a three-carbon intermediate (green lines in Figure 2). Several pathways with this design have been experimentally validated in plants. The pathways can be divided into those that rely on the *E. coli* glycolate catabolic pathway of glyoxylate condensation via tartronic semialdehyde (Carvalho, Kebeish/South AP1 and Wang pathways) [5, 6, 7, 16] and the β-hydroxyaspartate cycle (BHAC), which is the primary glycolate assimilation pathway in marine proteobacteria [17, 18]. The tartronic semialdehyde-based pathway generates glycerate that can be phosphorylated and re-enter the CBB cycle. The Kebeish/South AP1 and Wang pathways have been implemented in chloroplasts and the Carvalho pathway has been expressed in the peroxisome. The BHAC pathway generates a 4C product from two glycolate molecules, oxaloacetate (OAA), which can directly be incorporated into biomass via aspartate or can be used to replenish the CBB cycle. For this it must be decarboxylated to generate a three carbon CBB cycle intermediate, for instance 3PGA, via phosphoenolpyruvate-carboxykinase and enolase. So far the BHAC pathway has been implemented in peroxisomes of Arabidopsis [17]. The BHAC pathway could also theoretically be expressed in the plastid, which would facilitate CO_2 reassimilation, but this remains to be tested in plants. The partial decarboxylation APs all release 0.5 CO_2 per rubisco oxygenase reaction, the same as in native photorespiration, but are proposed to be more energy efficient as they all avoid the release and subsequent refixation of ammonium.

Complete decarboxylation pathways convert glycolate to CO_2 using either malate synthase and pyruvate dehydrogenase (Maier/SouthAP2/AP3) [7, 19], or oxalate oxidase (Shen)[20] (dark orange lines in Figure 2). The primary difference between the two options is the fate of reductant released from glyoxylate oxidation. Malate synthase generates NADPH and pyruvate de-

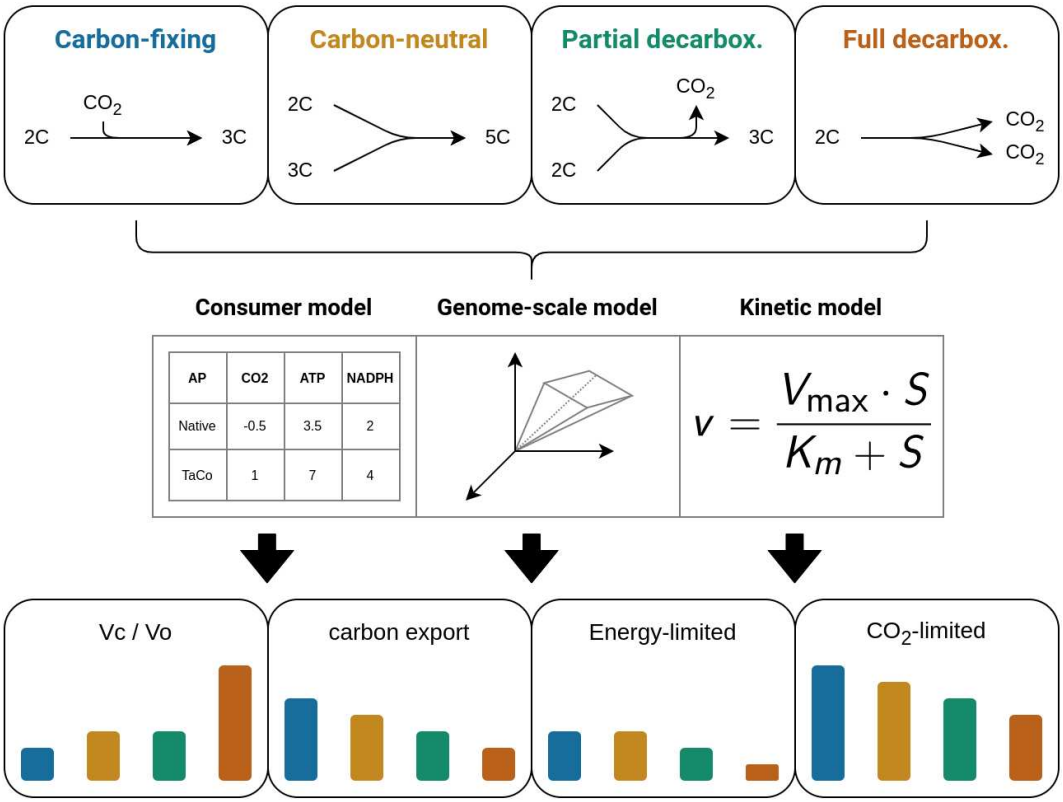


Figure 1: Graphical overview of this work. We analysed 12 alternative photorespiratory pathways (APs) that can be described using four general schemes regarding their carbon stoichiometry: carbon-fixing, carbon-neutral, partial decarboxylating and full decarboxylating pathways. The three modelling approaches we used are a simple stoichiometric consumer model, a genome-scale model solved using flux-balance analysis and a kinetic model to describe key reactions with higher fidelity. Modelling at these different scales allowed the mechanisms behind potential AP benefits to be evaluated.

hydrogenase generates NADH, whereas in the oxalate oxidase pathway two redox equivalents are transferred to water generating two H_2O_2. The CO_2 released by glycolate decarboxylation can re-enter the CBB cycle and therefore complete decarboxylating pathways can still fulfil the requirement of converting 2PG into a CBB cycle intermediate.

In both partial and complete decarboxylating APs and in native photorespiration, oxidation of glycolate is the first step (Figure 2 a, b). Variants exist on in which subcellular compartment this oxidation takes place, and on what the final electron acceptor is of the oxidation. In native photorespiration, the Carvalho pathway and BHAC (perox.) pathway, glycolate oxidase is located in the peroxisome and converts glycolate to glyoxylate generating H_2O_2. In the Maier/South AP2, Shen, and Wang pathways [5, 7, 19, 20], glycolate oxidase was relocated to the chloroplast potentially altering subcellular redox dynamics. Glycolate dehydrogenase from E. coli, which generates NADH, was used for the Kebeish and South AP1 pathways and glycolate dehydrogenase from the algae Chlamydomonas reinhardtii (CrGDH), with an as yet unknown electron acceptor, was used in the South AP3 pathway [7, 21]. The different electron acceptors of the glycolate oxidation reaction have effects on the energy state and redox metabolism in the chloroplast, potentially affecting stress responses to high-light.

Overall, the above-described APs all convert 2PG to a CBB cycle intermediate. Yet, they differ in the CO_2 stoichiometry, the energetic and redox costs, and their subcellular location.

A consumer model allows comparison of APs stoichiometries

To compare the ATP, redox equivalents and CO_2 stoichiometry of both the CBB cycle, photorespiration and APs it is convenient to define the pathways as closed cycles that regenerate one molecule of ribulose-1,5-bisphosphate (RuBP) and fix or release CO_2 and triose phosphates in a so-called consumer model [13]. In this way, photorespiration can be separated from the CBB cycle to directly compare energy and CO_2 stoichiometries. To calculate the ATP, redox equivalents and CO_2 costs of the APs the individual reaction steps starting and ending at RuBP were summed (Table 1).

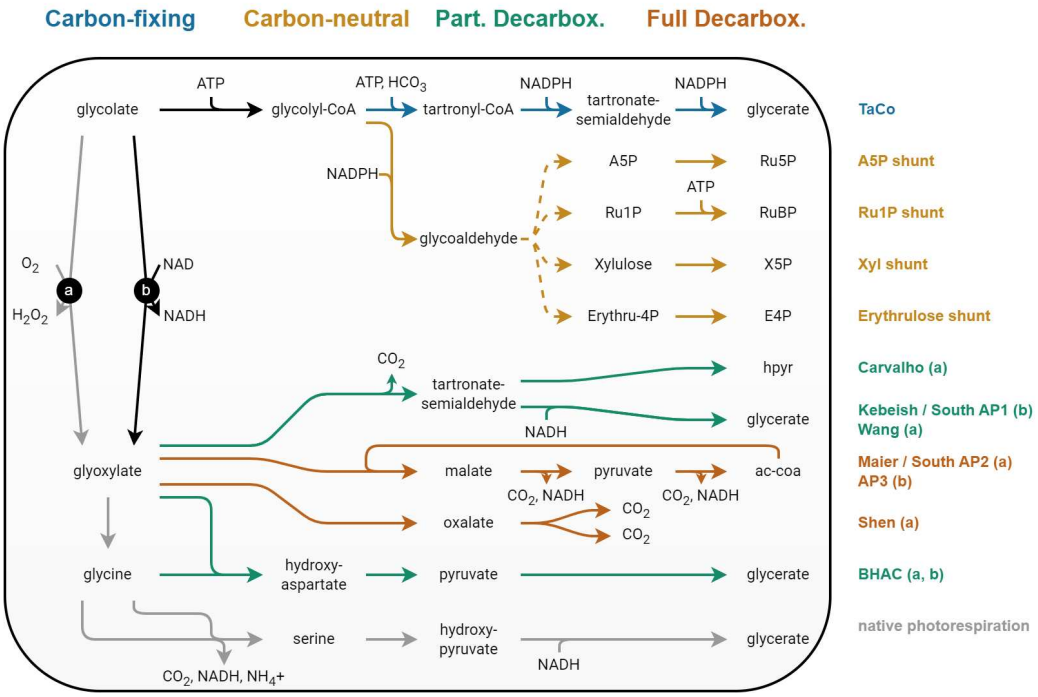


Figure 2: Schematic depiction of APs studied in this work. Grey arrows indicate wildtype reactions, colored arrows signify the classification regarding the carbon stoichiometry of the pathways: carbon-fixing, carbon-neutral, partial decarboxylating and full decarboxylating. Solid lines depict single reactions, while dashed lines depict intermediate steps that were omitted for clarity. The pathway variants **a** and **b** denote whether O_2 / hydrogen peroxide (H_2O_2) or Nicotinamide adenine dinucleotide (NAD) / Nicotinamide adenine dinucleotide (reduced) (NADH) are used as the redox pair for glycolate dehydrogenase.

AP	CO_2	ATP	Redox eq.	Total energy cost	ATP at net-zero CO_2	Redox eq. at net-zero CO_2	Total energy cost at net-zero CO_2
TaCo	1	7	4	17	4	2	9
Ara5P/Ru1P shunts	0	4	2	9	4	2	9
Xyl./Eryth. shunts	0	5	2	10	5	2	10
BHAC (plastid)	-0.5	3	1	5.5	4.5	2	9.5
Kebeish/South AP1	-0.5	3	1	5.5	4.5	2	9.5
BHAC (perox.)	-0.5	3	2	8	4.5	3	12
Carvalho	-0.5	3	2	8	4.5	3	12
Wang	-0.5	3	2	8	4.5	3	12
Photorespiration	-0.5	3.5	2	8.5	5	3	12.5
South AP3	-2	2	-2	-3	8	2	13
Maier/South AP2	-2	2	-1	-0.5	8	3	15.5
Shen	-2	2	1	4.5	8	5	20.5
CBB cycle	1	3	2	8			

Table 1: Consumer model allows comparison of APs stoichiometries. ATP, redox equivalent and CO_2 stoichiometry of AP defined as converting 2PG to RuBP (extended from [4]). Positive values represent consumption, negative values production. The total energy cost assumes 2.5 ATP per redox equivalent. To account for differences in CO_2 stoichiometry, the energy costs for 0 net change in CO_2 were calculated by assuming the CBB cycle can compensate for the release / uptake of CO_2 caused by photorespiration or APs. For example, for native photorespiration an additional 1.5 ATP and 1 redox equivalent must be spent to compensate for the 0.5 CO_2 released. For the TaCo pathway 3 ATP and 2 redox equivalents are spared as 1 CO_2 is already fixed by the AP.

The three complete decarboxylation pathways, South AP3, Maier/South AP2 and Shen, require the same amount of ATP but differ in their redox equivalent stoichiometry, e.g. production of H_2O_2 or NAD(P)H

(Table 1). Here, and in subsequent analyses, we assumed the Chlamydomonas reinhardtii glycolate dehydrogenase (CrGDH) used in the South AP3 pathway indirectly produces NADH, although the exact electron acceptor remains unknown [21]. The South AP3 and Maier/South AP2 pathways have the lowest total energy costs because they net generate NAD(P)H from the decarboxylation of glycolate (Table 1). Compared to native photorespiration, the partial decarboxylation pathways BHAC (perox.), Carvalho, and Wang pathways save 0.5 ATP and 0.5 NADPH by avoiding ammonia release, but require an additional NADH for reduction of tartronic-semialdehyde and do not gain 0.5 NADH from the oxidation of glycine. Therefore, avoiding the cost of ammonium release in these pathways only saves 0.5 ATP per rubisco oxygenase conversion when compared to native photorespiration. The Kebeish/South AP1 and BHAC (plast.) pathways have the same benefits from avoiding ammonia release but also produce an additional NADH from the use of glycolate dehydrogenase in place of the glycolate oxidase used by the Carvalho, Wang, and BHAC (perox) pathway. The carbon-neutral pathways require 0.5-1.5 more ATP than native photorespiration but require the same amount of redox equivalents (Table 1). The carbon-fixing TaCo pathway has the largest total energy cost as redox equivalents and ATP are required to fix CO_2 in this pathway (Table 1). Overall, there is a positive correlation between CO_2 released/taken up and the total energy cost of the pathway. Here, carbon-fixing pathways have the largest energy cost and CO_2-releasing pathways, which can generate redox equivalents from the oxidation of previously fixed carbon, have the lowest energy cost.

Net-zero CO_2 consumer model accounts for the cost of CO_2 release

The previous consumer model does not account for the increase in energetic cost caused by CO_2 release, which needs to be refixed by the CBB cycle, or the reduction in energetic costs if CO_2 is fixed by the APs. Accounting for the cost of modified CO_2 stoichiometry provides a different perspective and allows for more direct comparison between pathways. For example, in native photorespiration if we also account for the cost of fixing the additional 0.5 carbons required to compensate for the loss of CO_2 by glycine decarboxylase, then an additional cost of 1.5 ATP and 1 redox equivalent are required (with the simplifying assumption that the CBB cycle is able to fix CO_2 with no associated rubisco oxygenase or photorespiration). Therefore, the total cost for native photorespiration assuming no net release/uptake of CO_2, is 5 ATP and 3 redox equivalents per regeneration of RuBP (Table 1). Similar calculations were performed for all APs. Note that for the TaCo pathway, which fixes one CO_2, we calculated a benefit in terms of ATP and NADPH spared by not requiring CBB cycle flux.

The complete decarboxylation pathways now have the largest total energy costs as an additional 6 ATP and 4 NADPH are required to refix the 2 CO_2 that are lost by these pathways (Table 1). The partial decarboxylation pathways have a small additional energy cost for refixation of 0.5 CO_2 and the carbon-neutral pathways have no additional costs/benefits as no additional CO_2 must be fixed. In contrast, the TaCo pathway spares 3 ATP and 2 NADPH as it fixes an additional CO_2, reducing the total energy cost. Overall, from an energetic perspective the importance of accounting for CO_2 stoichiometry becomes apparent as large direct energy costs in CO_2-fixing pathways can be compensated for by the energy saving from additional CO_2 fixation.

Stoichiometric network modelling accounts for system-wide effects of APs

While the consumer model offers a useful description of the pathway stoichiometries, it does not capture the interplay of APs with the wider metabolic network, such as amino acid metabolism. Therefore, we extended our analysis by integrating the individual APs in a large-scale stoichiometric model of core plant metabolism based on an Arabidopsis leaf, including photosynthetic electron transport and subcellular compartments [22]. We used these models to investigate whether the APs can replace the complete flux of native photorespiration. Therefore, photorespiration was blocked by constraining the flux through glycine decarboxylase to zero. To ensure flux through photorespiration or an AP was required, we mimicked ambient CO_2 partial pressure by fixing the rubisco carboxylase:oxygenase ratio at 3:1 [23].

Carbon-fixing APs are more energy efficient

We first assessed the photosynthetic energy efficiency of the APs, which we defined as the CO_2 fixed per photon absorbed. For this, a sink reaction for glyceraldehyde-3-phosphate (GAP) was fixed to 1 $\mu mol\, s^{-1}$ and the optimisation objective set to minimisation of photon influx. Thus, the most energy efficient flux distribution was identified for each AP.

The photosynthetic energy efficiency of the APs can have largely different values, ranging from an 27 % increase down to a -54 % decrease relative to native photorespiration (Figure 3). The carbon-fixing TaCo pathway shows the largest increase, followed by the carbon-neutral Ara5P and Xylulose APs (Figure 3). The partial decarboxylation APs range from a small increase to almost no change and the complete decarboxylation APs show a large decrease in photosynthetic energy efficiency (Figure 3).

In terms of energetic efficiency, the net carbon exchanged by the AP has the greatest effect, with carbon-fixing APs being the most energy efficient and decarboxylating APs being the least energy efficient (Figure 3). The second most important factor determining the

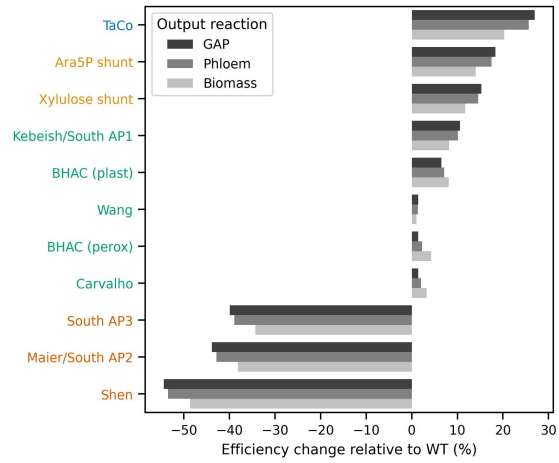


Figure 3: Change in photosynthetic energy efficiency (CO_2 fixed per photon) of APs relative to WT photorespiration for a rubisco carboxylase:oxygenase ratio fixed at 3:1 calculated by stoichiometric FBA modelling. Negative numbers denote a decrease in CO_2 fixed per photon relative to the WT, while positive numbers denote an increase. Model outputs were fixed to either glyceraldehyde 3-phosphate (GAP), phloem exudate or biomass represented by grey shading. AP label colours signify the amount of CO_2 the AP fixed or released by the alternative pathways; blue, 1; yellow, 0; green, -0.5; orange, -2. The absolute CO_2 fixed per photon range from 13.63 for GAP output to 17.67 for biomass, see supplementary Figure S2.

relative energy efficiency is the fate of redox equivalents or electrons from 2PG. Conversion of 2PG to CO_2 can generate three pairs of electrons which can be used to either reduce $NAD(P)^+$ to generate redox equivalents or water to generate H_2O_2. APs which generate more NAD(P)H (Figure 2, Table 1) therefore have an advantage in terms of energy efficiency (Figure 3).

APs can provide biosynthetic energy efficiency benefits

Photorespiration interacts with other metabolic pathways besides the CBB cycle. For example, carbon can be withdrawn from native photorespiration to provide one carbon units (CH_2-THF), glycine or serine which can result in release of less than 0.5 carbons per rubisco oxygenase reaction [11, 24, 25, 26, 27, 28]. Similarly, intermediates can also be withdrawn from APs and used to synthesise amino acids or other biomass precursors.

To calculate the impact that APs have on wider metabolism, we calculated the relative energy efficiency in the presence of APs with more complex cellular outputs including Arabidopsis phloem exudate or biomass [29, 30]. In general, for the carbon-fixing, carbon-neutral, and Kebeish/South AP1 pathways the benefit of the APs decreases as the output complexity increases (Figure 3). Phloem exudate and biomass contain amino acids which can already be efficiently synthesised by native photorespiration. Therefore, the benefit of these APs is smaller when the cell is producing amino acids compared to GAP alone because these APs do not generate amino acids or amino acid precursors as intermediates. In contrast, the BHAC and Carvalho pathways demonstrate an enhanced benefit when synthesising biomass or phloem exudate compared to GAP alone (Figure 3). Synthesis of amino acids with five-carbon backbones is more efficient in the presence of the BHAC and Carvalho pathways as additional separate decarboxylation steps and subsequent refixation of CO_2 are not required to synthesis these amino acids. Similarly, the complete decarboxylation pathways also show a benefit (i.e. decreased energy efficiency penalty) when synthesising biomass or phloem exudate in comparison to just GAP alone (Figure 3). Specifically, in the Maier/South AP2 and South AP3 pathways, carbon can be withdrawn as malate from malate synthase. Malate can then be used as the carbon backbone for generating aspartate and other derived amino acids. In this way, carbon is conserved and not released as CO_2 that must be refixed with an associated energetic cost.

APs can have some small additional benefits compared to native photorespiration when cells are synthesising biomass, and these will depend on the precise amino acid demands of the leaf. The WT model predicts here that approximately 6 % of the carbon entering native photorespiration is withdrawn as serine, substantially less than the 32 % reported for Tobacco leaves [11]. All the APs investigated here bypass the serine producing steps of native photorespiration and could therefore disadvantage the plant if serine is in high demand and other serine producing pathways are unable to compensate. Overall however, the greatest effect of the alternative pathways is on the efficiency of carbon fixation and any changes in the energy efficiency of synthesising biomass precursors is likely to be relatively minor.

APs alter the ATP and NADPH demand

The linear electron flow of photosynthesis supplies a fixed stoichiometry of ATP and redox equivalents that does not necessarily match the demand of the cell. Plants employ various mechanisms to match the supply to the demand, such as cyclic electron flow, but imbalances can potentially lead to photosynthetic inefficiency or damage [8]. Introducing APs can alter the ATP and redox demands of the cell with potentially positive or negative effects. We therefore quantified the ratio of ATP to NADPH demand of the cell in the presence of the APs to see whether the demand is shifted towards or away from the ratio supplied by linear electron flow. For this we identified the most energy-efficient flux distribution by forcing a GAP outflux of 1 $\mu molm^{-2}s^{-1}$ while minimising the photon input and then quantified the fluxes through ATP synthase and ferredoxin-NADP reductase to quantify the

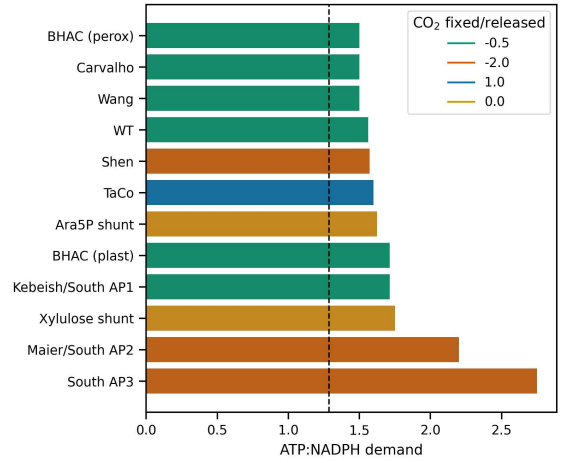


Figure 4: APs alter the ATP:NADPH demand of carbon fixation. The net ATP:NADPH demand of carbon fixation to glyceraldehyde 3-phosphate in the presence of APs calculated using the stoichiometric model and FBA, and assuming a rubisco carboxylase:oxygenase activity of 3:1. Dashed line represents the ATP:NADPH supplied by linear electron flow through the photosystems $\frac{9}{7}$.

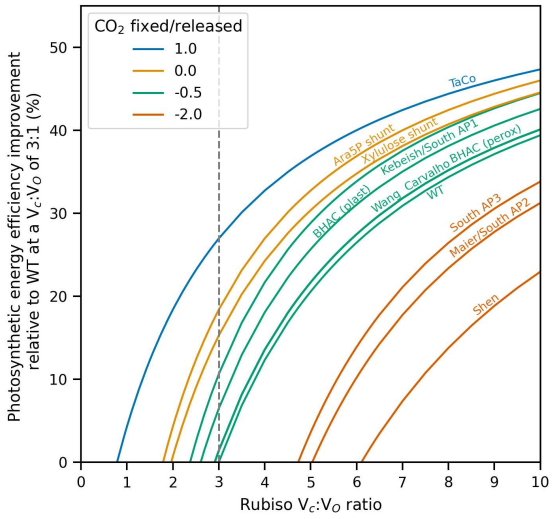


Figure 5: Complete decarboxylation pathways require an increased rubisco carboxylase:oxygenase ratio to achieve an increased photosynthetic energy efficiency relative to WT photorespiration. Effect of altered rubisco carboxylase:oxygenase ratio on photosynthetic energy efficiency of APs relative to WT photorespiration at a fixed rubisco carboxylase:oxygenase ratio of 3:1 (dashed line). Simulations were performed using the stoichiometric model and FBA with the objective set to maximisation of glyceraldehyde 3-phosphate production. Photosynthetic energy efficiency was defined as the CO_2 fixed per photon absorbed.

net ATP and reductant demand. Because the model represents autotrophic leaf metabolism these reactions are the primary source of ATP and NADPH and can therefore be used to calculate the net ATP:NADPH demand of the cell.

Here, only the Carvalho, Wang and peroxisomal BHAC pathways, with an ATP:NADPH demand of 1.5, are closer to the supply from linear electron flow ($\frac{9}{7}$ or 1.28) than that of a plant with WT photorespiration (1.56) (Figure 4). The lower ATP:NADPH demand in the Carvalho, Wang and BHAC (perox.) pathways is due to avoiding the need for ammonia reassimilation and the associated ATP cost. APs that capture the redox equivalents from complete glycolate decarboxylation as NADPH, such as the Maier and South AP3 pathways, significantly increase the demand of ATP relative to NADPH (up to 2.75 in South AP3) (Figure 4). This is caused by the increased NADPH supply from glycolate decarboxylation which decreases the NADPH that must be supplied from photosynthetic electron flow. In contrast, the Shen pathway, which also completely decarboxylates glycolate, does not capture the redox equivalents as NADPH and instead produces H_2O_2, resulting in an ATP:NADPH demand of 1.57 (Figure 4).

Overall, most of the APs cause relatively minor changes in the ATP:NADPH demand compared to native photorespiration with the notable exception of the complete decarboxylating Maier/South AP2 and South AP3 pathways which cause large increases in the ATP:NADPH demand (Figure 4).

Complete decarboxylation pathways require an increased rubisco carboxylation rate to have a benefit

Analysis so far has focused purely on photosynthetic energy efficiency and assumed no limitation of CO_2 diffusion either from outside the cell or between subcellular compartments. However, CO_2 diffusion poses a major limitation to photosynthesis in C3 plants [31]. The complete decarboxylation pathways decrease the photosynthetic energy efficiency relative to WT photorespiration if unlimited CO_2 diffusion is assumed. However, if CO_2 diffusion into the chloroplast is limited, releasing CO_2 specifically within the chloroplast and thereby increasing the carboxylation rate relative to the oxygenation rate could be beneficial. We therefore determined the increase in carboxylation rate relative to the oxygenation rate necessary to compensate for the increased energy cost of the complete decarboxylation pathways.

To model the effect of an increased chloroplast CO_2 concentration, we constrained the rubisco carboxylase:oxygenase ratio of rubisco to a range of values and calculated the photosynthetic energy efficiency relative to the WT in terms of CO_2 fixed per photon absorbed.

At a rubisco carboxylase:oxygenase ratio of 3:1 the complete decarboxylation pathways were less energet-

ically efficient than WT photorespiration (Figure 5). However, these pathways can become more energetically efficient than WT photorespiration if they are able to increase the chloroplast CO_2 concentration such that the rubisco (rubisco carboxylase:oxygenase) ratio reaches 4.8-6.1:1 (Figure 5). This represents a 60-103 % increase relative to the rubisco carboxylase:oxygenase ratio of 3:1 assumed for native photorespiration. The remaining partial decarboxylation, carbon-neutral or carbon fixing APs continue to show an increased photosynthetic efficiency relative to native photorespiration, even at lower rubisco carboxylase:oxygenase ratios (Figure 5).

Native CO_2 refixation capacity affects AP benefit

Another factor that needs to be taken into account is that in native photorespiration CO_2 set free by glycine decarboxylase activity in the mitochondria can diffuse out of the cell before it reaches the chloroplast, and is therefore only partly refixed. Plants have evolved mechanisms to recapture this CO_2 that would otherwise be lost, such as placing chloroplasts around the cell periphery [28, 32, 33, 34, 35] or relocating the decarboxylation step to the bundle sheath cells, as in C3-C4 intermediate photosynthesis [36, 37]. The exact amount of CO_2 refixation depends on both the plant species and the environmental conditions [32, 37]. APs can relocate the site of CO_2 release from the mitochondrion to the chloroplast and therefore potentially increase the proportion of refixed CO_2.

CO_2 diffusion is dependent on a series of resistances between the external and internal airspace and between subcellular organelles [28, 33, 34, 35]. These resistances were simplified in our stoichiometric model to a constraint on the fraction of CO_2 released in mitochondria that can be refixed by chloroplasts (refixation potential), with the remainder assumed to exit the leaf. We modelled the effect of different CO_2 refixation potentials in the WT plant on the relative carbon export of plants expressing APs compared to the WT. We assumed that relocation of CO_2 release by the APs to the chloroplast results in complete recapture of CO_2 released by the APs, therefore representing the maximum potential benefit. The input of CO_2 from outside the cell was fixed to the WT value to represent a CO_2 diffusion limited condition and prevent cells from compensating for CO_2 lost from photorespiration or APs by simply importing more CO_2. Photon input was constrained to the WT value representing an energy limited condition and the optimisation objective set to maximisation of GAP production.

In general, as the CO_2 refixation potential of the WT plant increases, the benefit of an AP which recaptures this otherwise lost CO_2 decreases (Figure 6A). The maximum carbon export increase of any AP that recaptures otherwise lost CO_2 is 20 % for a rubisco carboxylase:oxygenase ratio of 3:1 (Figure 6A). Under the energy limited condition modelled here, the

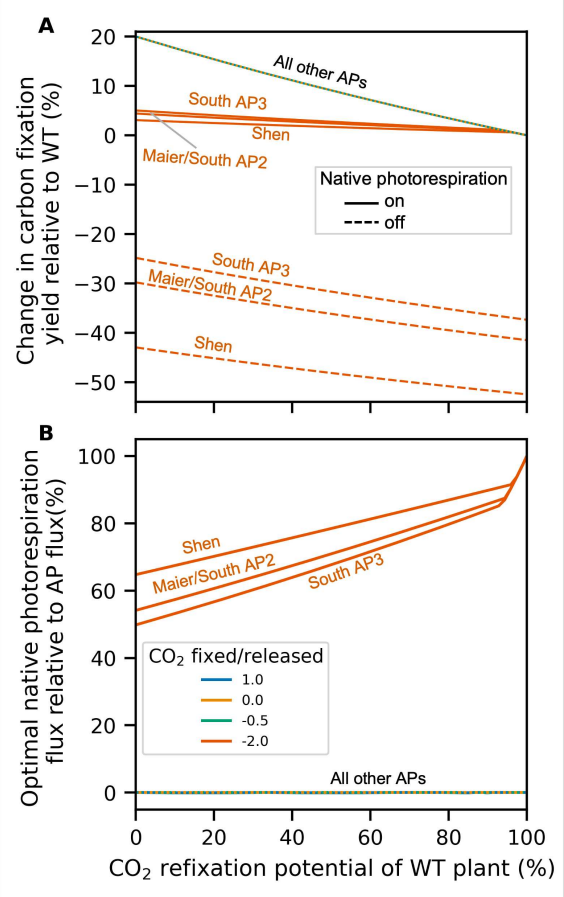


Figure 6: The benefit of an AP that recaptures otherwise lost CO_2, depends on how effective the WT plant is at already refixing CO_2 released by photorespiration. A) The effect of CO_2 refixation in WT plants on the relative benefit of photorespiratory APs to carbon fixation calculated using the stoichiometric model and FBA. Refixation potential is defined as the fraction of CO_2 released from photorespiration in the mitochondria that could enter the chloroplast rather than leaving the cell. The refixation potential of a WT cell was fixed between 0-100 % and the objective set to maximisation of glyceraldehyde 3-phosphate production (GAP). Rubisco carboxylase:oxygenase activity was constrained to 3:1. When simulating the APs, photon and CO_2 inputs were constrained to the WT value required to generate GAP at a rate of 1 $\mu mol\,s^{-1}$ representing an energy and CO_2 limited condition. Dashed lines are with flux though native photorespiration blocked; solid lines are with native photorespiration free to carry flux. B) The optimal flux through native photorespiration relative to the AP at varying CO_2 refixation potentials. Photon and CO_2 inputs were constrained as in A. 100 % means that there is no flux through the AP and 0 % means that the AP completely replaces photorespiration.

APs that are more energetically efficient than native photorespiration (carbon-fixing, carbon-neutral and partial decarboxylation pathways) can all achieve this maximum 20 % benefit (Figure 6A "All other APs"). In contrast, the complete decarboxylation pathways, which are less energy efficient than native photorespiration, fixed less CO_2 than the WT plant at all CO_2 refixation potentials (Figure 6A, dashed orange lines). Next we evaluated whether residual flux through the native photorespiratory pathway could be beneficial in combination with an AP. Native photorespiration potentially loses CO_2 from the cell by releasing it in the mitochondria, affecting the carbon export. APs can recapture this CO_2 by releasing it in the chloroplast, but can be more energetically expensive. Therefore, for the complete decarboxylation APs, there is a trade-off between the relatively more energy efficient native photorespiration and the more carbon efficient AP. We therefore repeated the previous analysis but allowed unlimited flux through native photorespiration and identified the optimal flux through native photorespiration and the AP under a CO_2 and energy limitation.

Under these energy- and CO_2-limited conditions, the optimal flux distribution for the more energy efficient pathways required zero flux through photorespiration (Figure 6B). In contrast, for the less energy efficient complete decarboxylation pathways, the optimal amount of native photorespiratory flux varied with the CO_2 refixation potential (Figure 6 B). In other words, residual flux though native photorespiration can compensate for the energy inefficiency of complete decarboxylation pathways while the APs act to recapture otherwise lost CO_2. Therefore, some residual photorespiratory flux can be beneficial in the presence of complete decarboxylation pathways, particularly when the CO_2 refixation potential is low (Figure 6 A, solid orange lines). The optimal flux through native photorespiration in combination with an AP may vary dynamically with the environmental conditions which can affect CO_2 diffusion and the refixation potential of the plant. Overall, assuming approximately 25 % of CO_2 is refixed in a WT plant, all AP designs are advantageous when both CO_2 and energy are limiting, and when operating in combination with native photorespiration (Figure 6A, solid lines).

Understanding CO_2 dynamics requires kinetic modelling

Stoichiometric modelling demonstrated the energetic and stoichiometric benefits of the AP designs, how flux through certain APs can support amino acid biosynthesis and the potential beneficial effects of avoiding CO_2 release in the mitochondria. However, the stoichiometric model required fixing the rubisco carboxylase:oxygenase ratio, whereas *in-vivo* it depends on the concentrations of O_2 and CO_2 which can vary dynamically. Therefore, to model the effect of varying CO_2 diffusion and CO_2 concentration more accurately, we developed a kinetic model of photosynthesis to test the different AP designs.

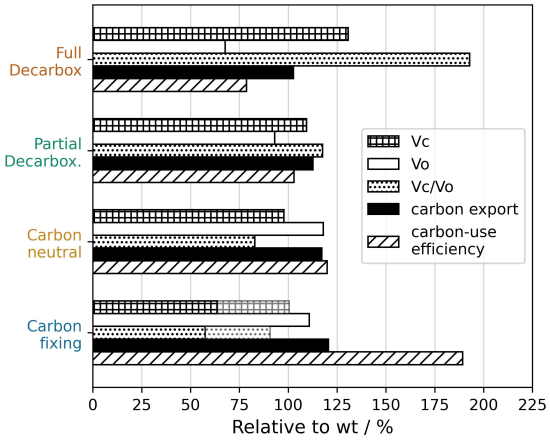


Figure 7: Steady-state performance indicators of plant metabolism simulated by the kinetic model for different AP designs, grouped by their CO_2 stoichiometry. The simulation was performed at reference conditions of 400 ppm atmospheric CO_2 and an illumination of 700 µmol/m^{2}s. Transparent bars signify the addition of secondary carboxylation steps to the measurement of rubisco carboxylation rate for the carbon-fixing pathways. The different AP designs are grouped by their CO_2 stoichiometry and only the pathway variant with the maximal carbon export of the group is shown.

The kinetic model was developed as a system of ordinary differential equations by combining models of the CBB cycle [38] and photorespiration [39], as well as a complete description of rubisco kinetics including carboxylation and oxygenation reactions [40]. CO_2 was modelled as a dynamic variable to capture the effect of the various APs on the CO_2 concentration in the chloroplast. A fixed proportion (25 %) of CO_2 released by glycine decarboxylase in the mitochondria was assumed to diffuse back into the chloroplast [32]. The APs were implemented on top of this WT model and are grouped by their CO_2 stoichiometry, with the best performing variant of each group shown in the following results. For a complete description of the model as well as a comparative analysis of intermediate model stages see supplementary material section S2.

Carbon-fixing pathways are more efficient at exporting carbons

We first evaluated the different AP designs at reference conditions of 400 ppm atmospheric CO_2 and an illumination of 700 µmol/m^{2}s. From this analysis we calculated the rubisco carboxylation and oxygenation rates (V_c and V_o) and rubisco carboxylase:oxygenase ratio. We further extended the evaluation by considering export of triose-phosphates and hexoses as a proxy for plant growth and defined the carbon-use

efficiency as the ratio of carbon export rate relative to the rubisco carboxylase rate.

The carbon-fixing pathways show a reduced rubisco carboxylation rate (less than 60 % of the WT) and slightly increased oxygenation rate, leading to a reduced rubisco carboxylase:oxygenase ratio relative to the WT (Figure 7). The complete decarboxylation APs show the exact opposite, with a higher carboxylation rate (more than 125 % of the WT) and lower oxygenation rate, leading to a rubisco carboxylase:oxygenase ratio of more than 175 % relative to the WT (Figure 7). Even if the secondary carboxylation steps like glycolyl-CoA carboxylase (GCC) in the carbon fixing TaCo pathway are counted in the total carboxylation rate, this general trend stays the same, even though now the total carbon fixation of the carbon-fixing APs is less decreased than the carbon neutral pathways (Figure 7, transparent area). Contrastingly, the carbon export is highest for the carbon-fixing APs (around 120 %) and lowest for the complete decarboxylation APs at slightly above 100 % (Figure 7). This is reflected by the carbon-use efficiency, which is highest for the carbon-fixing APs at around 190 % and lowest for the complete decarboxylation APs at around 80 % (Figure 7). This increase is carbon-use efficiency in carbon-neutral and carbon-fixing APs is caused by the fact that they generate intermediates, other than CO_2, that can enter the CBB cycle. By increasing the input of intermediates into the cycle, more carbon can be withdrawn without requiring additional fixation of CO_2 by rubisco. Therefore, carbon-fixing and carbon-neutral alternative photorespiratory pathways increase the carbon use efficiency of the CBB cycle resulting in more carbon exported per rubisco carboxylase reaction.

APs are most beneficial at low intracellular CO_2

The activity and effectiveness of the APs, which we measure as the improvement of carbon export, depends on the local CO_2 concentration in the chloroplast. The velocity at which this local CO_2 can be replenished is dependent on the external CO_2 concentration as well as the internal transport rate. We systematically evaluated all AP designs for both of these factors. The improvement of carbon export relative to the WT is highest for both low CO_2-transport rate (Figure 8D), as well as low external CO_2 concentration (Figure S1), in which the carbon-fixing pathways perform best (Figure 8D). Importantly, these are also the conditions where the absolute rates of carbon export are reduced (Figure 8B). The carbon-fixing APs consistently show the highest improvement relative to the WT, closely followed by the carbon-neutral APs and then by the partial decarboxylation APs (Figure 8D). The full decarboxylation APs show the lowest improvement and perform worse than the WT if the CO_2-transport rate is increased more than 10 % (Figure 8 D). The improvement to carbon export diminishes for all APs under very high CO_2 conditions (Figure S1).

High-light increases AP benefit

All APs require more energy equivalents than the CBB cycle per carbon export and the full decarboxylation pathways even more than native photorespiration. Therefore, it is important to assess how much energy supply is necessary, such that the benefits of the pathways outweigh this energy cost. As light is the primary source of energy for plants, and natural light conditions are continuously changing, the AP performance should be assessed at a variety of light-conditions. We did this by systematically scanning the effect of illumination on the carbon export, by simulating the steady-state for each photosynthetic photon flux density (PPFD). Here, we found all APs show the highest relative increase in carbon export in high-light conditions, with the carbon-fixing APs showing the highest increase in carbon export at 125 % relative to the WT (Figure 8C). The second-best option are carbon-neutral pathways, which in low-light conditions outperform the carbon-fixing pathways (Figure 8 C illumination below 400 µmol). The complete decarboxylation pathways show a beneficial effect for medium and high-light conditions compared to the WT, but are disadvantageous in low light conditions compared to the other APs (Figure 8C). These results demonstrate that in medium to high-light conditions the benefits of all APs outweigh the increased direct energy cost, while at low light the increased energy demand outweighs the benefits for the complete decarboxylation APs.

Environmental conditions require different AP designs

The previous results highlight that depending on the environmental conditions, which will affect CO_2 diffusion and light intensity, different APs should be utilised to maximise carbon export. We predicted the carbon export improvement of different APs for three distinct scenarios. First, a normal scenario, which corresponds to a well-watered, plant in a temperate climate with sufficient light. Next, a scenario with a well-watered plant in temperate climate but low light, which is thus energy-limited. For this we used an illumination of 250 µmol/m^{2}s. Lastly, a scenario of a hot and dry climate and high light, with partially closed stomata, which is thus CO_2-limited. This we represented by lowering the carbon transport rate to 90 % of the WT.

In both the normal and CO_2-limited scenario the carbon-fixing APs show the highest carbon export (Figure 8E, G). Under energy-limiting conditions carbon-fixing and carbon-neutral APs show similar carbon export (Figure 8F). In all cases, complete decarboxylation pathways show the lowest carbon export, with a negative effect in energy-limited conditions, a slightly positive effect in normal conditions and a positive effect in CO_2-limited conditions (Figure 8E-G). Thus,

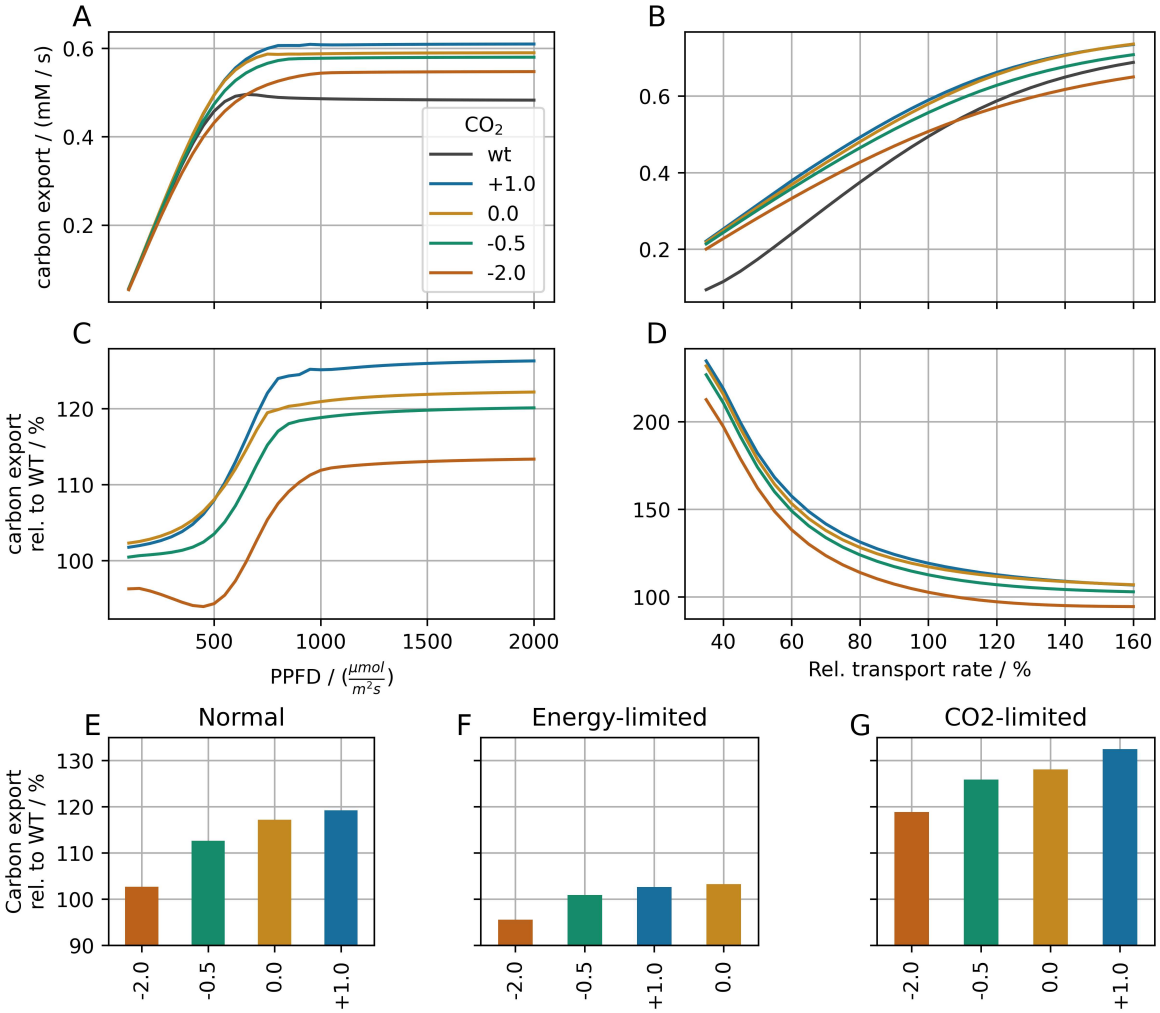


Figure 8: Steady-state carbon export of photorespiratory APs simulated by the kinetic model depending on photosynthetic photon flux density (PPFD) (A, C) or CO_2 transport rate (B, D). Carbon export is shown in both absolute terms (A, B) and relative to the WT under the same conditions (C, D). Three representative scenarios are highlighted for a well-watered plant in a temperate climate with sufficient light (700 µmol/m^2s) (E); a well-watered plant in temperate climate with low light (250 µmol/m^2s), which is thus energy-limited (F); and a hot and dry climate with low water supply and thus partially closed stomata (90 % of WT CO_2 transport rate), which is thus CO_2-limited (G). The different AP designs are grouped by their CO_2 stoichiometry (-2, -0.5, 0, 1) and only the pathway variant with the maximal carbon export of the group is shown.

kinetic-overview

in most scenarios the carbon-fixing APs are the preferred choice, while in energy-limited conditions the carbon-neutral APs might be more beneficial.

Discussion

Using various mathematical models of plant metabolism, we have comprehensively identified and assessed the mechanisms by which APs can enhance carbon assimilation and growth. To understand how environmental conditions and plant physiology can affect pathway benefits, we also quantitatively assessed the effects of light intensity and CO_2 availability on the different APs. We show that carbon-fixing APs, such as the TaCo pathway, have the greatest potential benefit over a range of conditions, and could provide an increase in carbon export from photosynthesis of over 20 %. Due to the initial exponential growth phase of plants, such a percentage could generate substantial gains in biomass over time [41].

APs use distinct mechanisms to improve growth

Our work showed that alternative photorespiratory pathways can provide benefits to plants via five different mechanisms (Table 2): (i) An energy efficiency benefit by metabolising the 2PG produced by the rubisco oxygenase reaction in a way that uses less ATP and reducing power compared to native photorespiration. (ii) A biosynthetic energy efficiency benefit if intermediates of the alternative pathway can be used for biosynthetic reactions. (iii) Recapture of CO_2 that could otherwise diffuse from the cell. (iv) An increased CO_2 concentration at the site of rubisco in the chloroplasts caused by altered CO_2 diffusion within the cell. (v) Increased carbon use efficiency of the CBB cycle caused by an additional input of CBB cycle intermediates, which enables more carbon to be exported per rubisco carboxylase reaction. As CO_2 diffusion differs depending on cell morphology and leaf physiology, these CO_2 related benefits can be different in different plant species. Thus, next to simply clearing the 2PG produced by the rubisco oxygenase reaction, alternative photorespiratory pathways can have broad effects on the plant in terms of energetics, biosynthesis and CO_2 dynamics and these must be considered at the system level to evaluate the effectiveness of any pathway.

Benefits of APs depend on environmental conditions

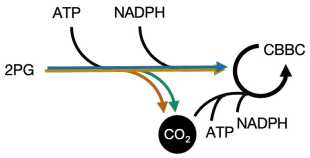
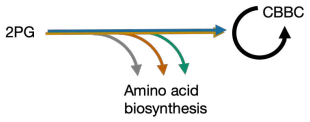
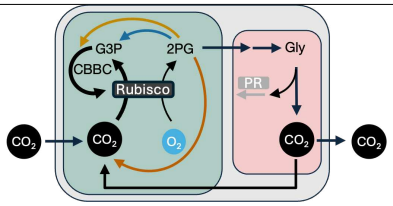
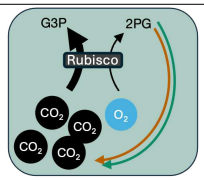
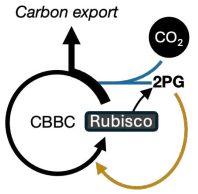
Importantly, the environmental conditions a plant grows in, such as light intensity, temperature, and water availability can affect the benefit of the different mechanisms that alternative pathways employ. Under low-light conditions, when energy is limiting, the energy efficiency of alternative pathways becomes critical to their effectiveness [12, 42]. Our models showed that alternative pathways that are more energy efficient than the native photorespiration will offer greater benefits in low-light, which is experimentally supported by growth enhancement in Arabidopsis expressing the Kebeish pathway and grown under low-light and short-day conditions [6]. In contrast, complete decarboxylation pathways, which are less energy efficient than native photorespiration, are predicted to show no benefits under these low-light conditions. Under high-light conditions energy efficiency no longer offers an advantage, and the pathways' effects on CO_2 diffusion and fixation become more important. As our analyses showed, all pathways perform best under high-light conditions (Figure 8A,C) suggesting that benefits from altered CO_2 diffusion have the biggest potential to increase plant growth. This finding is also supported experimentally with complete decarboxylating pathways including South AP3, Shen, and Wang all showing increased benefits under high-light conditions [5, 7, 20]. Overall, the alternative pathways provide the greatest benefit over native photorespiration under high-light, and CO_2-limiting conditions, with certain pathways also able to provide benefits under low-light conditions.

The benefits of alternative pathways are comparable to those predicted for C3-C4 intermediate or C4 metabolism [43]. Indeed, the conditions which favour these naturally evolved mechanisms are also those we identified as beneficial for the alternative pathways. However, from a metabolic engineering perspective, the alternative pathways described here offer advantages over introducing C3-C4 or C4 metabolism into a C3 plant, as they can require as few as three genes compared to the >15 genetic modifications potentially needed to support the biochemistry, leaf anatomy, and intercellular transport of C3-C4 or C4 photosynthesis. Therefore, alternative pathways offer a more easily implementable solution to the same problems that C3-C4 intermediate and C4 photosynthesis have evolved to address.

Applying new insights to previous results

By applying the insight we have gained, we can now try to better explain previous experimental observations. For example, complete decarboxylating pathways such as the Maier/South AP2, South AP3 and Shen bypasses experimentally showed growth benefits, but the molecular basis for these benefits was less clear, as previous computational analysis based on a kinetic model describing carbon fixation and subcellular CO_2 conductance did not predict an enhanced rate of photosynthesis [12]. Here, we quantitively demonstrate that complete decarboxylating pathways can indeed enhance photosynthesis by increasing the CO_2 concentration in the chloroplast and subsequently the rubisco carboxylase:oxygenase ratio, as well as by recapturing otherwise lost CO_2. These effects are consistent with experimental measurements of unchanged or decreased CO_2 compensation points in engineered plants expressing these pathways [7, 19, 20, 44]. Furthermore, for

Table 2: Alternative photorespiratory pathways can affect carbon assimilation and growth in multiple ways. *Only applies to pathways that relocate CO_2 release to chloroplasts i.e not Carvalho or BHAC(perox.) pathways which are peroxisomal. †Only applies to BHAC and Carvalho pathways not Wang, Kebeish/South AP1. 2PG, 2-phoshpoglycolate; CBBC, Calvin-Benson-Bassham cycle; G3P, glyceraldehyde 3-phosphate.

Direct or indirect effect of pathway		Pathway effect applies to (carbon stoichiometry)				Interactions with physiology and environment
		-2	-0.5	0	1	
i) Energy efficiency			•	•	•	Most relevant when energy is limiting i.e. low-light
ii) Biosynthetic efficiency		•	†			Depends on biomass composition and if leaves are growing or mature
iii) CO_2 recapture		•	*	•	•	Depends on; mitochondria/chloroplast arrangement, leaf CO_2 diffusion, chloroplast CO_2 permeability
iv) Increase chloroplast CO_2 concentration		•	*			Depends on chloroplast CO_2 permeability
v) Increase CBB cycle carbon use efficiency				•	•	High-light and low CO_2 scenarios provide the greatest relative increase in carbon export

complete decarboxylating and some partial decarboxylating pathways, we demonstrated additional increases in energy efficiency in cells synthesizing amino acids for biomass synthesis or phloem exudate that have not previously been identified. Thus, through our work, we can now better explain the reasons why in previous studies a particular alternative pathway has generated benefits.

Some aspects of APs require future investigation

Yet, despite improving our understanding of alternative photorespiratory pathways, a number of experimental observations remain unexplained. For example, the South AP3 pathway expressed in tobacco shows increased CO_2 assimilation even at very high intracellular CO_2 [7]. Decarboxylation pathways reduce rubisco oxygenase activity by increasing the local CO_2 concentration at increased energetic costs. At very high intracellular CO_2 this mechanism cannot offer any additional benefit, as CO_2 concentration in the chloroplast should already be high. The mechanism of the benefit of South AP3 pathway even at high CO_2 remains unexplained and may relate to the reported beneficial effects of expressing glycolate dehydrogenase alone, which could affect the efficiency of photosynthetic electron transport [6, 45, 46, 47]. Additionally, our models predicted that partial decarboxylating pathways should outperform full decarboxylating pathways across conditions. However, comparison of the partial decarboxylating South AP1 and the complete decarboxylating South AP3 expressed in tobacco in

greenhouse trials, showed greater yield increases in the AP3 pathway [7]. In this context, it is important to realise that our model predictions reflect best-case scenarios, requiring optimal expression and kinetics of the pathway enzymes. While a discrepancy between our model predictions and experimental implementations of alternative pathways could point to shortcomings in our models, an alternative explanation for the discrepancy could be that an engineered plant might not yet have optimal enzyme expression levels. Thus, such discrepancy might also indicate potential for further improvements of plant performance.

Conclusion

With this work, we have improved our mechanistic understanding of how alternative photorespiratory pathways can enhance carbon assimilation and growth, and used this to predict that the carbon fixing TaCo pathway represents the best option for increasing crop yields over a range of conditions. In future, crops, or specific cultivars, should be screened to identify those with limitations in CO_2 refixation capacity or CO_2 diffusion that make them most likely to benefit from engineering with alternative pathways to native photorespiration. The models presented here could also be used to evaluate or develop new alternative pathway designs that may further increase yield gains e.g. by targeting specific growth scenarios, such as the juvenile or adult growth stages, or specific crop species. With the ability to more rationally engineer alternative photorespiratory pathways into suitable crops, and identify their optimal growing conditions, our work will hopefully contribute to realising the maximum impact of alternative photorespiratory pathways for improving crop yields.

Methods

Stoichiometric model

A stoichiometric model of core plant metabolism based on an Arabidopsis leaf, PlantCoreMetabolism_v3, was curated starting from a previously described model (PlantCoreMetabolism_v1_2, [22]). The model is available in SBML format as an XML file at https://gitlab.com/gain4crops/2024-paper. Python 3, COBRApy [48] and the CPLEX solver were used for FBA optimisations. All code required to reproduce the results is available at https://gitlab.com/gain4crops/2024-paper as Jupyter notebook files. Output fluxes to phloem exudate were defined from Arabidopsis phloem composition [29]. Output fluxes to biomass were defied as described in the AraGEM model [30]. Net ATP demands were calculated by quantifying the flux through both plastidic and mitochondrial ATP synthase. As the model is autotrophic, all ATP must ultimately be generated by either plastidic ATP synthase, or by mitochondrial ATP synthase using NADH generated in the plastid. To calculate the reductant demand the flux through proton pumping mitochondrial NADH-dehydrogenase (which is used for generating ATP) was subtracted from the flux through plastidic ferredoxin-NADP reductase.

Kinetic model

The ordinary differential equation (ODE) model was built mainly using two previously published models of the CBB cycle [38, 49] and photorespiration (PR) [39] and developed using Python-based software modelbase [50]. Rubisco kinetics including carboxylation and oxygenation reactions were described based on the rate equation from Witzel 2010 [40]. Energy metabolites, ATP and NADPH, were implemented as dynamic variables as in Matuszyńska 2019 [51] and their production modelled using a simplified light-dependent reaction with an additional quenching reaction to account for the different ATP and NADPH demands of the system. Thioredoxin based redox regulation of CBB cycle enzymes was linked to the energy status of the model via an NADPH-thioredoxin reductase reaction based on the description by Saadat 2021 [52]. CO_2 was modelled dynamically with CO_2 input from the atmosphere described with a diffusion equation. As refixation of respired and photorespired CO_2 was shown to range between 24-38 % in wheat and rice [32], a static CO_2 refixation of 25 % from mitochondiral glycine decarboxylation was assumed with the remaining CO_2 lost to the atmosphere. For APs that relocate the CO_2 to the chloroplast we assumed 100 % of this could potentially be refixed. Ammonia was also modelled as a dynamic variable along with ammonium assimilation into glutamate and the associated energy costs. The model was built in a stepwise manner, with each iteration being compared to the previous one to ensure that the changes made were valid (see supplementary material section S2 for a complete description).The final model was used as a reference point, called WT, and the APs were implemented on top of this model. Native photorespiration was deactivated in the presence of the APs by setting the Vmax of the first native photorespiratory enzyme that was not used by the respective AP to zero. The AP designs were grouped by their CO_2 stoichiometry and the best performing pathway of each group is shown in the results. All code required to reproduce the results is available at https://gitlab.com/gain4crops/2024-paper.

Funding

We appreciate funding by the European Union H2020 Program (project GAIN4CROPS, GA no. 862087, to E.N.S, M.v.A., O.E, A.P.M.W and M.H.) and the Deutsche Forschungsgemeinschaft (Cluster of Excellence for Plant Sciences [CEPLAS] under Germany's Excellence Strategy EXC-2048/1 under project ID 390686111 to O.E and A.P.M.W).

References

[1] Thomas D Sharkey. "Estimating the rate of photorespiration in leaves". In: *Physiologia Plantarum* 73.1 (1988), pp. 147–152.

[2] Berkley J Walker et al. "The costs of photorespiration to food production now and in the future". In: *Annual review of plant biology* 67 (2016), pp. 107–129.

[3] Marco Betti et al. "Manipulating photorespiration to increase plant productivity: recent advances and perspectives for crop improvement". In: *Journal of Experimental Botany* 67.10 (2016), pp. 2977–2988.

[4] Edward N Smith et al. "Improving photosynthetic efficiency toward food security: Strategies, advances, and perspectives". In: *Molecular plant* (2023).

[5] Li-Min Wang et al. "A synthetic photorespiratory shortcut enhances photosynthesis to boost biomass and grain yield in rice". In: *Molecular Plant* 13.12 (2020), pp. 1802–1815.

[6] Rashad Kebeish et al. "Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in Arabidopsis thaliana". In: *Nature biotechnology* 25.5 (2007), pp. 593–599.

[7] Paul F South et al. "Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field". In: *Science* 363.6422 (2019), eaat9077.

[8] Deserah D Strand and Berkley J Walker. "Energetic considerations for engineering novel biochemistries in photosynthetic organisms". In: *Frontiers in Plant Science* 14 (2023), p. 1116812.

[9] Shimon Rachmilevitch, Asaph B. Cousins, and Arnold J. Bloom. "Nitrate assimilation in plant shoots depends on photorespiration". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.31 (2004), pp. 11506–11510. ISSN: 00278424.

[10] Rong Li, Maya Moore, and John King. "Investigating the regulation of one-carbon metabolism in Arabidopsis thaliana". In: *Plant and Cell Physiology* 44.3 (2003), pp. 233–241.

[11] Xinyu Fu et al. "Integrated flux and pool size analysis in plant central metabolism reveals unique roles of glycine and serine during photorespiration". In: *Nature Plants* 9 (2023), pp. 169–178. ISSN: 20550278. DOI: 10.1038/s41477-022-01294-9.

[12] Chang-Peng Xin et al. "The benefits of photorespiratory bypasses: how can they work?" In: *Plant Physiology* 167.2 (2015), pp. 574–585.

[13] Devin L Trudeau et al. "Design and in vitro realization of carbon-conserving photorespiration". In: *Proceedings of the National Academy of Sciences* 115.49 (2018), E11455–E11464.

[14] Marieke Scheffen et al. "A new-to-nature carboxylation module to improve natural and synthetic CO2 fixation". In: *Nature Catalysis* 4.2 (2021), pp. 105–115.

[15] Daniel G Marchal et al. "Machine Learning-Supported Enzyme Engineering toward Improved CO2-Fixation of Glycolyl-CoA Carboxylase". In: *ACS Synthetic Biology* 12.12 (2023), pp. 3521–3530.

[16] Josirley de FC Carvalho et al. "An engineered pathway for glyoxylate metabolism in tobacco plants aimed to avoid the release of ammonia in photorespiration". In: *BMC biotechnology* 11 (2011), pp. 1–17.

[17] Marc-Sven Roell et al. "A synthetic C4 shuttle via the β-hydroxyaspartate cycle in C3 plants". In: *Proceedings of the National Academy of Sciences* 118.21 (2021), e2022307118.

[18] Lennart Schada von Borzyskowski et al. "Marine Proteobacteria metabolize glycolate via the β-hydroxyaspartate cycle". In: *Nature* 575.7783 (2019), pp. 500–504.

[19] Alexandra Maier et al. "Transgenic introduction of a glycolate oxidative cycle into A. thaliana chloroplasts leads to growth improvement". In: *Frontiers in Plant Science* 3 (2012), p. 18890.

[20] Bo-Ran Shen et al. "Engineering a new chloroplastic photorespiratory bypass to increase photosynthetic efficiency and productivity in rice". In: *Molecular plant* 12.2 (2019), pp. 199–214.

[21] Mohamed H Aboelmy and Christoph Peterhansel. "Enzymatic characterization of Chlamydomonas reinhardtii glycolate dehydrogenase and its nearest proteobacterial homologue". In: *Plant physiology and biochemistry* 79 (2014), pp. 25–30.

[22] Sanu Shameer, R George Ratcliffe, and Lee J Sweetlove. "Leaf energy balance requires mitochondrial respiration and export of chloroplast NADPH in the light". In: *Plant Physiology* 180.4 (2019), pp. 1947–1961.

[23] Fangfang Ma et al. "Isotopically nonstationary 13C flux analysis of changes in Arabidopsis thaliana leaf metabolism due to high light acclimation". In: *Proceedings of the National Academy of Sciences* 111.47 (2014), pp. 16967–16972.

[24] Kenneth R Hanson and Richard B Peterson. "The stoichiometry of photorespiration during C3-photosynthesis is not fixed: evidence from combined physical and stereochemical methods". In: *Archives of biochemistry and biophysics* 237.2 (1985), pp. 300–313.

[25] Kenneth R Hanson and Richard B Peterson. "Regulation of photorespiration in leaves: evidence that the fraction of ribulose bisphosphate oxygenated is conserved and stoichiometry fluctuates". In: *Archives of biochemistry and biophysics* 246.1 (1986), pp. 332–346.

[26] Asaph B Cousins et al. "Peroxisomal malate dehydrogenase is not essential for photorespiration in Arabidopsis but its absence causes an increase in the stoichiometry of photorespiratory CO2 release". In: *Plant physiology* 148.2 (2008), pp. 786–795.

[27] Asaph B Cousins et al. "Peroxisomal hydroxypyruvate reductase is not essential for photorespiration in Arabidopsis but its absence causes an increase in the stoichiometry of photorespiratory CO 2 release". In: *Photosynthesis Research* 108 (2011), pp. 91–100.

[28] Florian A Busch. "Photorespiration in the context of Rubisco biochemistry, CO2 diffusion and metabolism". In: *The Plant Journal* 101.4 (2020), pp. 919–939.

[29] TL Wilkinson and AE Douglas. "Phloem amino acids and the host plant range of the polyphagous aphid, Aphis fabae". In: *Entomologia Experimentalis et Applicata* 106.2 (2003), pp. 103–113.

[30] Cristiana Gomes de Oliveira Dal'Molin et al. "A multi-tissue genome-scale metabolic modeling framework for the analysis of whole plant systems". In: *Frontiers in plant science* 6 (2015), p. 111599.

[31] John R Evans et al. "Resistances along the CO2 diffusion pathway inside leaves". In: *Journal of experimental botany* 60.8 (2009), pp. 2235–2248.

[32] Florian A Busch et al. "C3 plants enhance rates of photosynthesis by reassimilating photorespired and respired CO2". In: *Plant, cell & environment* 36.1 (2013), pp. 200–212.

[33] Danny Tholen et al. "Variable mesophyll conductance revisited: theoretical background and experimental implications". In: *Plant, Cell & Environment* 35.12 (2012), pp. 2087–2103.

[34] Danny Tholen, Gilbert Ethier, and Bernard Genty. "Mesophyll conductance with a twist". In: *Plant, Cell & Environment* 37.11 (2014), pp. 2456–2458.

[35] Xinyou Yin and Paul C Struik. "Simple generalisation of a mesophyll resistance model for various intracellular arrangements of chloroplasts and mitochondria in C 3 leaves". In: *Photosynthesis Research* 132 (2017), pp. 211–220.

[36] Urte Schlüter et al. "Brassicaceae display variation in efficiency of photorespiratory carbon-recapturing mechanisms". In: *Journal of Experimental Botany* 74.21 (2023), pp. 6631–6649. ISSN: 0022-0957. DOI: 10.1093/jxb/erad250.

[37] Olav Keerberg et al. "C2 photosynthesis generates about 3-fold elevated leaf CO2 levels in the C3–C4 intermediate species Flaveria pubescens". In: *Journal of Experimental Botany* 65.13 (July 2014), pp. 3649–3656. ISSN: 0022-0957. DOI: 10.1093/jxb/eru239. URL: https://doi.org/10.1093/jxb/eru239 (visited on 09/17/2024).

[38] Mark G. Poolman, David A. Fell, and Simon Thomas. "Modelling photosynthesis and its control". In: *Journal of Experimental Botany* 51.suppl_1 (Feb. 2000), pp. 319–328. DOI: 10.1093/jexbot/51.suppl_1.319.

[39] Akiho Yokota, Hiroshi Komura, and Shozaburo Kitaoka. "Refixation of photorespired CO2 during photosynthesis in Euglena gracilis Z". In: *Agricultural and biological chemistry* 49.11 (1985), pp. 3309–3310.

[40] Franziska Witzel, Jan Götze, and Oliver Ebenhöh. "Slow deactivation of ribulose 1, 5-bisphosphate carboxylase/oxygenase elucidated by mathematical models". In: *The FEBS Journal* 277.4 (2010), pp. 931–950.

[41] Miko U.F. Kirschbaum. "Does enhanced photosynthesis enhance growth? Lessons learned from CO2 enrichment studies". In: *Plant Physiology* 155.1 (2011), pp. 117–124. DOI: 10.1104/pp.110.166819.

[42] Christoph Peterhansel, Christian Blume, and Sascha Offermann. "Photorespiratory bypasses: how can they work?" In: *Journal of Experimental Botany* 64.3 (2013), pp. 709–715.

[43] Chandra Bellasio and Graham D. Farquhar. "A leaf-level biochemical model simulating the introduction of C2 and C4 photosynthesis in C3 rice: gains, losses and metabolite fluxes". In: *New Phytologist* 223 (2019), pp. 150–166. ISSN: 14698137. DOI: 10.1111/nph.15787.

[44] Amanda P Cavanagh et al. "Alternative pathway to photorespiration protects growth and productivity at elevated temperatures in a model crop". In: *Plant biotechnology journal* 20.4 (2022), pp. 711–721.

[45] Greta Nölke et al. “The expression of a recombinant glycolate dehydrogenase polyprotein in potato (S olanum tuberosum) plastids strongly enhances photosynthesis and tuber yield”. In: *Plant Biotechnology Journal* 12.6 (2014), pp. 734–742.

[46] Jyoti Dalal et al. “A photorespiratory bypass increases plant growth and seed yield in biofuel crop Camelina sativa”. In: *Biotechnology for Biofuels* 8.1 (2015), pp. 1–22.

[47] Anum Zeb Abbasi et al. “Expression of cyanobacterial genes enhanced CO2 assimilation and biomass production in transgenic Arabidopsis thaliana”. In: *PeerJ* 9 (2021), pp. 1–33.

[48] Ali Ebrahim et al. “COBRApy: constraints-based reconstruction and analysis for python”. In: *BMC systems biology* 7 (2013), pp. 1–6.

[49] Gosta Pettersson and Ulf Ryde-Pettersson. “A mathematical model of the Calvin photosynthesis cycle”. In: *European Journal of Biochemistry* 175.3 (1988), pp. 661–672. ISSN: 14321033. DOI: 10.1111/j.1432-1033.1988.tb14242.x.

[50] Marvin van Aalst, Oliver Ebenhöh, and Anna Matuszyńska. “Constructing and analysing dynamic models with modelbase v1. 2.3: a software update”. In: *BMC bioinformatics* 22 (2021), pp. 1–15.

[51] Anna Matuszyńska, Nima P. Saadat, and Oliver Ebenhöh. “Balancing energy supply during photosynthesis – a theoretical perspective”. In: *Physiologia Plantarum* 166.1 (May 2019), pp. 392–402. ISSN: 13993054. DOI: 10.1111/ppl.12962.

[52] Nima P Saadat et al. “Computational analysis of alternative photosynthetic electron flows linked with oxidative stress”. In: *Frontiers in plant science* 12 (2021), p. 750580.

[53] Moritz E Beber et al. “eQuilibrator 3.0: a database solution for thermodynamic constant estimation”. In: *Nucleic acids research* 50.D1 (2022), pp. D603–D609.

3.3 Explaining the photosynthetic Gibbs effect

The paper "Explaining the photosynthetic Gibbs effect" investigates the asymmetric distribution of radioactive labels in sugars produced by photosynthesis. I was involved in this project during every step of the research process, including writing the manuscript.

Publication	Explaining the photosynthetic Gibbs effect
Author position	First
Status	pre-print
doi	...

Explaining the photosynthetic Gibbs effect

Marvin van Aalst [1] and Oliver Ebenhöh [1,2]

[1] Institute of Theoretical and Quantitative Biology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[2] Cluster of Excellence on Plant Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

The photosynthetic Gibbs effect describes an asymmetric distribution of radioactive labels in sugars produced by photosynthesis. Previous arguments regarding whether these findings are in agreement with the Calvin-Benson-Bassham cycle have mostly been qualitative. Here we present a quantitative analysis of the photosynthetic Gibbs using a mathematical model of the Calvin-Benson-Bassham cycle. We show that the photosynthetic Gibbs effect is in agreement with the Calvin-Benson-Bassham cycle, providing a more detailed explanation as to why.

Correspondence: oliver.ebenhoeh@hhu.de
Code: https://gitlab.com/qtb-hhu/photosynthesis-task-force/2022-explaining-the-photosynthetic-gibbs-effect

Introduction

The Calvin-Benson-Bassham (CBB) cycle is a critical biochemical pathway integral to global carbon fixation [1]. Various photosynthetic organisms, including plants, utilize this cycle to assimilate carbon dioxide, ultimately converting it into sugars [2]. Central to this process is the enzyme rubisco, which catalyzes the carboxylation of ribulose-1,5-bisphosphate. The enzyme rubisco is arguably the most abundant enzyme on Earth, responsible for over 99 % of global carbon dioxide fixation [3, 4]. Given its pivotal role, the biochemical mechanisms underlying inorganic carbon fixation have been the focus of extensive scientific research since the inception of metabolic studies.

In the initial phase of elucidating the CBB cycle, Otto Kandler and Martin Gibbs experimentally tested Melvin Calvin's proposed carbon fixation model by employing radioactive carbon labeling. Unexpectedly, their findings revealed an asymmetric distribution of radioactive labels within starch-derived hexoses produced during photosynthesis. This asymmetry was consistent across various experimental conditions and independent of the botanical origin of the hexoses.

The observed asymmetries, now referred to as the 'Gibbs effect', demonstrated a specific labeling sequence in starch-derived hexoses: the fourth carbon position consistently acquired the radioactive label prior to the third, the first position was labeled before the sixth, and the second carbon was labeled before the fifth. In the following, these asymmetries will be abbreviated by a fraction, with the numerator showing the position labeled first, e.g. C4/C3.

In the debate regarding whether these findings are in agreement with the CBB cycle reaction scheme, the arguments have largely been qualitative and speculative in nature [5]. Thus, we present a quantitative examination of the photosynthetic Gibbs effect, building on previous work [5]. This not only gives us the ability to give an explanation for the patterns observed by Gibbs, but also to make falsifiable predictions about the label distribution of all other compounds in the CBB cycle.

Gibbs effect readily explained by CBB cycle structure

To quantitatively study the Gibbs effect with theoretical models, we implemented and adapted a published kinetic model of the CBB cycle [6, 7]. We then transformed the model to reflect all possible isotopomers using the Python package modelbase [8]. For this, every intermediate is represented by all possible 2^n isotopomer versions, where n denotes the number of carbon atoms. Likewise, all reactions are represented by all possible versions transforming the corresponding isotopomers. The resulting scheme is depicted in Fig. 1.

Due to the rapid equilibrium assumption of the fast reversible reactions in the model showed an instant equilibration of the labels in the respective intermediates, see supplementary Fig. S1. Thus, in contrast to the original model, we here explicitly determined the disequilibrium of the involved reactions. This involved the identification of forward and backward rate constants of the reversible reactions. The forward and backward rate constants can be uniquely determined from measured concentrations and the equilibrium constants [9, 10].

With this we performed simulations over time, assuming a constant supply of fully labeled carbon dioxide. To avoid biases by arbitrary choices of time points to compare the asymmetry, we took the integral of the difference between the labelled carbons as a measurement of the asymmetry.

Fig. 2 shows the labeling patterns of fructose-1,6-bisphosphate (FBP) and fructose 6-phosphate (F6P) (see supplementary Fig. S2 for all other metabolites). In both hexoses the fourth carbon is labeled before the third. However, while in F6P the first carbon is labeled before the sixth and the second carbon before the fifth, this pattern is reversed in FBP. The label asymmetries in F6P originally described in experiments by Gibbs can be observed without any further parameter adjustments to the model.

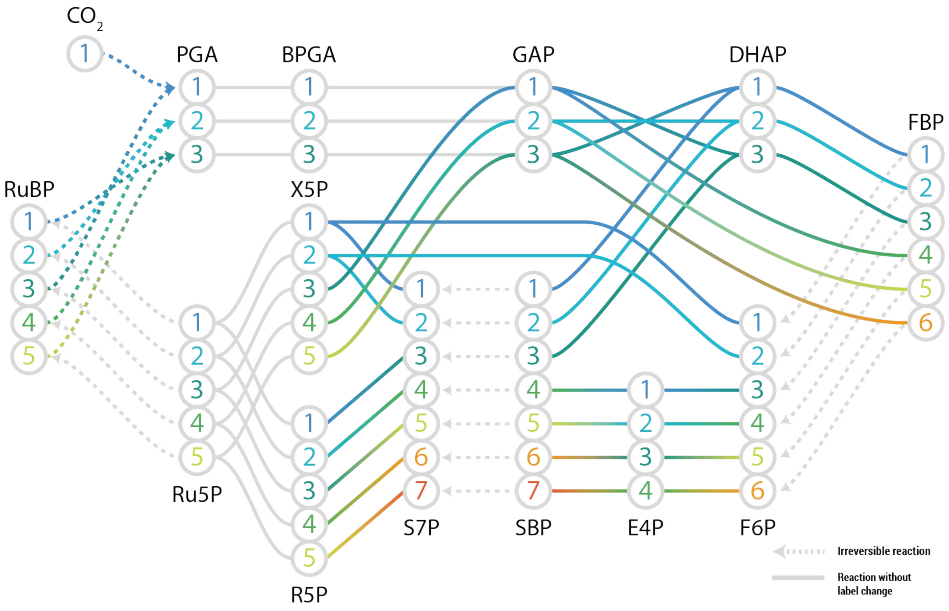


Figure 1: The path of carbon atoms in the CBB cycle. Shown are the carbon positions of CBB cycle intermediates and the mappings of the corresponding reactions. Reactions changing the label position between the reactants are coloured, irreversible reactions are drawn with dashed lines. Triose phosphate exporters, pathways to starch production and enzyme names are omitted for clarity.

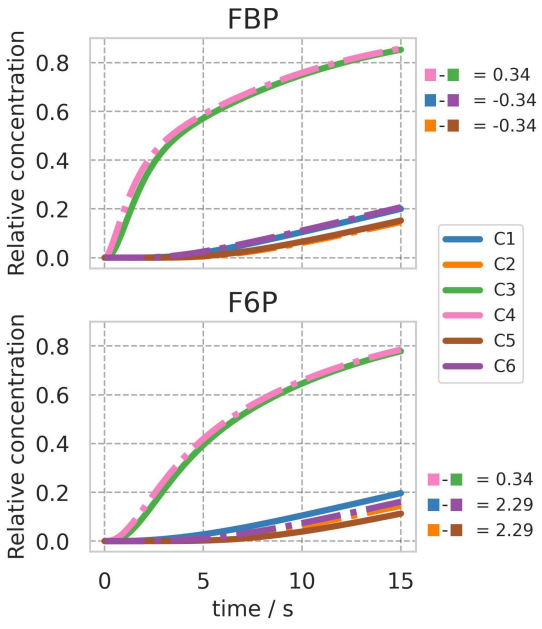


Figure 2: Time development of label distribution in FBP and F6P shown relative to the concentration of the respective metabolite. C1 to C6 denote the respective carbon atoms.

In FBP however, the C1/C6 and C2/C5 asymmetries are reversed. To investigate what causes the contrast between the C1/C6 and C2/C5 label asymmetries in FBP and F6P, we systematically varied the enzyme activity of all CBB cycle enzymes. We then calculated the respective label asymmetries for F6P, shown in Fig. 3 for a selection of enzymes (see supplementary Fig. S4 for all CBB cycle enzymes). The C4/C3 asymmetry in F6P is mostly affected by triose-phosphate isomerase (TPI), which shows an antiproportional effect. For example, at half the TPI activity the C4/C3 asymmetry is 150 % relative to the wildtype and at twice the TPI acitivity the asymmetry is 57 %. Increasing sedoheptulose-bisphosphatase (SBPase) also slightly decreases the C4/C3 asymmetry, which at fivefold activity is at 89 % of the wildtype asymmetry. As with the C4/C3 asymmetry, an increase in TPI activity also decreases the C1/C6 and C2/C5 asymmetries. This is also the case for SBPase and phosphoribulosekinase (PRK), which reduce the C2/C5 and C1/C6 asymmetries. At fivefold SBPase activity both C2/C5 and C1/C6 asymmetries are at 31 % of the wildtype activity. Similarly, at fivefold PRK activity both asymmetries are at 16 % of the wildtype asymmetry. Contrastingly, increasing transketolase increases the C1/C6 and C2/C5 asymmetries, which at double the activity are at 200 % of their respective wildtype asymmetry.

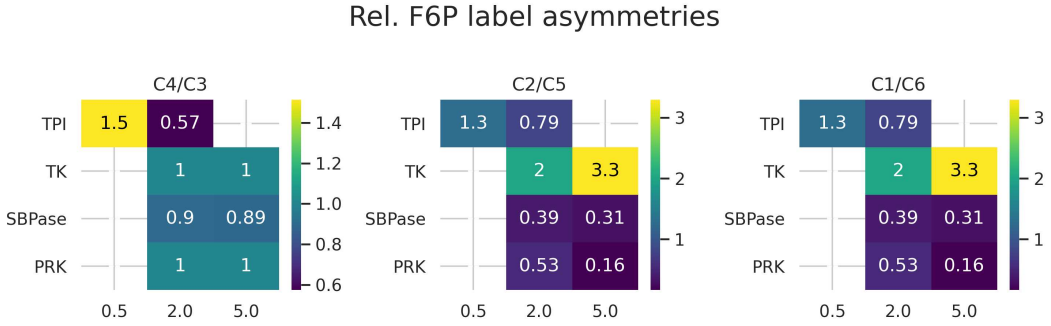


Figure 3: Systematic scan of relative F6P label asymmetries dependent on the activity of selected CBB cycle enzymes. The asymmetries are calculated as the integral of the difference between the respective time series and shows relative to the wildtype asymmetries. The x-axis shows the scaling factor of the $V_{\max}$ and the y-axis denotes the respective enzyme.

Octuloses might influence labeling patterns

The model in its current form readily explains the photosynthetic Gibbs effect. However, there is a long standing debate about additional reactions catalysed by the promiscuous enzymes transketolase and aldolase [11]. These additional reactions allow production of D-glycero-D-altro-octulose 8-phosphate (O8P), D-erythro-D-gluco-octose α-1,8-bisphosphate (OBP) and aldehydo-D-arabinose 5-phosphate (A5P) [12, 13]. All these metabolites have been detected experimentally in both photosynthetic and heterotrophic organisms [14, 15]. However, the significance of the flux through the reactions producing those metabolites is still debated [16]. We thus sought to quantify the effect those reactions might have on the labeling distribution. For this, we included the reactions in Table 1. This included ways to produce A5P or consume OBP further for which we included an A5P isomerase which catal-

Non-standard transketolase reactions

Substrates	Products
F6P + R5P	S7P + E4P
X5P + G6P	O8P + GAP
F6P + G6P	O8P + E4P
S7P + G6P	O8P + R5P

Neutral transketolase reactions

Substrates	Products
X5P + GAP	X5P + GAP
F6P + E4P	F6P + E4P
S7P + R5P	S7P + R5P
O8P + G6P	O8P + G6P

Non-standard aldolase reactions

Substrates	Products
DHAP + A5P	OBP

Table 1: Stoichiometries of extended transketolase and aldolase reactions.

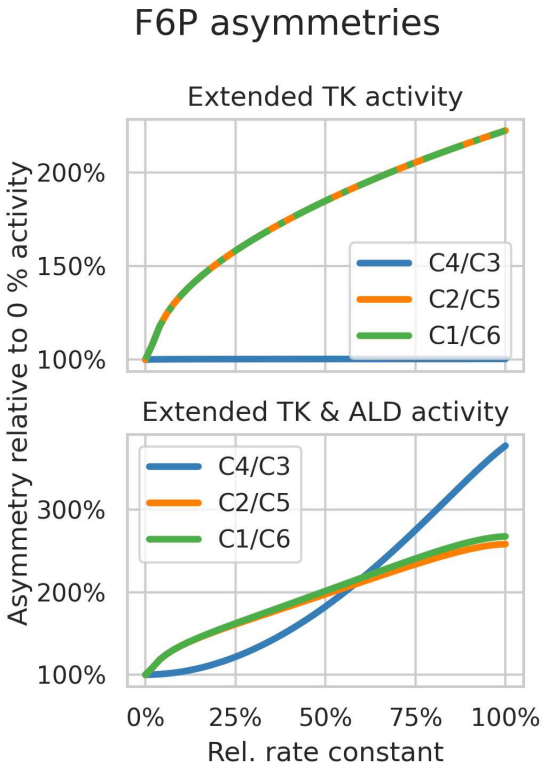


Figure 4: Relative F6P asymmetries dependent on the rate constant the extended transketolase reactions (upper panel) or both transketolase and aldolase reactions (lower panel). Rate constants are shows relative to to the respective rate constant of the standard CBB cycle reaction.

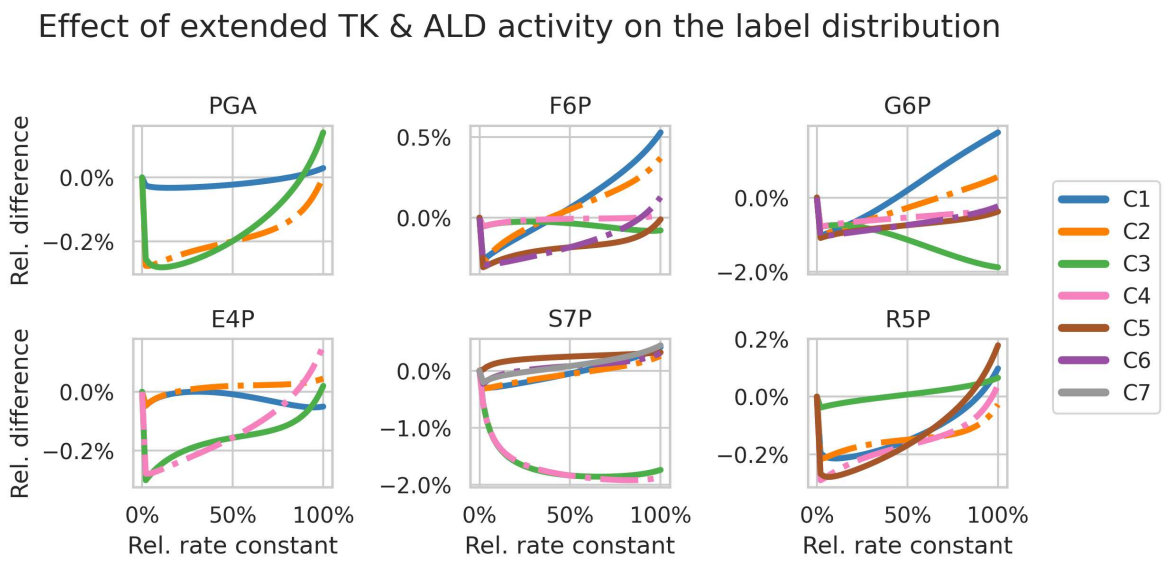


Figure 5: Effect of extended transketolase and aldolase on the labeling pattern of selected CBB cycle intermediates. The effect is shown as the integral of the difference between the time course of the respective carbon at a given relative rate constant against the time course at wildtype conditions. The rate constants are shown relative to the value of their respective wildtype reaction rate constants.

yses the reaction A5P <=> RU5P and an OBPase, which catalyzes the reaction OBP + H2O <=> O8P + Pi. The respective rate equations of transketolase and aldolase were adapted using reversible kinetics to ensure thermodynamically sound behaviour [10]. We assumed OBPase to be irreversible and thus adapted the rate equation of SBPase instead. For all the newly added reactions we added scaling factors, as we could not find any kinetic data for them.

We systematically scanned the label distributions for model variants containing the extended transketolase reactions, the extended aldolase reactions or both. The C1/C6 and C2/C5 asymmetries are increased more than 200 % if the relative activity of the extended transketolase reactions matches the one for the standard ones, see upper panel of Fig. 4. If the extended aldolase reaction is also scaled proportionally, the C4/C3 asymmetry increases to over 300 % and the C1/C6 asymmetry is also further increased, see lower panel of Fig. 4.

As the inclusion of extended transketolase and aldolase reactions can increase the Gibbs effect we sought to further quantify the general effect on the label distribution in the CBB cycle. The effect of the extended transketolase and aldolase reactions on the labeling patterns of select CBB cycle intermediates is shown in Fig. 5 systematically for rate constants between 0-100 % of the respective wildtype rate constants (see supplementary Fig. S7 for all CBB cycle intermediates). There we took the integral of the relative difference between the time course at a specific relative rate constant and the one at zero activity as a metric for the change of a given label. Even at 100 % relative rate constant of the extended transketolase and aldolase reactions, the magnitude of the relative difference is less than 5 % (see glucose-1-phosphate (G1P) in supplementary Fig. S7).

Of special interest are the four carbons in the CBB cycle which with the additional reactions require fewer enzymatic steps to be labelled. For example, with the addition of the extended reactions the first carbon of glucose-6-phosphate (G6P) only requires three instead of seven enzymatic steps to be labelled starting from DHAP. This can be achieved by the reaction sequence of aldolase producing OBP, OBPase producing O8P and transketolase running backwards. Similar pathways can be found for the first carbon of F6P, requiring seven instead of nine steps, the second carbon of G6P, requiring 11 instead of 13 steps and the first carbon of sedoheptulose-7-phosphate (S7P), requiring eight instead of nine steps. At 100 % relative rate constant the first carbon of G6P is labelled 1 % more. Similarly, the first carbon of F6P is labelled 0.5 % more and the label of the second carbon of G6P and the first carbon of S7P are also increased.

However, these are not the only effects the addition of the extended reaction has on the label distribution. For a low relative rate constant, the third carbon of PGA is labelled less, which is alleviated at high relative activitiy by aldolase activity (see supplementary Fig. S5 for a comparison to just the transketolase activity). This effect propagates to the fourth and third carbon of erythrose-4-phosphate (E4P), the sixth and seventh carbon of the heptuloses and the fifth and

fourth carbon of the pentoses respectively. Similarly, the increased C4/C3 asymmetry in FBP can be observed in the second and first carbon of E4P.

At a high relative rate constant the third carbon of G6P the change in the label of the third carbon is more negative than the one of F6P. The third carbon of G6P can only be derived from the third carbon of F6P or the second carbon of A5P. This suggests that the same mechanism causing the increase in the labeling of the first carbon decreases the labeling of this carbon.

Carbons three and four of sedoheptulose-1-7-bisphosphate (SBP) and S7P are affected differently by an increased relative rate constant of the extended reactions. While there is almost no change in those carbons in SBP (see supplementary Fig. S7), they are reduced by roughly 2 % in S7P. In the extended transketolase reactions the third and fourth S7P carbons are derived from the first and second carbons of ribose-5-phosphate (R5P), which also happens in the standard transketolase reactions. Thus, this effect is caused by an increase in transketolase activity in general and not by the extended reactions per se.

Discussion & Conclusions

Our model, without any parameter fitting, qualitatively reproduces the asymmetries experimentally measured by Gibbs. That is, we can observe that in F6P the fourth carbon is labeled before the third, the first before the sixth and the second before the fifth, see Fig. 2.

Bassham noted in 1964 that the asymmetry in the 4 and 3 positions can be explained by the different pool sizes of the triose phosphates [17]. We argue here that regardless of the pool sizes the C4/C3 asymmetry will always be observed. It is true that TPI activity and thus different pool sizes of GAP and DHAP can change the magnitude of the asymmetry. However, the set of reactions necessary to label the fourth carbon of FBP is a strict subset of the reactions necessary to label the third carbon of FBP. As the relative outflux of a label from one metabolite cannot exceed the relative amount of label in the respective pool, it is thus structurally impossible for the third carbon to be labeled more than the fourth.

While the C4/C3 asymmetry is present in FBP as well, the C1/C6 and C2/C5 asymmetries are reversed. This difference can readily be explained by transketolase activity, as Bassham already noted in 1964 [17]. In order for the first and sixth carbon of FBP to be labeled, the third carbon of GAP needs to be labeled. If CO_2 is the only label influx to the system, this requires the label to be passed through the entire cycle once, via the third carbon of SBP, which ends up in the third carbon of GAP. The difference between the labeling of the first and sixth position can then be explained exactly like the C4/C3 asymmetry above.

In F6P however, there exists a second influx route for labels into the first two carbons via transketolase. This just requires six reactions from DHAP to take place, while the first label in FBP requires ten reactions of which four are shared with the alternative route. As Fig. 3 shows, increased transketolase activity increases the C1/C6 label asymmetry, giving more weight to this explanation. While this asymmetry in principle can be reversed by very low transketolase activity the model does not yield stable solutions for those cases. An analogous explanation can be given for the C2/C5 asymmetry.

We next extended the model to include non-standard transketolase and aldolase reactions, capable of producing octuloses. While octuloses have been detected experimentally, the significance of the flux through them still debated [14, 15]. As kinetic parameters for these reactions are unknown we systematically scanned possible rate constants between 0-100 % of the rate constants of the respectively standard transketolase and aldolase reactions. The inclusion of extended transketolase and aldolase reactions increased the magnitude of the Gibbs effect by up to three-fold, see Fig. 4. However, when the relative changes of all the labels in the CBB cycle was assessed, we found no change larger than ± 5 % (Fig. 5). This suggests both that small changes in labelling can lead to large changes in asymmetries and that inclusion of this larger scheme in approaches like MFA might not be necessary, as the difference is likely going to be small [18]. However, it is theoretically possible for the rate constants to be higher than the bounds we chose. More experimental data is required to increase the confidence in the predictions made.

In order to estimate the fluxes through the extended reactions we identified asymmetries and label patterns as useful targets. One such target is the first carbon of G6P, which with the addition of the extended reactions just requires three instead of seven enzymatic steps from DHAP to be labelled - all of which are extended reactions. Thus, deviation from the expected labelling can be a useful hint for the total flux through the extended reactions. Similarly, the third carbon of G6P is predicted to be labelled less due to and only due to the extended reactions. Here, again, the total flux can be estimated from a deviation from the expected label.

However, as the labelling pattern of carbons three and four of S7P shows, those estimates require a good estimate of the transketolase activity in general, as changes in transketolase activity already cause changes in a multitude of labelling patterns. Thus, when assessing the differences caused by the extended reactions it is important to distinguish them from changes caused by increased total transketolase and aldolase activity. Further theoretical work is required to isolate these two effects.

Methods

$$k^+ = \frac{\frac{v}{\prod_i S_i}}{1 - \frac{Q}{K_{eq}}}$$

and respectively

$$k^- = -\frac{\frac{v}{\prod_j P_j}}{1 - \frac{K_{eq}}{Q}}$$

With the new obtained mass action rate parameters the corresponding reactions were brought into standard mass action format

$$v = k^+ \prod_i S_i - k^- \prod_j P_j \tag{1}$$

Similiarly, in order to fit the regularised reactions that follow the canonical expression

$$v = \frac{V_{Max} \cdot [S]}{[S] + K_m \cdot R(S)} \tag{2}$$

the V_{Max} values were obtained using

$$V_{Max} = \frac{v\left([S] + K_m \cdot R(S)\right)}{[S]} \tag{3}$$

and the experimentally measured concentrations.

Funding

This work was 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.).

Funding

We appreciate funding by the European Union H2020 Program (project GAIN4CROPS, GA no. 862087, to M.v.A. & O.E) and the Deutsche Forschungsgemeinschaft (Cluster of Excellence for Plant Sciences [CEPLAS] under Germany's Excellence Strategy EXC-2048/1 under project ID 390686111 to O.E).

References

[1] John A Raven. "Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments". In: *Aquatic Microbial Ecology* 56.2-3 (2009), pp. 177–192.

[2] James A Bassham and Melvin Calvin. *The path of carbon in photosynthesis*. Springer, 1960.

[3] R.John Ellis. "The most abundant protein in the world". In: *Trends in Biochemical Sciences* 4.11 (Nov. 1979), pp. 241–244. ISSN: 0968-0004. DOI: 10.1016/0968-0004(79)90212-3. URL: https://www.sciencedirect.com/science/article/pii/0968000479902123.

[4] John A Raven. "Rubisco: still the most abundant protein of Earth?" In: *New Phytologist* 198.1 (2013), pp. 1–3.

[5] Oliver Ebenhöh and Stephanie Spelberg. "The importance of the photosynthetic Gibbs effect in the elucidation of the Calvin-Benson-Bassham cycle." In: *Biochemical Society transactions* 46.1 (Feb. 2018), BST20170245. ISSN: 1470-8752. DOI: 10.1042/BST20170245. URL: http://www.ncbi.nlm.nih.gov/pubmed/29305411%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5818666%20http://biochemsoctrans.org/lookup/doi/10.1042/BST20170245.

[6] Gosta Pettersson and Ulf Ryde-Pettersson. "A mathematical model of the Calvin photosynthesis cycle". In: *European Journal of Biochemistry* 175.3 (1988), pp. 661–672. ISSN: 14321033. DOI: 10.1111/j.1432-1033.1988.tb14242.x.

[7] Mark G. Poolman, David A. Fell, and Simon Thomas. "Modelling photosynthesis and its control". In: *Journal of Experimental Botany* 51.suppl_1 (Feb. 2000), pp. 319–328. ISSN: 1460-2431. DOI: 10.1093/jexbot/51.suppl_1.319. URL: https://academic.oup.com/jxb/article-lookup/doi/10.1093/jexbot/51.suppl%7B%5C_%7D1.319.

[8] Marvin van Aalst, Oliver Ebenhöh, and Anna Matuszyńska. "Constructing and analysing dynamic models with modelbase v1. 2.3: a software update". In: *BMC bioinformatics* 22.1 (2021), pp. 1–15.

[9] J A Bassham and G H Krause. "Free energy changes and metabolic regulation in steady-state photosynthetic carbon reduction." In: *Biochimica et Biophysica Acta* 45.868 (1969), pp. 207–221. ISSN: 00052728. DOI: 10.1016/0005-2728(69)90048-6. URL: http://linkinghub.elsevier.com/retrieve/pii/0005272869900486%7B%5C%%7D5Cnhttp://dx.doi.org/10.1016/0005-2728(69)90048-6.

[10] Moritz E Beber et al. "eQuilibrator 3.0: a database solution for thermodynamic constant estimation". In: *Nucleic Acids Research* 50.D1 (Nov. 2021), pp. D603–D609. ISSN: 0305-1048. DOI: 10.1093/nar/gkab1106. eprint: https://academic.oup.com/nar/article-pdf/50/D1/D603/42057662/gkab1106.pdf. URL: https://doi.org/10.1093/nar/gkab1106.

[11] Thomas D Sharkey. "Pentose phosphate pathway reactions in photosynthesizing cells". In: *Cells* 10.6 (2021), p. 1547.

[12] MG Clark, JF Williams, and PF Blackmore. "The transketolase exchange reaction in vitro." In: *Biochemical Journal* 125.1 (1971), p. 381.

[13] BL Horecker, Francesco Paoletti, and John F Williams. "Occurrence and significance of octulose phosphates in liver". In: *Annals of the New York Academy of Sciences* 378.1 (1982), pp. 215–224.

[14] John F Williams and John K MacLeod. "The metabolic significance of octulose phosphates in the photosynthetic carbon reduction cycle in spinach". In: *Photosynthesis Research* 90 (2006), pp. 125–148.

[15] Ian L Flanigan, John K MacLeod, and John F Williams. "A re-investigation of the path of carbon in photosynthesis utilizing GC/MS methodology. Unequivocal verification of the participation of octulose phosphates in the pathway". In: *Photosynthesis Research* 90 (2006), pp. 149–159.

[16] Anna Stincone et al. "The return of metabolism: biochemistry and physiology of the pentose phosphate pathway". In: *Biological Reviews* 90.3 (2015), pp. 927–963.

[17] J A Bassham. "Kinetic Studies of the Photosynthetic Carbon Reduction Cycle". In: *Annual Review of Plant Physiology* 15.1 (June 1964), pp. 101–120. ISSN: 0066-4294. DOI: 10.1146/annurev.pp.15.060164.000533. URL: http://www.annualreviews.org/doi/10.1146/annurev.pp.15.060164.000533.

[18] Yujue Wang et al. "Metabolic flux analysis—linking isotope labeling and metabolic fluxes". In: *Metabolites* 10.11 (2020), p. 447.

3.4 Approximating carbon fixation - how important is the Calvin-Benson cycle steady-state assumption?

The paper "Approximating carbon fixation - how important is the Calvin-Benson cycle steady-state assumption?" investigates whether a steady-state assumption for the CBB cycle is realistic in the context of fluctuating light environments. I was involved in this project during every step of the research process, including writing the manuscript.

Publication	Approximating carbon fixation - how important is the Calvin-Benson cycle steady-state assumption?
Author position	First
Status	pre-print
doi	10.1101/2022.11.18.517021

Approximating carbon fixation - how important is the Calvin-Benson cycle steady-state assumption?

Marvin van Aalst [1], **Oliver Ebenhöh** [1,2], **and Berkley J. Walker** [3,4]

[1] Institute of Theoretical and Quantitative Biology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany; [2] Cluster of Excellence on Plant Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany; [3] Department of Energy-Plant Research Laboratory, Michigan State University, East Lansing, MI, USA; [4] Department of Plant Biology, Michigan State University, East Lansing, MI, USA

Plants use light energy to produce ATP and redox equivalents for metabolism. Since during the course of a day plants are exposed to constantly fluctuating light, the supply of ATP and redox equivalents is also fluctuating. Further, if the metabolism cannot use all of the supplied energy, the excess absorbed energy can damage the plant in the form of reactive oxygen species. It is thus reasonable to assume that the metabolism downstream of the energy supply is dynamic and as being capable of dampening sudden spikes in supply is advantageous, it is further reasonable to assume that the immediate downstream metabolism is flexible as well. A flexible metabolism exposed to a fluctuating input is unlikely to be in metabolic steady-state, yet a lot of mathematical models for carbon fixation assume one for the Calvin-Benson-Bassham (CBB) cycle. Here we present an analysis of the validity of this assumption by progressively simplifying an existing model of photosynthesis and carbon fixation.

mathematical model | photosynthesis | calvin cycle | steady-state

Correspondence: ***marvin.van.aalst@hhu.de***

Introduction

The light reactions of photosynthesis must provide the chemical energy (ATP and NADPH) needed for CO_2 fixation through the CBB cycle under rapidly fluctuating light availability. Light fluctuates across many time scales under natural conditions, greatly complicating the challenge plants have in harvesting sufficient light energy for optimal rates of carbon fixation [1, 2, 3, 4]. Plants possess a myriad of physiological responses to changing light intensity such as variable stomatal conductance, regulation of CBB enzymes and possibly even the activation of photorespiratory genes [5]. While these factors are critical for understanding the integrated response of net assimilation to fluctuating light, we have focused this investigation on dissecting out the important assumptions for modeling the CBB and photosynthetic electron transfer chain (pETC) using reaction kinetic models.

There have been many excellent metabolic models representing the CBB and associated pETC activity using various frameworks that can represent steady-state or dynamic behavior [6, 7, 8, 9, 10]. In this paper we use a metabolic definition of steady-state, specifically that under steady-state metabolite pool sizes are constant as well as all input and output fluxes. We use a corresponding metabolic definition for dynamic in that it is a condition where metabolite pool sizes and internal fluxes are changing. Note that this is a slightly different definition of a physiological dynamic model, which can represent components like the slow relaxation of non-photochemical quenching, changes in stomatal conductance and rubisco activation state [11]. In this work we ask how important it is to consider the dynamic response of the CBB during light fluctuations and how well the commonly applied steady-state assumption can represent metabolism.

The fundamental aim of mathematical modeling is to enhance understanding of the system studied, as B.D. Hahn already noted in 1993 [12]. Thirty years later much effort is still being put into building ever more complicated models with high fidelity, which are increasingly difficult to understand. Notably, Hahn considered models with 17-31 non-linear differential equations to be large while nowadays genome-scale modeling techniques are frequently employed, which can contain thousands of reactions [13, 14, 15, 16]. To revisit the spirit of trying to understand the key features of a system we employ model reduction of a previously published model of the pETC and the CBB cycle to elucidate the main processes controlling carbon fixation in C3 plants under dynamic conditions [17]. The resulting simplified models can serve as robust alternatives in situations where few parameters are known and the research question is only concerned with carbon fixation rate, as their predictions are in very good agreement with the predictions of the original complex model. The three consecutive reductions we employ are first replacing the dynamic system behavior with a steady-state approximation,

then replacing the steady-state system with a polynomial fitted to the predicted carbon dioxide fixation rate and lastly reducing the amount of data that is fed into the model. We find that a good prediction of carbon fixation rate requires a remarkably low amount of model fidelity as well as data, while prediction of other fluxes requires a much better description of the underlying biochemistry.

Results

To determine the required fidelity to represent the CBB cycle, pETC and photo-protective mechanisms, we first compared simulations of a detailed ordinary differential equation (ODE) model with and without a simplifying metabolic steady-state assumption [17, 7]. For this first simplification we removed the dynamic change of metabolite concentrations by precalculating steady-state fluxes of an ODE model in response to realistic field conditions using incoming photosynthetically active photon flux density (PPFD) measured in one minute intervals at a National Ecological Observatory Network site in Washington state, USA [18]. We then compared the simulated rate of carbon fixation of the dynamic and the steady-state model for a 6 hour window from a typical summer day (shown in Figure 1). There is generally good agreement between the simulation of carbon fixation (ribulose-1,5-bisphosphate carboxylase-oxygenase (rubisco) flux) between the two approaches. While the relative error in short periods can exceed 50 %, the total error of predicting the rubisco flux is 0.49 %.

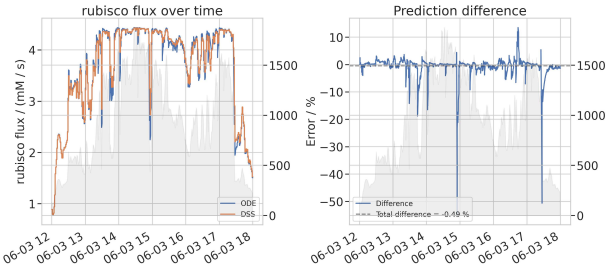


Figure 1. Comparison of ODE and steady-state approximated model predictions over a dynamic light signal (grey area). The left subplot shows the rubisco flux predicted by the ODE and the steady-state approximated model respectively over a time course of 6 hours. The right subplot shows the error of the steady-state approximated model rubisco flux predictions relative to ODE model predictions.

To understand why the steady-state model behaves similarly to the dynamic model with regard to total carbon fixation despite the lack of dynamic interaction we investigated how well other fluxes are predicted and how much the concentrations change over the course of the experiment. Shown in Figure 2 is the absolute total difference between the simulations of key CBB, pETC and acclimation fluxes and the relative standard deviation (coefficient of variation) of representative concentrations in the dynamic model. The remaining fluxes and concentrations are

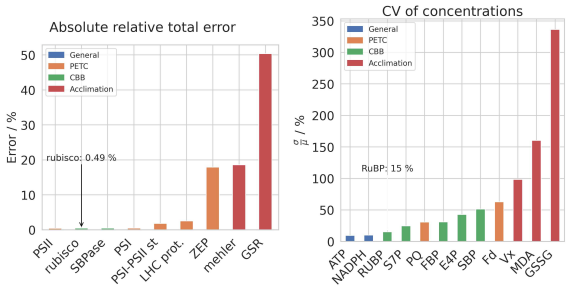


Figure 2. Left: absolute total error of steady-state approximated model predictions compared to ODE model predictions. Right: relative standard deviation (coefficient of variation) of ODE model predicted metabolite concentrations over a course of a 6-hour experiment.

displayed in supplementary figures S1 and S2 respectively and the relative standard deviation for the fluxes in supplementary Figure S3. In contrast to the the small difference between the rubisco flux in the simulations, the difference between rapid acclimation response mechanisms such as the xanthophyll cycle (zeaxanthin epoxidase) or reactions of the water-water cycle (glutathione reductase) is between 20 % and 50 %. Similarly, the relative standard deviation of the rubisco substrate ribulose-1,5-bisphosphate (RuBP) is comparatively small, varying 15 % of its mean value compared to the large relative standard deviation of some metabolites that take part in the rapid acclimation response, e.g. glutathione-disulfide (GSSG) and dehydroascorbate (DHA) which vary nearly up to 350 % of their mean value. Notably, other CBB cycle intermediates like sedoheptulose-1-7-bisphosphate (SBP) and sedoheptulose-7-phosphate (S7P) vary up to 50 % while the difference between the flux predictions of sedoheptulose-bisphosphatase (SBPase) is comparable to the one of rubisco.

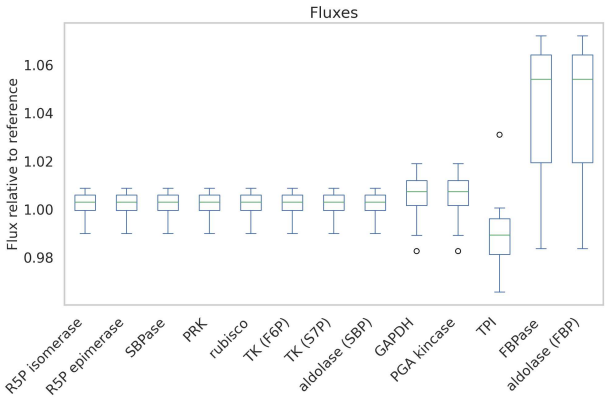


Figure 3. Box-plot of CBB cycle fluxes in model collection with randomly perturbed kinetic parameters relative to the fluxes in model with reference parameters.

To understand how alternative parameterizations of the model that lead to similar carbon fixation rates would effect internal fluxes of the CBB cycle we generated 100,000 sets of randomly perturbed kinetic parameters. For these parameter-sets we calculated the steady state flux and then selected the ones for which the carbon fixation flux was within 1 % of the original

model, which was the case for 321 of them. Figure 3 shows a box-plot of simulated CBB cycle fluxes of the selected parameter sets relative to the dynamic model, the remaining fluxes are shown in supplementary Figure S7. Some reactions of the CBB cycle must vary less than 10 % to ensure a similar flux, while e.g. triose-phosphate isomerase (TPI), fructose-1,6-bisphophatase (FBPase) and fructose-bisphosphate aldolase (ALDOA) can vary more than 30 %. Reactions related to triose phosphate export and storage can vary upwards of 100 %.

Given the stability of the RuBP concentration during our simulations across irradiances and the insensitivity of rubisco flux to parameterization of many CBB model reactions, we hypothesized that it is possible to vastly simplify the CBB cycle and still get a reasonable prediction. To describe the carbon fixation rate as being only dependent on irradiance, we replaced the explicit CBB model with a 4th degree polynomial function which we fitted over simulated carbon fixation fluxes. The agreement between these two models was very good, with a very small root-mean-square error (RMSE) of 0.02 $\frac{mM}{s}$ (shown in supplementary Figure S4). When simulated over the same data shown in Figure 1, the difference between the total rubisco flux of the polynomial model and the dynamic model ODE fluxes is 0.64 % and thus comparable to the difference of 0.49 % between the steady-state model and the dynamic model (shown in supplementary Figure S5).

So far we have used high temporal resolution data in our simulations, but data availability, lack of computational power or technical difficulties due to multi-scale modeling can require the use of low temporal resolution data. A common practice to reduce the temporal resolution is to simply average the PPFD values over longer time periods. In our experiment, an increase from one minute to 60 minute steps increased the difference of carbon fixation from 0.49 % to roughly 7 % (see supplementary Figure S10).

While an increase in error with fewer data was to be expected, we hypothesized that the simulation can be improved substantially by using an alternative averaging approach. If carbon assimilation responded linearly to irradiance, carbon assimilation over the average irradiance of a time period would equal the average of carbon assimilation over that time period. In other words, irradiance could be averaged over a time period with fluctuating light and simulate an equal total amount of carbon fixation as the fluctuating light period. However, as carbon fixation responds non-linearly to irradiance, the simulated carbon fixation of the average irradiance does not lead to the average of the simulated carbon fixation over fluctuating light. In the case of the saturating response of carbon fixation to irradiance, simple averaging would lead to over-estimates over periods of fluctuating irradiance.

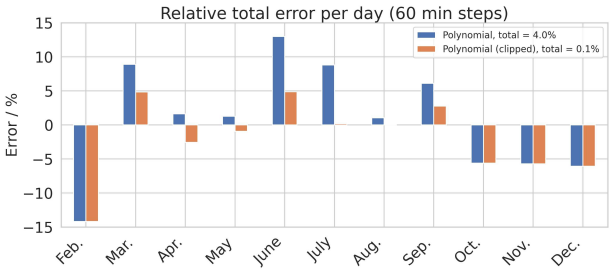


Figure 4. Total prediction error of carbon fixation of the steady-state approximated model (60 minute resolution) and the polynomial model (60 minute resolution) relative to the ODE model (1 minute resolution). Results are shown for both the mean PPFD value of the raw data and the mean value of the data clipped at either PPFD 900 (steady-state approximated model) or 1000 (polynomial model).

To produce representative PPFD values that avoid the bias of simple averaging over each time period, we clipped the input data by capping saturating PPFD values to reduce their effect before calculating the mean. As this approach led to promising results with our single-day data we expanded the analysis to a representative day for each month of the entire year, and then simulated the carbon fixation rate of this day for each month with one minute steps for the ODE model and 60 minute steps for the polynomial model.

For the polynomial model we included both the prediction with the mean of the raw data and the mean of the clipped data. Figure 4 shows the total error of the rubisco flux prediction of the polynomial model relative to the ODE model per month. As indicated in the figure legend, the total error per year of the polynomial model relatives to the ODE model is 4.0 % for raw data and 0.1 % for data clipped at a PPFD of 1000. The absolute of the error of the polynomial model for October to February is between 5 and 15 %, while it is lower for the summer months (except June), however if the data is clipped at a PPFD of 1000 there is an improved fit relative to the polynomial model for the winter months.

Discussion & Conclusions

Our simulations reveal that accounting for the dynamic change of metabolite concentrations under fluctuating light results in only small differences in the predicted carbon fixation rate relative to assuming a metabolic steady state in a combined model including the CBB and light reactions. Specifically, in Figure 1 we compared predictions of an ODE model with a simplified version in which we removed the dynamic change of metabolite concentrations and instead precalculated the steady-state fluxes for a range of PPFD inputs. Despite the rapid light fluctuations in the input data, our results show that the predicted total carbon fixation essentially stays the same, even though the difference of the prediction of the dynamic state can exceed 50 % for some time steps. This means

that for the purpose of predicting total carbon fixation, our ODE model of carbon fixation can be greatly simplified using steady-state assumptions with little difference - saving time and energy by avoiding computationally costly numerical integration. We expect this to hold for other ODE models of biological systems that are structured similarly. Next we investigated how such a simplified model, which does a mechanistically poor job at representing the underlying biochemistry, does a good job at representing carbon fixation. For this we turned to other biochemical predictions of the model.

While total carbon fixation was predicted well in the steady-state model, the fluxes of rapid acclimation mechanisms (e.g. violaxanthin deepoxidase) were predicted poorly, with relative errors exceeding 50 % (see Figure 2). These observations can be explained by the amount of dynamic change in concentration over time of specific metabolite pools that directly influence carbon fixation. For example, the RuBP pool is comparatively stable (varying 15 % of its mean concentration), while the GSSG pool can vary over 350 % of its mean concentration (see Figure 2). Since the RuBP pool comprises the only dynamic substrate for carbon fixation in the original model it has the largest effect on rates of rubisco carboxylation, reflected by the small relative error in this reaction of 0.49 %. The higher errors of acclimation and pETC intermediates in contrast contribute very little to the error of rubisco carboxylation, as they mostly depend on prior perturbations of the system but always tend towards stabilizing the system towards the respective steady state. Notably, other CBB cycle intermediates like SBP can also show a variation of more than 50 % of their mean concentration, but these errors are not reflected in the downstream concentrations of RuBP, resulting in less differences in rates of rubisco carboxylation.

One explanation for the different variation over time in metabolite pool sizes is that multiple flux distributions can lead to the same carbon fixation rate, thus allowing a wider range of concentrations for certain CBB cycle intermediates. This buffering effect can be seen in Figure 3, where in models with perturbed kinetic parameters but similar rates of carbon fixation some reactions like FBPase can vary more than a third of their reference flux. This suggests that the CBB cycle is structured such that temporary changes in illumination are buffered by the cycle for RuBP concentration to remain relatively constant, resulting in similar rates of carbon fixation, effectively working as a low-pass filter (see supplementary figures S8 and S13 and supplementary Figure S16 to Figure S45). Due to this stability, carbon fixation can be accurately simulated, even if the underlying metabolism is not. While the stability of RuBP is advantageous for downstream metabolism this also implies that alternative energy sinks like quenching mechanisms and the water-water cycle (WWC) are required to dissipate further excess energy. The reactions with the most variation relate to sucrose and starch partitioning (e.g. FBPase, FBP aldolase and TPI), suggesting that while carbon fixation is simulated accurately, downstream carbon partitioning may not be. These findings suggest that carbon fixation can be simulated accurately in an even more simplified CBB cycle, providing downstream carbon partitioning is not of interest. Our findings that an explicit model of the CBB cycle can be replaced with a simple polynomial model with little sacrifice, see Figure 4, are in line with many crop-systems models that represent net carbon assimilation using a simple radiation use efficiency (e.g.

The discussion above highlights that metabolic models of the CBB cycle can produce similar rates of carbon fixation despite mechanistically curedely simplified assumptions of the underlying biochemistry. This finding indicates that care is needed when validating a model using only carbon fixation rates, since any model that keeps RuBP concentration stable can lead to realistic carbon fixation rates. Possible improvements include using predictions of other metabolite pools ([6]) or fluorescence data if the model also contains the pETC ([17]). Further work can test the ability to make steady-state assumptions with more complex models of carbon assimilation that include for example photorespiration, CO2 as a dynamic variable or dynamic temperature. Photorespiration may present an interesting case since large pools of glycine accumulate during photorespiratory induction, effectively decreasing relative rates of glycine decarboxylation (and increasing net carbon exchange) during this transient by up to 40 % [19].

We show that temporally high-resolution PPFD data is not required to give a good prediction of total carbon fixation as reducing the amount of data by 60-fold still led to an prediction error of less than 1 % if over modified averaging approach is used, see Figure 4. The ability to properly aggregate data is important, since often all data needed to produce and validate a model are not available on the same time resolutions. For example, in canopy scale predictions, irradiance values are available at time scales of seconds, while eddy covariance data is usually presented over 30 minute time steps. Our modeling indicates that clipping saturating values before averaging improved prediction error by $\approx$ 14-fold, see Figure 4. While these findings demonstrate the value of clipping saturating values before averaging, considering the saturating kinetics of carbon fixation to PPFD, this method could be valuable for any high-resolution data which needs to be averaged over a time-step in a process with saturating kinetics.

Methods

We performed our analyses in Python 3.10, utilising the common packages NumPy, pandas, SciPy and Matplotlib for general data analysis as well as modelbase and Assimulo for building and integrating the ODE model [20, 21, 22, 23, 24, 25, 26]. All code used to generate the publication results and figures is publicly available on our GitLab repository https://gitlab.com/qtb-hhu/photosynthesis-task-force/2022-how-realistic-is-the-cbb-ssa. We obtained field observation PPFD data measured in one minute intervals in Washington state (latitude 45.790835, longitude -121.933788) by the National Ecological Observatory Network (NEON) for which we identified which day had the highest data coverage for each month, which was the 25th, and the ODE model from Saadat 2021 [18, 17]. Due to model instabilities for PPFD values below 30, we clamped the minimum of the data to 30.

Steady-state model

We calculated the steady state fluxes of the ODE model for PPFD values between the minimal and maximal values found in the dataset with step size 1. Then we simulated the model by looking up the steady state flux of the PPFD value rounded to the nearest integer.

Approximations

We fitted the $V_{\max}$ and K_m parameters of the Michaelis-Menten function using the SciPy `minimize` function and the L-BFGS-B algorithm [23, 27]. The polynomial fit was performed using NumPy's `polyfit` function [21].

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy EXC 2048/1, Project ID: 390686111 (O.E.), EU's Horizon 2020 research and innovation programme under the Grant Agreement 862087 (M.v.A.) and also supported by the U.S. Department of Energy Office of Science, Basic Energy Sciences under Award DE-FG02-91ER20021 (B.J.W.).

References

References

[1] Robert W Pearcy. "Sunflecks and photosynthesis in plant canopies". In: *Annual review of plant biology* 41.1 (1990), pp. 421–453.

[2] Sarah M Assmann and Xi-Qing Wang. "From milliseconds to millions of years: guard cells and environmental responses". In: *Current opinion in plant biology* 4.5 (2001), pp. 421–428.

[3] Rebecca A Slattery et al. "The impacts of fluctuating light on crop performance". In: *Plant physiology* 176.2 (2018), pp. 990–1003.

[4] Alejandro Morales and Elias Kaiser. "Photosynthetic acclimation to fluctuating irradiance in plants". In: *Frontiers in Plant Science* 11 (2020), p. 268.

[5] Xinyu Fu and Berkley J Walker. "Dynamic response of photorespiration in fluctuating light environments". In: *Journal of Experimental Botany* (2022).

[6] Anne Arnold and Zoran Nikoloski. "A quantitative comparison of Calvin–Benson cycle models". In: *Trends in Plant Science* 16.12 (2011), pp. 676–683. ISSN: 1360-1385. DOI: https://doi.org/10.1016/j.tplants.2011.09.004. URL: https://www.sciencedirect.com/science/article/pii/S1360138511001956.

[7] Anna Matuszyńska, Nima P Saadat, and Oliver Ebenhöh. "Balancing energy supply during photosynthesis–a theoretical perspective". In: *Physiologia plantarum* 166.1 (2019), pp. 392–402.

[8] Ladislav Nedbal, Jan Červený, and Henning Schmidt. "Scaling and integration of kinetic models of photosynthesis: Towards comprehensive e-photosynthesis". In: *Photosynthesis in silico*. Springer, 2009, pp. 17–29.

[9] 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". In: *Plant physiology* 145.2 (2007), pp. 513–526.

[10] A Laisk et al. "A mathematical model of the carbon metabolism in photosynthesis. Difficulties in explaining oscillations by fructose 2, 6-bisphosphate regulation". In: *Proceedings of the Royal Society of London. B. Biological Sciences* 237.1289 (1989), pp. 389–415.

[11] Samuel H Taylor and Stephen P Long. "Slow induction of photosynthesis on shade to sun transitions in wheat may cost at least 21% of productivity". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1730 (2017), p. 20160543.

[12] B.D. Hahn. "Photosynthesis and photorespiration: Modelling the essentials". In: *Journal of Theoretical Biology* 151.1 (1991), pp. 123–139. ISSN: 0022-5193. DOI: https://doi.org/10.1016/S0022-5193(05)80147-X. URL: https://www.sciencedirect.com/science/article/pii/S002251930580147X.

[13] Jaime Milstein and Hans J Bremermann. "Parameter identification of the Calvin photosynthesis cycle". In: *Journal of Mathematical Biology* 7.2 (1979), pp. 99–116.

[14] B Hahn. "A mathematical model of leaf Calvin cycle". In: *Annals of Botany* 54 (1984), pp. 325–339.

[15] Brian D Hahn. "A mathematical model of photorespiration and photosynthesis". In: *Annals of Botany* 60.2 (1987), pp. 157–169.

[16] A Laisk and David Alan Walker. "Control of phosphate turnover as a rate-limiting factor and possible cause of oscillations in photosynthesis: a mathematical model". In: *Proceedings of the Royal society of London. Series B. Biological sciences* 227.1248 (1986), pp. 281–302.

[17] Nima P. Saadat et al. "Computational Analysis of Alternative Photosynthetic Electron Flows Linked With Oxidative Stress". In: *Frontiers in Plant Science* 12 (2021). ISSN: 1664-462X. DOI: 10.3389/fpls.2021.750580. URL: https://www.frontiersin.org/article/10.3389/fpls.2021.750580.

[18] National Ecological Observatory Network (NEON). *Photosynthetically active radiation (PAR) (DP1.00024.001)*. en. 2022. DOI: 10.48443/8A01-0677. URL: https://data.neonscience.org/data-products/DP1.00024.001/RELEASE-2022.

[19] Xinyu Fu et al. "Integrated flux and pool size analysis in plant central metabolism reveals unique roles of glycine and serine during photorespiration". In: (2022).

[20] Python Core Team. *Python: A dynamic, open source programming language*. Python version 3.10. Python Software Foundation. 2022. URL: https://docs.python.org/3/whatsnew/3.10.html.

[21] Charles R Harris et al. "Array programming with NumPy". In: *Nature* 585.7825 (2020), pp. 357–362. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2649-2. URL: https://doi.org/10.1038/s41586-020-2649-2.

[22] Wes McKinney. "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 51–56.

[23] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* (2020). DOI: https://doi.org/10.1038/s41592-019-0686-2.

[24] John D Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in science & engineering* 9.3 (2007), p. 90.

[25] Marvin van Aalst, Oliver Ebenhöh, and Anna Matuszyńska. "Constructing and analysing dynamic models with modelbase v1. 2.3: a software update". In: *BMC bioinformatics* 22.1 (2021), pp. 1–15.

[26] Christian Andersson, Claus Führer, and Johan Åkesson. "Assimulo: A unified framework for ODE solvers". In: *Mathematics and Computers in Simulation* 116 (2015), pp. 26–43. ISSN: 0378-4754. DOI: https://doi.org/10.1016/j.matcom.2015.04.007. URL: https://www.sciencedirect.com/science/article/pii/S0378475415000701.

[27] Ciyou Zhu et al. "Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization". In: *ACM Trans. Math. Softw.* 23.4 (Dec. 1997), pp. 550–560. ISSN: 0098-3500. DOI: 10.1145/279232.279236. URL: https://doi.org/10.1145/279232.279236.

3.5 What controls carbon sequestration in plants under which conditions?

The paper "What controls carbon sequestration in plants under which conditions" investigates which reactions involved in carbon fixation (e.g. CBB cycle and PETC) carry control over the carbon fixation rate under different environmental conditions, challenging overly simplifying results that claim the control is always on the same reaction, e.g. SBPase. I was involved in this project during every step of the research process, including writing the manuscript.

Publication	What controls carbon sequestration in plants under which conditions?
Author position	Second
Status	peer-reviewed and published
doi	10.1016/j.biosystems.2023.104968

BioSystems 231 (2023) 104968

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

What controls carbon sequestration in plants under which conditions?

Tim Nies [a,*], Marvin van Aalst [a], Nima Saadat [a,b], Josha Ebeling [a], Oliver Ebenhöh [a,b]

[a] Institute of Theoretical and Quantitative Biology, Heinrich Heine University, Düsseldorf, 40225, North Rhine-Westphalia, Germany
[b] Cluster of Excellence on Plant Sciences, Heinrich Heine University, Düsseldorf, 40225, North Rhine-Westphalia, Germany

ARTICLE INFO

Dataset link: https://gitlab.com/qtb-hhu/photosynthesis-task-force/2023-50-years-of-mca

Keywords:
Metabolic control theory
Carbon sequestration
Electron transport
Mathematical model

ABSTRACT

Photosynthetic organisms use photosynthesis to harvest sunlight and convert the solar energy into chemical energy, which is then used to reduce atmospheric carbon dioxide into organic molecules. This process forms the basis of all life on Earth, and stands at the beginning of the food chain which feeds the world population. Not surprisingly, many research efforts are currently ongoing aiming at improving growth and product yield of photosynthetic organisms, and several of these activities directly target the photosynthetic pathways. Metabolic Control Analysis (MCA) shows that, in general, the control over a metabolic flux, such as carbon fixation, is distributed among several steps and highly dependent on the external conditions. Therefore, the concept of a single 'rate-limiting' step is hardly ever applicable, and as a consequence, any strategy relying on improving a single molecular process in a complex metabolic system is bound to fail to yield the expected results. In photosynthesis, reports on which processes exert the highest control over carbon fixation are contradictory. This refers to both the photosynthetic 'light' reactions harvesting photons and the 'dark' reactions of the Calvin–Benson–Bassham Cycle (CBB cycle). Here, we employ a recently developed mathematical model, which describes photosynthesis as an interacting supply–demand system, to systematically study how external conditions affect the control over carbon fixation fluxes.

1. Introduction

Photosynthesis classically has been divided into two parts. The 'light' reactions supply energy and reduction equivalents to the 'dark' reactions of the Calvin–Benson–Bassham (CBB) cycle (Bassham et al., 1950, 1954), where carbon dioxide is fixed to form reduced carbon compounds used as building blocks in other metabolic processes. The CBB cycle (demand side) is one of the most critical pathways on Earth that plants and many other photosynthetic organisms use. Current estimates indicate that over 99% of global carbon dioxide is fixed by the key enzyme of the CBB cycle ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) (Raven, 2009). However, the CBB cycle is not the only carbon fixation mechanism employed by photosynthetic organisms, especially phototrophic prokaryotes. For instance, green sulfur bacteria fix carbon dioxide via a reversed tricarboxylic acid cycle or filamentous anoxygenic photorophs use a carbon fixation pathway known as hydroxypropionate pathway for autotrophic growth (Fuchs, 2011). In this paper, we focus on carbon fixation by the CBB cycle only.

To guarantee efficiency and prevent the formation of toxic reactive oxygen species, the supply (PETC) and demand (CBB cycle) of energy and redox equivalents must be coordinated (Matuszyńska et al., 2019). However, the habitats of photosynthetic organisms are usually characterized by a high fluctuation of abiotic factors, such as light intensity and CO_2 concentration (Kaiser et al., 2018), which makes balancing the photosynthetic electron transport chain (PETC, supply side) and the CBB cycle challenging. Therefore, versatile regulatory mechanisms that coordinate carbon fixation and the PETC and adapt both processes to external conditions have evolved. Examples of regulatory mechanisms include non-photochemical quenching, the thioredoxin-dependent redox control of CBB cycle enzymes, and regulated changes in stomatal conductance (Farquhar and Sharkey, 1982; Muller et al., 2001; Geigenberger et al., 2017). These processes are currently targets of research activities aiming to increase plant performance and crop yield (Kaiser et al., 2019).

Considering the importance of the PETC and CBB cycle it is unsurprising that much effort has been spent studying their kinetics regulation, and control by experimental and theoretical methods. Various mathematical models have been developed aiming at providing a theoretical framework to analyse which factors determine the efficiency of carbon fixation (Hahn, 1986, 1987; Pettersson and Ryde-Pettersson, 1988; Poolman et al., 2000; Jablonsky et al., 2011). Kinetic models of the CBB cycle established, e.g., the importance of the sedoheptulose-1,7-bisphosphatase (SBPase) for controlling carbon assimilation and provided theoretical explanations for a wide range of observed kinetic properties of RuBisCO (Poolman et al., 2000; Raines et al., 2000;

* Corresponding author.
E-mail address: tim.nies@hhu.de (T. Nies).

https://doi.org/10.1016/j.biosystems.2023.104968
Received 25 April 2023; Received in revised form 28 June 2023; Accepted 29 June 2023
Available online 5 July 2023

Witzel et al., 2010). Appropriate theoretical tools are needed to study control in metabolic networks, e.g., the CBB cycle and PETC. Metabolic control analysis (MCA) is a theoretical framework developed in the 1970s, which is continuously improved and generalized (Heinrich and Rapoport, 1974; Kacser et al., 1995; Heinrich and Schuster, 1996; Dourado and Lercher, 2020; Wilken et al., 2022). A major purpose of MCA is to quantify the influence that single enzymes have over the steady-state properties of metabolic networks. A central concept is the *control coefficient* which describes how small changes in activities of single steps affect stationary metabolites and fluxes. Because control coefficients depend on the dynamics of the interactions of all components, they are systemic properties. MCA has been repeatedly applied to study the control of reaction steps in plant metabolic pathways (Rohwer, 2012) with examples including applications to the benzoid pathway, sucrose accumulation, the CBB cycle, the electron transport chain, and combinations of these (Uys et al., 2007; Colón et al., 2010; Ebenhöh et al., 2011; Matuszyńska et al., 2019; Saadat et al., 2021).

Here we present an *in silico* analysis of how external conditions affect the control over carbon fixation fluxes. We focused in particular on the effect of two environmental parameters, light intensity, and CO_2 concentration, to assess how these factors affect the control of carbon fixation. Classically many studies stress the importance of RuBisCO and its activation processes as highly influential on photosynthetic efficiency (Stitt and Schulze, 1994). But is RuBisCO always the main controlling factor? For our analysis, we employ a published kinetic model of photosynthesis (see Fig. S1) that combines the PETC, the CBB cycle, and the Ascorbate-Gluthatione (ASC-GSH) cycle implemented in Python using the modelbase software package (Saadat et al., 2021; van Aalst et al., 2021). This model was originally used to study the importance of cyclic electron transport around photosystem I for photoprotection and also includes regulatory mechanisms, such as non-photochemical quenching, state transitions between the two photosystems, and redox regulation of CBB cycle enzymes through the thioredoxin system. We began our analysis by using the stoichiometric structure of this combined model to conduct a reaction correlation analysis (Poolman et al., 2007), followed by the investigation of control on carbon fixation by different key processes in photosynthesis. We identified a condition-dependent shift in control and determined its structural origin using a robustness analysis with sampled parameter sets. Using the results from the robustness analysis, we could show that some reactions exert control in an either-or relationship while others may exert control simultaneously. With this work, we contribute to elucidating the control in photosynthesis and its dependence on external conditions.

2. Results

2.1. Reaction correlation analysis

We begin our analysis by studying the constraints on stationary fluxes, which are imposed by the stoichiometry of the network alone. For this, we calculate reaction correlation coefficients (Poolman et al., 2007) for the previously published photosynthesis supply–demand model (Saadat et al., 2021). These coefficients provide a generalization of the concept of enzyme subsets. Reactions within one enzyme subset are strictly coupled in the sense that they always carry fluxes in a fixed proportion. For such reactions, the reaction correlation coefficient is ± 1, and the corresponding row vectors of the kernel matrix of the stoichiometry matrix are parallel. Reaction correlation coefficients generalize this idea by essentially calculating the angle between row vectors of the kernel matrix (for details, see Poolman et al. (2007)), and therefore indicate how strong reactions are correlated as a result of structural constraints of a network. Fig. 1 presents a metabolic tree constructed by hierarchical clustering using dissimilarities calculated by the reaction correlation coefficients (see Methods). The matrix of reaction correlation coefficients ϕ can be found in the supplement (Fig. S2).

There exist ten enzyme subsets containing more than one reaction (table S1). Generally, reactions that function as regulatory control mechanisms (such as deprotonation/protonation of PsbS, and the xanthophyll cycle necessary for non-photochemical quenching, or the reactions involved in the redox control by thioredoxin), each form an enzyme subset. Three subsets are connected to the electron transport chain and ROS scavenging. The first subset consists of the reactions mediated by cytochrome b_6f and photosystem I, which are therefore strictly coupled. The second contains the Mehler reaction and ascorbate peroxidase, while the third consists of the remaining reactions of the ASC-GSH cycle. Four subsets can be assigned to the CBB cycle and starch synthesis. The largest subset contains eight reactions, including RuBisCO, the transketolase reactions, and the reactions of the regeneration phase leading to the formation of ribulose-1,5-bisphosphate.

In Fig. 1, clusters of reactions, which are highly correlated but not strictly coupled, are indicated by different colours. We identify clusters that are associated with key metabolic functions, such as the PETC, the ASC-GSH cycle, and the CBB cycle. The CBB cycle is split into two pronounced subclusters, which can be associated with carbon fixation and carbon export, respectively. The same is true for the PETC, for which we can distinguish between the linear (photosystems II/I and cytochrome b_6f) and the cyclic electron flow. Interestingly, RuBisCO and PSII are grouped in different clusters, which suggests that these two processes are decoupled. This is unexpected because the electrons obtained by photosystem II are mainly used by the CBB cycle to fix carbon, therefore, one would expect highly correlated fluxes. This observation can be explained by considering that the stoichiometric analysis applied here is based on the full null space and completely ignores constraints on the stationary flux solutions by the kinetic parameters.

A kinetic model, however, drastically restricts the possible stationary fluxes by the specific parameter values of the dynamic equations. Therefore, actually observable steady-state fluxes will represent only a small subset of the complete null space. The processes decoupling reaction rates of RuBisCO and PSII, such as proton leak, terminal oxidases, etc., are constrained by their kinetic parameters to carry only relatively small fluxes, such that, in fact, the correlation between RuBisCO and PSII rates should be high for realistic conditions. To test this assumption, we repeated the correlation analysis based on steady-state fluxes sampled from the kinetic model, in which the reference values of the rate constants have been randomly varied by a factor between 0.5 and 2. The rationale behind this is that now the stationary fluxes are restricted to solutions which are close to a reference state and therefore reflect a more physiologically relevant subset of the null space. The resulting tree is depicted in Fig. 1B. As expected, the fluxes of the CBB cycle and the PETC are now strongly correlated. We decided to use a factor of 2 as we assume it to be a realistic range achievable for biotechnological manipulation and short-term evolutionary processes. To test whether a higher factor would drastically change our results we repeated all analyses with a factor 5 (see Figs. S3 to S9). Generally the analyses with a higher factor result in similar trends, but show larger variations. This is because, at least for some parameters, a factor of 5 is relatively high, leading to entirely different system behaviours than usually expected.

2.2. The control on carbon fixation switches between environmental conditions

We use metabolic control analysis (MCA) to quantify the control that individual molecular processes exert on the performance of the system, measured by the net carbon fixation rate, in different environmental conditions. We found that overall flux control is exerted mostly by one of four steps: photosystem I and II in the electron transport chain and RuBisCO and SBPase in the CBB cycle. In Fig. 2 we depict the four

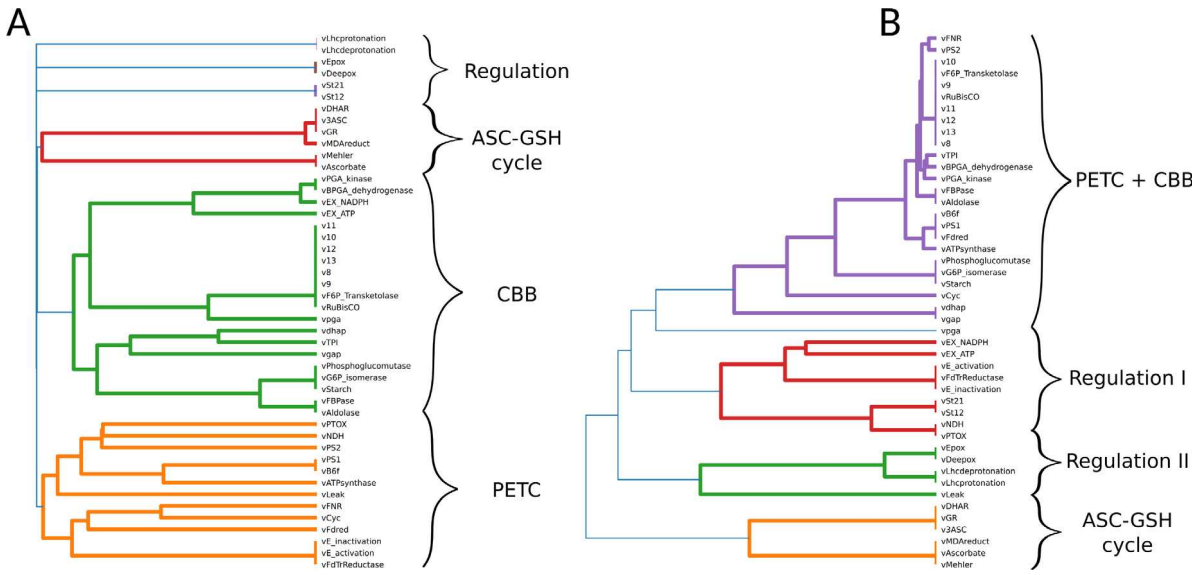


Fig. 1. Metabolic trees, showing different metabolic clusters. The left tree was obtained by reaction correlation analysis. The right tree was obtained by steady-state flux correlation analysis for 10000 random parameter sets in low light intensity and low CO_2 concentration conditions for the kinetic model of photosynthesis. Abbreviations in the cluster annotations: PETC — photosynthetic electron transport chain, CBB — Calvin–Benson–Bassham cycle, ASC-GSH cycle — ascorbate-glutathione cycle.

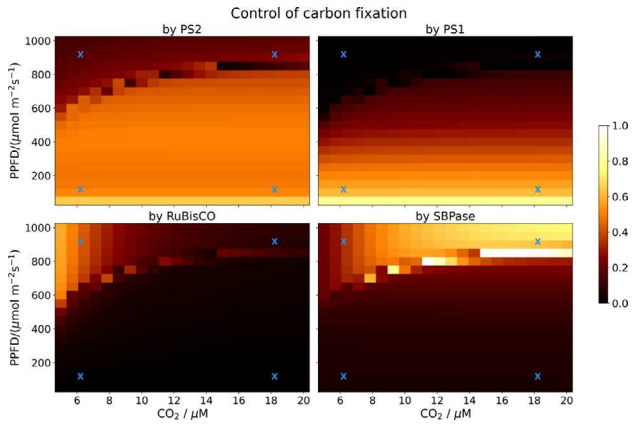


Fig. 2. The flux control of PSII (top left), PSI (top right), RuBisCO (bottom left), and SBPase (bottom right) on carbon fixation in light intensities ranging from 50 to 1000 $\mu mol/m^2/s$ and in CO_2 concentrations ranging from 5 to 20 μM. The control coefficients are indicated by the heat map with dark areas indicating low and light areas high control. The blue crosses are the reference points for the further analyses indicating low CO_2/low light (6 μM, 100 $\mu mol/m^2/s$), low CO_2/high light (6 μM, 900 $\mu mol/m^2/s$), high CO_2/low light (18 μM, 100 $\mu mol/m^2/s$), and high CO_2/high light (18 μM, 900 $\mu mol/m^2/s$).

flux control coefficients on the overall carbon fixation rate for light intensities ranging from 50 to 1000 $\mu mol/m^2/s$and for CO_2 concentrations between 6 and 20 μM, corresponding to atmospheric concentrations between approximately 170 and 700 ppm. It is clearly visible that there are two distinct light regimes. A sharp transition between control by the light reactions (low light) to control by the dark reactions (high light) can be observed. Interestingly, the curve separating these two regimes corresponds to the limit where the quenching capacity reaches its maximum, and the lumen becomes highly acidic (see Fig. S10). As we observed and discussed previously (Saadat et al., 2021), this transition marks the saturation of the photosynthetic system, above which increasing light no longer facilitates higher carbon fixation rates. At the transition, the carbon fixation rate (and many other rates and intermediate concentrations) is not a smooth function of the incident light intensity. Therefore, the numerical differentiation employed to calculate the control coefficients may lead to imprecise results and as a consequence, the coefficients very close to the transition should be interpreted with care.

In the low light regime, both photosystems have substantial control whereas the CBB cycle enzymes exert almost no control. This changes drastically for higher light intensities, where PSI exerts practically no control and PSII only a small but distinguishable control, whereas the CBB cycle enzymes now control the carbon fixation rate. In high light, a gradual shift in control from RuBisCO to SBPase can be observed as CO_2 concentrations increase. SBPase has the highest control in high light intensities and high CO_2 concentrations, while RuBisCO is the dominant reaction in high light intensities and low CO_2 concentrations. High control of RuBisCO in high light and low CO_2 concentrations has also been found experimentally (Stitt and Schulze, 1994).

In summary, this analysis shows that the control on carbon fixation switches from photosystem I in low light to photosystem II in medium light intensities to SBPase and RuBisCO in high light intensities, where RuBisCO control dominates in low and SBPase control in high CO_2 concentrations. For our further analyses, we define four reference conditions for low CO_2/low light (6 μM, 100 $\mu mol/m^2/s$), low CO_2/high light (6 μM, 900 $\mu mol/m^2/s$), high CO_2/low light (18 μM, 100 $\mu mol/m^2/s$), and high CO_2/high light (18 μM, 900 $\mu mol/m^2/s$). These conditions are indicated by blue crosses in Fig. 2.

2.3. Control of photosynthetic intermediates

Besides the carbon fixation rate, also the states of the intermediates in the photosynthetic electron transport chain and the CBB cycle are important determinants for the efficiency and status of the photosynthetic system at large. In particular, poised redox levels of the electron carriers are indicative of the efficient functioning of the PETC, the concentrations of ATP and NADPH are important as ubiquitous energy and redox equivalents, and the CBB cycle intermediates must be above a certain level to ensure the cycle runs efficiently (Matuszyńska et al., 2019). Moreover, various mechanisms ensure that in particular in high light, photodamage by reactive oxygen species (ROS) is minimized.

The electron carriers behave as expected (see Figs. S11 to S14). In general, upstream reactions have a positive control on their redox state, while downstream reactions exert a negative control. For example, the redox state of plastoquinone is strongly positively controlled by PSII, slightly positive by the cyclic electron flow (which feeds back electrons from ferredoxin to plastoquinone), and negatively or not at

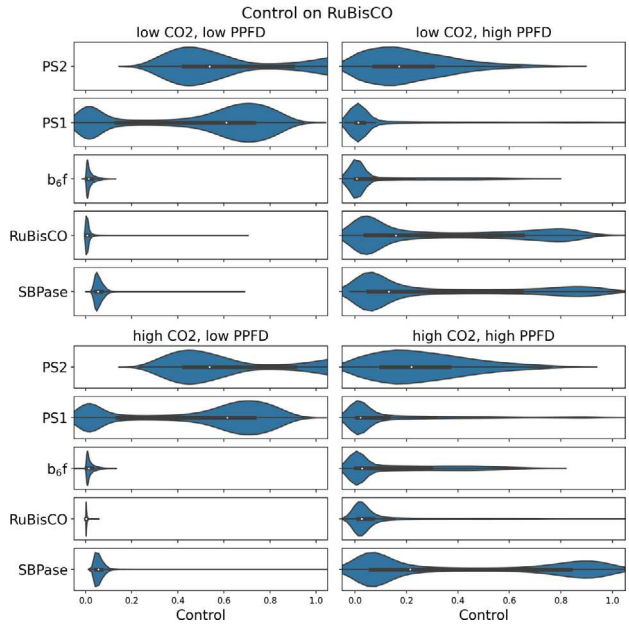


Fig. 3. Distribution of flux control coefficients of key reactions in the PETC and CBB cycle on carbon fixation over 10000 sets of randomly perturbed parameters with a factor between 0.5 and 2. The shaded area shows the frequency of flux control coefficients. In all cases, the area was scaled to the maximum (including values outside the shown range) for clarity. The white dot indicates the median of the distribution and the bold part of the central line denotes the range between upper and lower quartile. Top left shows results under low CO_2/low light conditions, top right under low CO_2/high light, bottom left under high CO_2/low light and bottom right under CO_2/high light conditions. Only flux control coefficients between 0 and 1 are shown.

all by downstream processes, such as PSI or the CBB enzymes (see figure S11). The only electron carrier that is more reduced when the CBB cycle enzymes are increased is plastocyanin, which is in agreement with previous model analyses (Saadat et al., 2021) (Fig. S12). An interesting observation is that under low light, both ferredoxin and NADPH are less reduced if PSI activity is enhanced, although ferredoxin is a direct product of PSI (Figs. S13 and S14). A possible explanation for this counter-intuitive finding is that the cyclic electron flow is strongly increased with increasing PSI activity (Fig. S15) and that, together with the increased CBB activity (see above) this leads to a slight reduction of these two electron carriers. The control of ATP levels is complex (Fig. S16). For example, increased PSII leads to reduced ATP levels in very low light, increased in intermediate light (still below the quencher saturation threshold), and a slight reduction again for high light conditions. However, steady-state energy levels range between 0.6 and 0.8 (fraction of ATP in the adenosine phosphate pool), which are in the range of measured values (Stitt et al., 1982). An interesting effect is observed when calculating the control on the total phosphates in CBB intermediates. Apparently, enhancing the fixation process (RuBisCO) leads to a reduction in CBB intermediates, whereas enhancing the recycling phase (SBPase) leads to an increase, except for very low light intensities (Fig. S17). ROS (simulated as stationary H_2O_2 concentrations) levels respond as expected. Increasing the photosystems leads to higher levels while increasing the cyclic electron flow the b_6f complex activity or the CBB cycle lead to reduced levels (Fig. S18.)

2.4. Robustness of the control on carbon fixation in multiple environmental conditions

Control coefficients quantify the strength of control of individual processes in a metabolic network. They are system-wide properties and as such depend on the specific values of the kinetic parameters of the involved enzymatic reactions. Therefore, they should not be considered a rigid value independent of all choices in the model-building process or of varying external conditions. The control on RuBisCO, an essential enzyme for carbon fixation, by other reaction steps in the CBB cycle, PETC, or ASC-GSH cycle, is interesting for broadening our understanding of sequestering carbon in photosynthetic organisms.

In order to determine if the previously observed shift in control (Fig. 2) is a consequence of the structural design of the PETC and the CBB, we performed a robustness analysis. For this, we varied parameters by multiplying a randomly selected factor between 0.5 and 2 to generate 10000 perturbed parameter sets. For each parameter set, we analysed the control exerted by PSII, cytochrome b_6f, RuBisCO, FBPase, and SBPase on carbon fixation in the four reference conditions for low/high CO_2/light as defined above. Fig. 3 shows the distributions of flux control coefficients on carbon fixation by selected reactions. With most parameter sets, the photosystems had a much higher control on carbon fixation in low light intensities in both CO_2 concentration conditions than reactions in the CBB cycle. In low light intensity, cytochrome b_6f and RuBisCO have almost no control, and SBPase, while detectable, has only a minor influence. Investigating the correlation of control coefficients for both photosystems under low-light conditions reveals that these two processes indeed share the main flux control in a proportion, which depends on the exact parameter values (Fig. 4).

The control of photosystems on carbon fixation is drastically reduced in high-light conditions. Especially photosystem I has lost almost all its control. As a general trend, the distribution of flux control coefficients in high light is broader than in low light. The main controlling steps are now on the demand side of photosynthesis, in particular on RuBisCO and SBPase. SBPase is, besides photosystem II, the controlling reaction for carbon fixation in high light intensity and high CO_2 conditions. At the same time, RuBisCO and SBPase are the main factors in high light intensity and low CO_2 concentrations, with RuBisCO having a slightly higher influence. Correlating the control coefficients of these two central CBB enzymes shows for most randomly selected parameter sets RuBisCO does not exert any considerable control under high light and high CO_2 concentrations, while under high light but low CO_2 the control can be on either of these enzymes (or none of the two) but only for a few parameter sets the control is shared between the two enzymes (Fig. 5).

Overall, our robustness analysis, in which we randomly varied parameters by a factor between 0.5 and 2, confirms our previous observations, namely that the control shifts from the photosystems in low light to the CBB cycle enzymes in high light. Under the latter conditions, RuBisCO exerts a higher control if ambient CO_2 concentrations are low, while SBPase is the controlling step under high CO_2 concentrations. This indicates that the shift in control is less a kinetic, parameter-dependent, effect but rather a structural property of photosynthesis.

3. Discussion & conclusions

Photosynthesis is a supply–demand system. The supply (PETC) and demand (CBB cycle) sides must be coordinated to ensure efficient photosynthesis. Considering the often rapidly and unpredictably changing light intensities (Kaiser et al., 2018) plants are exposed to in natural environments, maintaining such coordination appears challenging. It is plausible to assume that the present environmental-dependent regulatory mechanisms controlling carbon fixation have evolved to be highly efficient, considering the direct effect that carbon capture has on plant growth and fitness. This work presents an *in silico* analysis of the control over carbon fixation in different environmental conditions. For this, we used a published model of photosynthesis that combines the supply side (PETC), the demand side (CBB cycle), and the Ascorbate-Glutathion cycle (Saadat et al., 2021).

Such a supply–demand photosynthesis model allows for quantifying the control that individual processes have on the overall carbon fixation

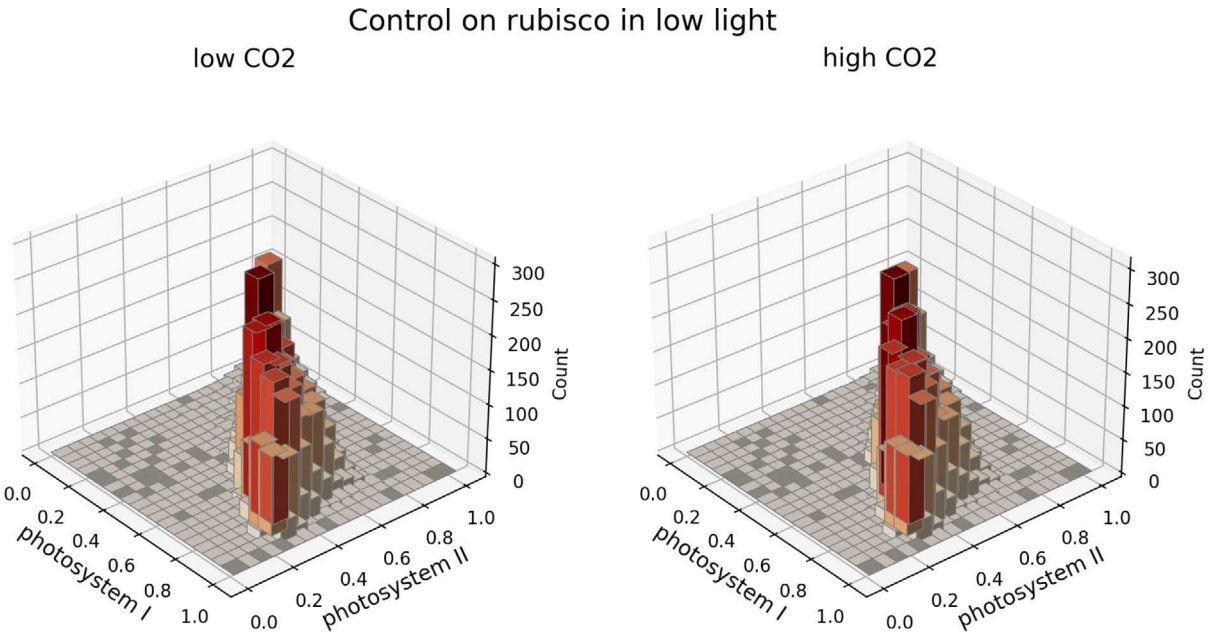


Fig. 4. Correlation of control coefficients for both photosystems on carbon fixation under low-light conditions represented as a 3D histogram. The *z*-axis indicates how many control coefficients fall into a specific numerical range. The calculation is based on 10000 randomly generated parameter sets as described in the text.

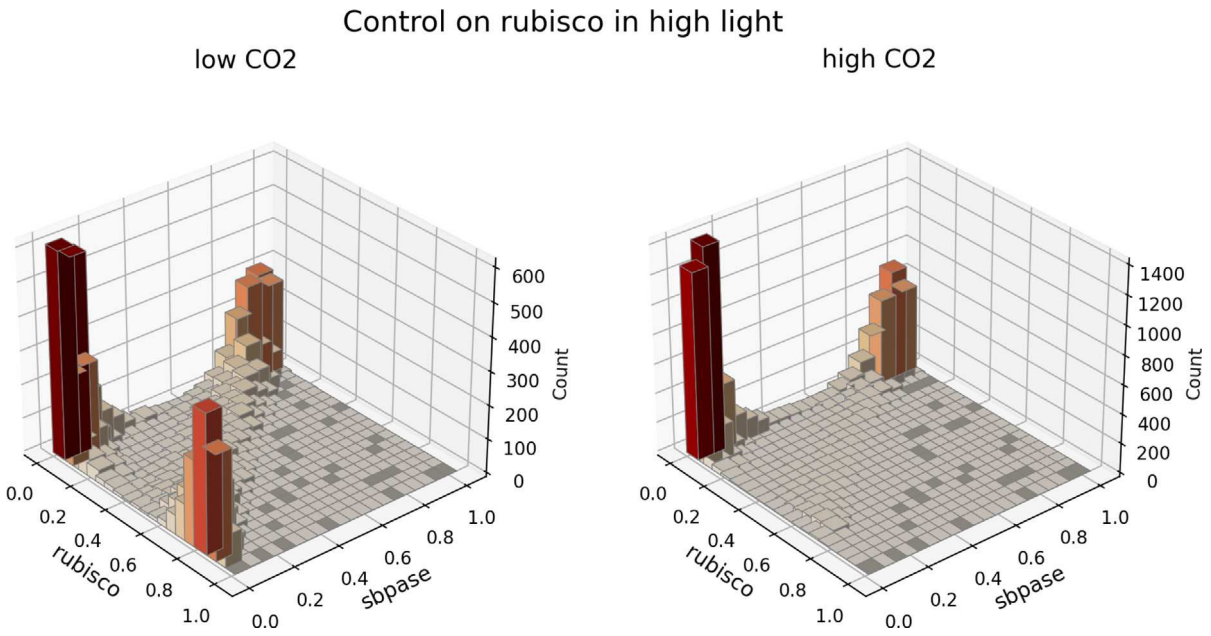


Fig. 5. Control coefficients of carbon fixation by RuBisCO vs. SBPase under high-light conditions represented as a 3D histogram. The *z*-axis indicates how many control coefficients fall into a specific numerical range. The calculation is based on 10000 randomly generated parameter sets as described in the text.

rate. We focused in particular on photosystems I and II, cytochrome b_6f, RuBisCO, and SBPase, which were reported to exert control on carbon fixation under several conditions (Poolman et al., 2000; Johnson and Berry, 2021; Raines et al., 2000). Using Metabolic Control Analysis, we quantified the control of these single steps on carbon fixation for different simulated environmental conditions. By simultaneously varying the light intensity and CO_2 concentration, we could show that the control shifts from the photosystems in low light intensities to RuBisCO and SBPase in high light intensities but then from RuBisCO in low to SBPase in high CO_2 concentrations (Fig. 2). The shift of the control confirms that most of the reactions previously reported to control the flux are indeed critical for regulating carbon fixation. However, whether PSII, PSI, RuBisCO, or SBPase is the main controlling factor strongly depends on the external conditions. In our photosynthesis model, a relatively sharp threshold marks the transition between a supply- and a demand-controlled situation (see Fig. 2). This threshold, separating 'low' and 'high' light conditions, occurs when the quenching mechanism reaches its maximal capacity (see Fig. S10). This results in a reduction of most electron carriers and a sharp accumulation of protons in the lumen. The PETC still operates at a fast rate, so ATP and NADPH production is no longer limiting carbon fixation. It is an open question whether this sharp transition is a feature of the specific model that was used for this analysis or whether this is actually a systemic property of photosynthesis. We assume that the transition from non-saturated to saturated quencher is not as sharp *in vivo* as suggested by the model, but that the principle feature, namely that high light intensity results in a

shift of control to the demand reactions of the CBB cycle, is a structural feature of the photosynthetic supply–demand system. The continuous transition under high light between the RuBisCO and SBPase-mediated control suggests that at high CO_2 concentrations, the carboxylation by RuBisCO is not determining carbon fixation rate, but rather the distribution of the intermediates in the CBB cycle through SBPase. To test whether the shift in control is a kinetic property of the rates in photosynthesis or follows from the structure resulting from the interconnections of the PETC and CBB cycle, we performed a robustness analysis by randomly varying kinetic parameters by a factor between 0.5 and 2. Fig. 3 illustrates that, at least in our model representation, the control shift is a property that occurs with many parameter sets. This observation indicates that the shift of control indeed seems to be a structural feature, and rather independent of the specific parameter values.

Interesting patterns emerge by correlating the flux control coefficients obtained by the robustness analysis. In the low light regime, the control is shared mostly among the two photosystems (Fig. 4 and Fig. S19), where often one of the photosystems exhibits a higher control than the other. Which photosystem exerts the higher control apparently depends on the specific numerical parameter values. Additionally, the fact that both photosystems always have clearly non-zero control for all parameter sets in low light underlines the importance of the PETC for carbon fixation as a limiting factor in these conditions. The light-driven photosystems ultimately determine the flow of electrons through the PETC and the translocation of protons into the lumen, hence the production of ATP and reduction equivalents required by the CBB cycle. Correlating the control coefficients quantifying the importance of RuBisCO and SBPase under high light shows a drastically different picture. Fig. 5 reveals that typically carbon fixation is either controlled by RuBisCO or by SBPase, but the control is rarely shared. This is especially pronounced in high light intensity and low CO_2 concentration conditions. Fig. 5 also reveals that for a substantial number of parameter combinations, neither RuBisCO nor SBPase exerts control over the carbon fixation rate. A closer inspection reveals that in these cases, the control lies, in fact, with the photosystems. In fact, correlating the total control (sum of control coefficients of the individual processes) of the supply reactions with the total control of the demand reactions reveals that the control lies either on the supply side or on the demand side, but is rarely shared between both sides (Fig. S20). An interesting observation is that even in high light conditions, the model exhibited control by the light reactions for a substantial fraction of parameter sets. A possible explanation for the observation that also in high light for many parameter sets the control lies on the photosystems is that variations of the parameters can lead to scenarios where our selected 'high light' condition is actually not perceived as saturating light. In order to test this hypothesis, we relate the control exhibited by the dark reaction to the simulated stationary lumen pH (Fig. S21). This analysis shows that whenever the control is on the dark reactions, the pH is low, indicating that light (and the quencher) is saturated, whereas low control by the dark reactions is associated with a high lumenal pH, indicating non-saturating conditions.

Some experimental and theoretical studies claim that cytochrome b_6f controls the photosynthetic flux (Stiehl and Witt, 1969; Johnson and Berry, 2021). In contrast, our analysis suggests that cytochrome b_6f exhibits a considerable control only for very few parameter sets (Fig. 3). However, when we systematically decrease the activity of cytochrome b_6f, also in our model cytochrome b_6f can become a rate-controlling step (Fig. S22). These considerations show that seemingly conflicting reports on the control of cytochrome b_6f are not necessarily contradictory. In fact, the parameters describing the composition and kinetic properties of the photosynthetic apparatus have an important influence on the strength of control.

Most concentration control coefficients behave as expected. For the electron carriers, upstream reactions exert positive and downstream reactions negative control. We obtained an initially counter-intuitive result only for ferredoxin and NADPH, as they are both less reduced when PSI activity is enhanced in low light. This observation might be explained through an increased cyclic electron flow with a concomitant increase in CBB cycle activity. The cyclic electron flow is an integral part of the photosynthetic machinery adjusting the ATP/NADPH ratio in the PETC and, hence, is an essential regulatory mechanism. Responding to the ATP/NADPH ratio required by the demand reactions, the effects of other processes can be reduced or even reverted, when compared to a system without cyclic electron flow. The regulatory effects of CEF may also explain the complex patterns in the control that some processes have on ATP concentrations. These results demonstrate that control in a complex system is often non-trivial, and altering reaction rates may result in counterintuitive effects.

Exploring a previously published supply–demand photosynthesis model with metabolic control analysis, we could resolve seemingly contradictory statements about which reactions have the strongest control on carbon fixation. We showed that basically all reactions previously reported exerting a strong control can indeed have high flux control under some conditions. It is important to note that all results have been obtained from a single, imperfect model. The model does not, for example, include the important process of photorespiration or stomatal aperture. It is unclear how far the interpretation of the results and the derived conclusions can be generalized. Still, the strength of theoretical analyses is that also with simplified and imperfect models, general features can be identified and novel hypotheses derived. For example, the general pattern observed in our analysis of how the control shifts between key enzymes and complexes depending on light intensity and CO_2 concentrations is plausible and generally applicable. By understanding the principles of how regulation depends on environmental conditions, new data can be interpreted in a highly informed manner. Additionally, the existence of different physiological states and the fact that the control in photosynthesis shifts might necessitate the adaption of experimental protocols aiming for the improvement of photosynthesis. For instance, improvement strategies could differ depending on the typical physiological states of the photosynthetic organism. With our study, we aimed at demonstrating the usefulness of systematic model analyses with Metabolic Control Analysis in understanding metabolic regulation in complex networks.

4. Methods

4.1. Model description

For the *in silico* analyses, we used a previously published model of photosynthesis (Saadat et al., 2021). This model (see Fig. S1) combines mechanistic descriptions of the PETC, and the CBB cycle, supplying and consuming ATP and NADPH. The model includes the regulation of CBB enzymes via thioredoxin and mechanisms responsible for producing and scavenging ROS around PSI. The scavenging of ROS is mediated by a module representing the ASC-GSH cycle.

4.2. Metabolic control analysis

The flux ($C^{J_j}_{v_k}$) and concentration ($C^{S_j}_{v_k}$) coefficients are defined, as

$$C^{J_j}_{v_k} = \frac{v_k}{J_j}\frac{\partial J_j/\partial p}{\partial v_k/\partial p}, \tag{1}$$

and

$$C^{S_j}_{v_k} = \frac{v_k}{S_j}\frac{\partial S_j/\partial p}{\partial v_k/\partial p}, \tag{2}$$

where the steady state fluxes and concentrations are denoted as J_j and S_j, respectively. p is a kinetic parameter affecting only the reaction k with rate v_k directly. In the computational analyses, the control coefficients were numerically approximated using central difference and varying the parameter p by ±1%.

4.3. Reaction dendrogram and reaction correlation coefficients

A set of flux vectors satisfying the steady state condition,

$$\mathbf{Nv} = \mathbf{0}, \tag{3}$$

defines the null space of the stoichiometric matrix. A set of base vectors summarized in the kernel matrix **K**, in which they form the columns, span the null space. The kernel matrix can be obtained by the relation,

$$\mathbf{NK} = 0. \tag{4}$$

The reaction correlation coefficients were calculated following (Poolman et al., 2007). The kernel matrix was orthonormalized using the Gram–Schmidt process implemented in the sympy (Meurer et al., 2017) package. For a pair of reactions and corresponding row vectors in the kernel matrix $\mathbf{k}_i$, and $\mathbf{k}_j$, the reaction correlation coefficients calculate as,

$$\phi_{ij} = \frac{\mathbf{k}_i \mathbf{k}_j^T}{\sqrt{\mathbf{k}_i \mathbf{k}_i^T}\sqrt{\mathbf{k}_j \mathbf{k}_j^T}}. \tag{5}$$

.

The dissimilarity matrix Δ_{ij}, describing the angle between the row vectors of the kernel matrix **K**, was obtained using the reaction correlation matrix

$$\Delta_{ij} = cos^{-1}(\phi_{ij}). \tag{6}$$

Hierarchical clustering, using Δ_{ij}, was conducted with the WPGMA algorithm implemented in the scipy package (Virtanen et al., 2020).

4.4. Robustness analysis

To analyse whether our previous results were due to the general properties of the system or due to the choice of parameters we performed a control analysis scan over 10000 sets of randomly perturbed parameters. The varied parameters were, (1) the total concentration of photosystem II and (2) photosystem I, (3) the rate constant of cytochrome b6f, (4) the rate constant determining the rate of cyclic electron flow and (5) the Mehler reaction, (6) the maximum velocities of RuBisCO, (7) FBPase, and (8) SBPase, as well as the rate constants of the (9) MDAR and (10) DHAR-catalyzed reactions. The model parameters were randomly multiplied by a factor 2^x, where x was drawn from a uniform random distribution between −1 and 1. This results in a multiplication by a factor between 0.5 and 2. This analysis was performed for low/high CO2 and light conditions. To reduce the amount of unrealistic simulation we only used those parameter sets that lead to a 5 times higher or lower RuBP steady state concentration for further downstream analyses.

Abbreviations

MCA — Metabolic control analysis, CBB — Calvin–Benson–Bassham-cycle, PETC — photosynthetic electron transport chain, ROS — reactive oxygen species, ASC-GSH — ascorbate-glutathione cycle, SBPase — sedoheptulose-1,7-bisphosphatase, RuBisCO — ribulose-1,5-bisphosphate carboxylase/oxygenase

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG), Germany, project ID 391465903/GRK 2466 (T.N.), the Deutsche Forschungsgemeinschaft (DFG), Germany under Germany's Excellence Strategy EXC 2048/1, Project ID: 390686111 (N.S., O.E.) and EU's Horizon 2020 research and innovation programme under the Grant Agreement 862087 (M.v.A.).

CRediT authorship contribution statement

Tim Nies: Visualization, Formal analysis, Writing. **Marvin van Aalst:** Visualization, Formal analysis, Writing. **Nima Saadat:** Writing and discussions. **Josha Ebeling:** Visualization, Formal analysis. **Oliver Ebenhöh:** Initial idea, Conceptualization, Funding acquisition, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code can be found at https://gitlab.com/qtb-hhu/photosynthesis-task-force/2023-50-years-of-mca.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.biosystems.2023.104968.

References

Bassham, J.A., Benson, A.A., Calvin, M., 1950. The path of carbon in photosynthesis VIII. The role of malic acid. J. Biol. Chem. 185 (2), 781–787. http://dx.doi.org/10.1016/S0021-9258(18)56368-7.

Bassham, J.A., Benson, A.A., Kay, L.D., Harris, A.Z., Wilson, A.T., Calvin, M., 1954. The path of carbon in photosynthesis. XXI. The cyclic regeneration of carbon dioxide acceptor1. J. Am. Chem. Soc. 76 (7), 1760–1770.

Colón, A.M., Sengupta, N., Rhodes, D., Dudareva, N., Morgan, J., 2010. A kinetic model describes metabolic response to perturbations and distribution of flux control in the benzenoid network of Petunia hybrida. Plant J. 62 (1), 64–76. http://dx.doi.org/10.1111/j.1365-313X.2010.04127.x.

Dourado, H., Lercher, M.J., 2020. An analytical theory of balanced cellular growth. Nature Commun. 11 (1), 1226. http://dx.doi.org/10.1038/s41467-020-14751-w.

Ebenhöh, O., Houwaart, T., Lokstein, H., Schlede, S., Tirok, K., 2011. A minimal mathematical model of nonphotochemical quenching of chlorophyll fluorescence. Biosystems 103 (2), 196–204. http://dx.doi.org/10.1016/j.biosystems.2010.10.011.

Farquhar, G.D., Sharkey, T.D., 1982. Stomatal conductance and photosynthesis. Annu. Rev. Plant Physiol. 33 (1), 317–345. http://dx.doi.org/10.1146/annurev.pp.33.060182.001533.

Fuchs, G., 2011. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? Annu. Rev. Microbiol. 65, 631–658. http://dx.doi.org/10.1146/annurev-micro-090110-102801.

Geigenberger, P., Thormählen, I., Daloso, D.M., Fernie, A.R., 2017. The unprecedented versatility of the plant thioredoxin system. Trends Plant Sci. 22 (3), 249–262. http://dx.doi.org/10.1016/j.tplants.2016.12.008.

Hahn, B.D., 1986. A mathematical model of the Calvin cycle: analysis of the steady state. Ann. Botany 57 (5), 639–653. http://dx.doi.org/10.1093/oxfordjournals.aob.a087147.

Hahn, B.D., 1987. A mathematical model of photorespiration and photosynthesis. Ann. Botany 60 (2), 157–169. http://dx.doi.org/10.1093/oxfordjournals.aob.a087432.

Heinrich, R., Rapoport, T.A., 1974. A linear steady-state treatment of enzymatic chains: general properties, control and effector strength. Eur. J. Biochem. 42 (1), 89–95. http://dx.doi.org/10.1111/j.1432-1033.1974.tb03318.x.

Heinrich, R., Schuster, S., 1996. The Regulation of Cellular Systems. Chapman & Hall, London.

Jablonsky, J., Bauwe, H., Wolkenhauer, O., 2011. Modeling the Calvin-Benson cycle. BMC Syst. Biol. 5 (1), 1–13. http://dx.doi.org/10.1186/1752-0509-5-185.

Johnson, J., Berry, J., 2021. The role of cytochrome b6f in the control of steady-state photosynthesis: a conceptual and quantitative model. Photosynth. Res. 148 (3), 101–136. http://dx.doi.org/10.1007/s11120-021-00840-4.

Kacser, H., Burns, J.A., Kacser, H., Fell, D., 1995. The control of flux. Biochem. Soc. Trans. 23 (2), 341–366.

Kaiser, E., Correa Galvis, V., Armbruster, U., 2019. Efficient photosynthesis in dynamic light environments: a chloroplast's perspective. Biochem. J. 476 (19), 2725–2741. http://dx.doi.org/10.1042/BCJ20190134.

Kaiser, E., Morales, A., Harbinson, J., 2018. Fluctuating light takes crop photosynthesis on a rollercoaster ride. Plant Physiol. 176 (2), 977–989. http://dx.doi.org/10.1104/pp.17.01250.

Matuszyńska, A., Saadat, N.P., Ebenhöh, O., 2019. Balancing energy supply during photosynthesis–a theoretical perspective. Physiol. Plant. 166 (1), 392–402. http://dx.doi.org/10.1111/ppl.12962.

Meurer, A., Smith, C.P., Paprocki, M., Čertík, O., Kirpichev, S.B., Rocklin, M., Kumar, A., Ivanov, S., Moore, J.K., Singh, S., Rathnayake, T., Vig, S., Granger, B.E., Muller, R.P., Bonazzi, F., Gupta, H., Vats, S., Johansson, F., Pedregosa, F., Curry, M.J., Terrel, A.R., Roučka, Š., Saboo, A., Fernando, I., Kulal, S., Cimrman, R., Scopatz, A., 2017. SymPy: symbolic computing in Python. PeerJ Comput. Sci. 3, e103. http://dx.doi.org/10.7717/peerj-cs.103.

Muller, P., Li, X.P., Niyogi, K.K., 2001. Non-photochemical quenching. A response to excess light energy. Plant Physiol. 125 (4), 1558–1566. http://dx.doi.org/10.1104/pp.125.4.1558.

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

Poolman, M.G., Fell, D.A., Thomas, S., 2000. Modelling photosynthesis and its control. J. Exp. Bot. 51 (suppl_1), 319–328. http://dx.doi.org/10.1093/jexbot/51.suppl_1.319.

Poolman, M.G., Sebu, C., Pidcock, M.K., Fell, D.A., 2007. Modular decomposition of metabolic systems via null-space analysis. J. Theoret. Biol. 249 (4), 691–705. http://dx.doi.org/10.1016/j.jtbi.2007.08.005.

Raines, C.A., Harrison, E.P., Ölçer, H., Lloyd, J.C., 2000. Investigating the role of the thiol-regulated enzyme sedoheptulose-1, 7-bisphosphatase in the control of photosynthesis. Physiol. Plant. 110 (3), 303–308. http://dx.doi.org/10.1111/j.1399-3054.2000.1100303.x.

Raven, J.A., 2009. Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. Aquat. Microb. Ecol. 56 (2–3), 177–192. http://dx.doi.org/10.3354/ame01315.

Rohwer, J.M., 2012. Kinetic modelling of plant metabolic pathways. J. Exp. Bot. 63 (6), 2275–2292. http://dx.doi.org/10.1093/jxb/ers080.

Saadat, N.P., Nies, T., van Aalst, M., Hank, B., Demirtas, B., Ebenhöh, O., Matuszyńska, A., 2021. Computational analysis of alternative photosynthetic electron flows linked with oxidative stress. Front. Plant Sci. 12, 750580. http://dx.doi.org/10.3389/fpls.2021.750580.

Stiehl, H., Witt, H.T., 1969. Quantitative treatment of the function of plastoquinone in photosynthesis. Z. Naturforsch. B 24 (12), 1588–1598. http://dx.doi.org/10.1515/znb-1969-1219.

Stitt, M., Lilley, R.M., Heldt, H.W., 1982. Adenine nucleotide levels in the cytosol, chloroplasts, and mitochondria of wheat leaf protoplasts. Plant Physiol. 70 (4), 971–977. http://dx.doi.org/10.1104/pp.70.4.971.

Stitt, M., Schulze, D., 1994. Does rubisco control the rate of photosynthesis and plant growth? An exercise in molecular ecophysiology. Plant Cell Environ. 17 (5), 465–487. http://dx.doi.org/10.1111/j.1365-3040.1994.tb00144.x.

Uys, L., Botha, F.C., Hofmeyr, J.H.S., Rohwer, J.M., 2007. Kinetic model of sucrose accumulation in maturing sugarcane culm tissue. Phytochemistry 68 (16–18), 2375–2392. http://dx.doi.org/10.1016/j.phytochem.2007.04.023.

van Aalst, M., Ebenhöh, O., Matuszyńska, A., 2021. Constructing and analysing dynamic models with modelbase v1. 2.3: a software update. BMC Bioinformatics 22 (1), 1–15. http://dx.doi.org/10.1186/s12859-021-04122-7.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods 17 (3), 261–272. http://dx.doi.org/10.1038/s41592-019-0686-2.

Wilken, S.E., Besançon, M., Kratochvíl, M., Kuate, C.A.F., Trefois, C., Gu, W., Ebenhöh, O., 2022. Interrogating the effect of enzyme kinetics on metabolism using differentiable constraint-based models. Metab. Eng. 74, 72–82. http://dx.doi.org/10.1016/j.ymben.2022.09.002.

Witzel, F., Götze, J., Ebenhöh, O., 2010. Slow deactivation of ribulose 1, 5-bisphosphate carboxylase/oxygenase elucidated by mathematical models. FEBS J. 277 (4), 931–950. http://dx.doi.org/10.1111/j.1742-4658.2009.07541.x.

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

The paper "Computational Analysis of Alternative Photosynthetic Electron Flows Linked With Oxidative Stress" investigates the ROS producing and scavenging reactions around PSI, shining 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. I was involved in this project during every step of the research process.

Publication	Computational Analysis of Alternative Photosynthetic Electron Flows Linked With Oxidative Stress
Author position	Third
Status	peer-reviewed and published
doi	10.3389/fpls.2021.750580

frontiers in Plant Science

ORIGINAL RESEARCH
published: 22 October 2021
doi: 10.3389/fpls.2021.750580

Computational Analysis of Alternative Photosynthetic Electron Flows Linked With Oxidative Stress

*Nima P. Saadat[1], Tim Nies[1], Marvin van Aalst[1], Brandon Hank[1], Büsra Demirtas[1], Oliver Ebenhöh[1,2] and Anna Matuszyńska[1,2]**

[1] Institute of Quantitative and Theoretical Biology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany, [2] CEPLAS - Cluster of Excellence on Plant Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

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)

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, Ebenhöh 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

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$_{b_6f}$, and Photosystem I (PSI) to the final electron acceptor NADP$^+$ 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_2 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 $O_2^{-\cdot}$, 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_2O_2) 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_2O_2 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 disproportion 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; PPFD, 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.

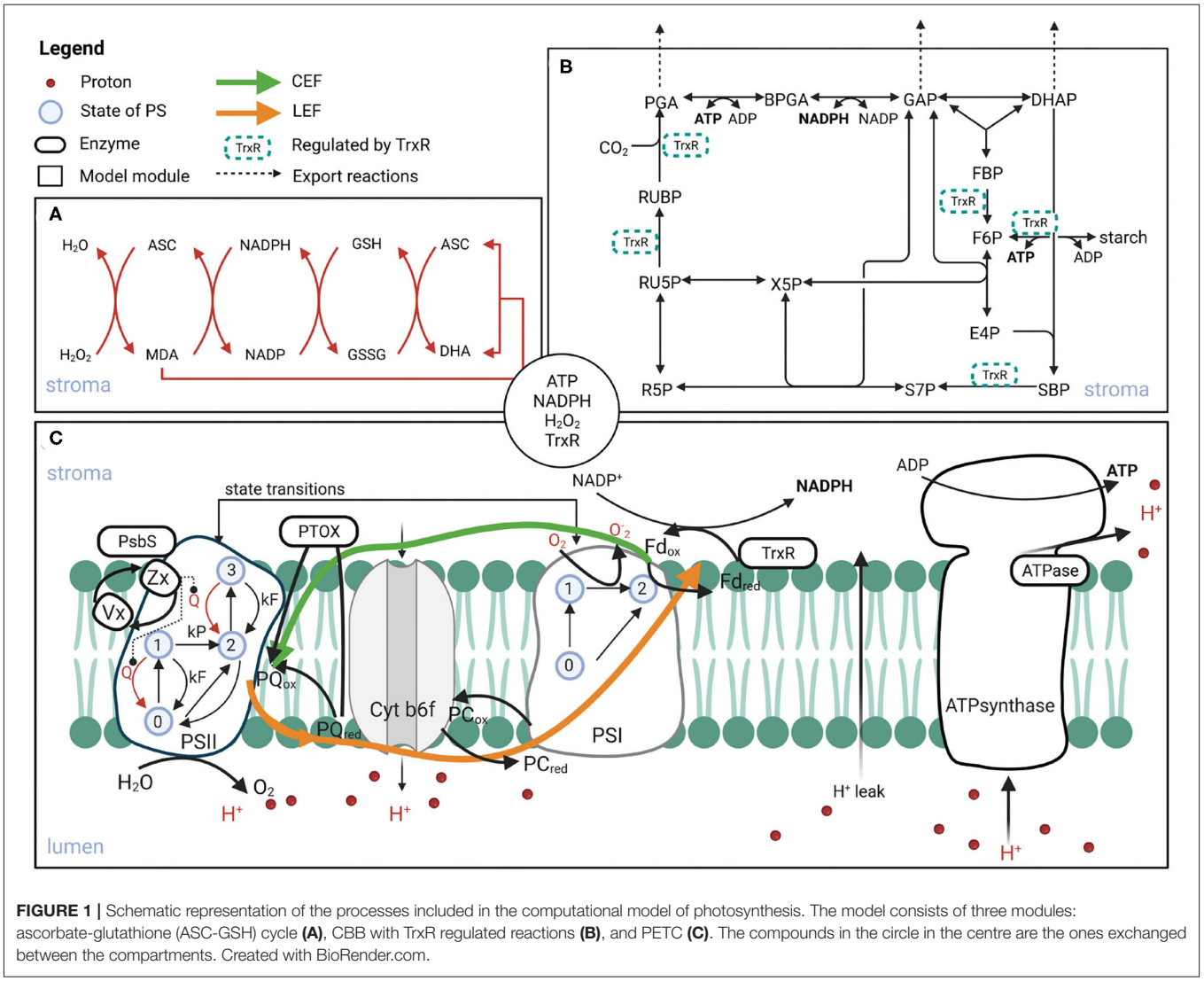


FIGURE 1 | Schematic representation of the processes included in the computational model of photosynthesis. The model consists of three modules: ascorbate-glutathione (ASC-GSH) cycle **(A)**, CBB with TrxR regulated reactions **(B)**, and PETC **(C)**. The compounds in the circle in the centre are the ones exchanged between the compartments. Created with BioRender.com.

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

$$H_2O + 2h\nu \rightarrow 2e^- + \frac{1}{2}O_2 + 2H^+_{\text{lumen}}, \quad (1)$$

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 lumenal compartments. The concentrations in the lumen are expressed in mmol (mol Chl)$^{-1}$ and inside the stroma in mM. To convert the concentrations of ATP, NADPH and H_2O_2 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)$^{-1}$ corresponds to $3.2 \cdot 10^{-5}$ 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 ($C^{J_j}_{v_k}$) and concentration ($C^{S_j}_{v_k}$) control coefficients are defined as

$$C^{J_j}_{v_k} = \frac{v_k}{J_j}\frac{\partial J_j/\partial p}{\partial v_k/\partial p}, \quad (2)$$

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

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 ±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 μmol m^{-2}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 (H_2O_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

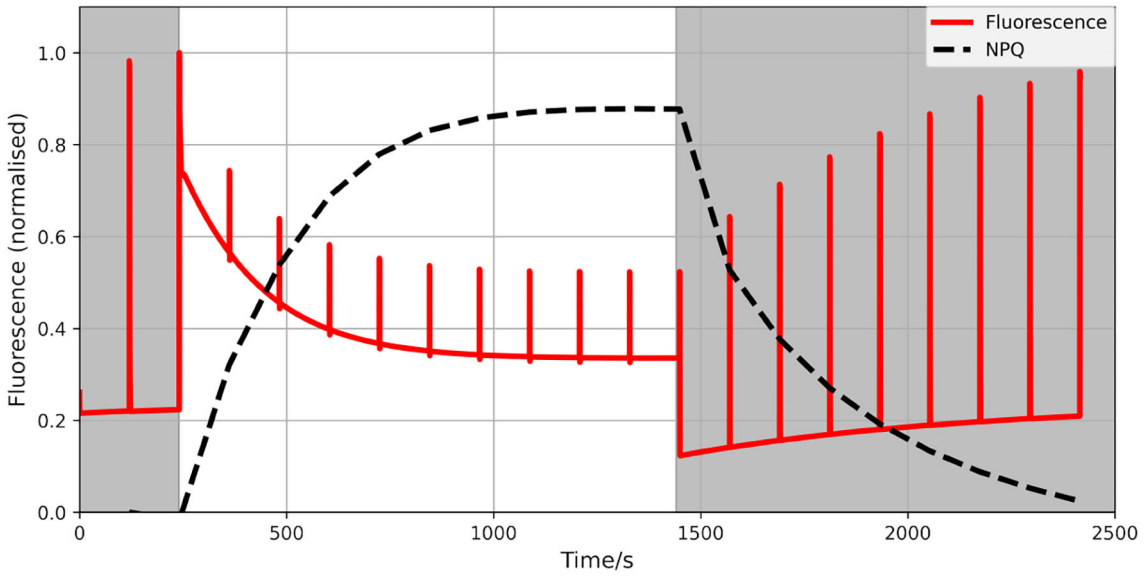


FIGURE 2 | Calculated fluorescence trace (red) of a PAM experiment of a generic photosynthetic organism without CEF. A standard dark-light-dark protocol was simulated (dark phase shaded). We used 1,000 μmol m^{-2}s^{-1} photosynthetic photon flux density (PPFD) as the intensity of the actinic light and 40 μmol m^{-2}s^{-1} PPFD for dark/dimmed light with pulses every 2 min. The calculated NPQ is marked in black (dashed). It exhibits well-known dynamics of excitation and relaxation.

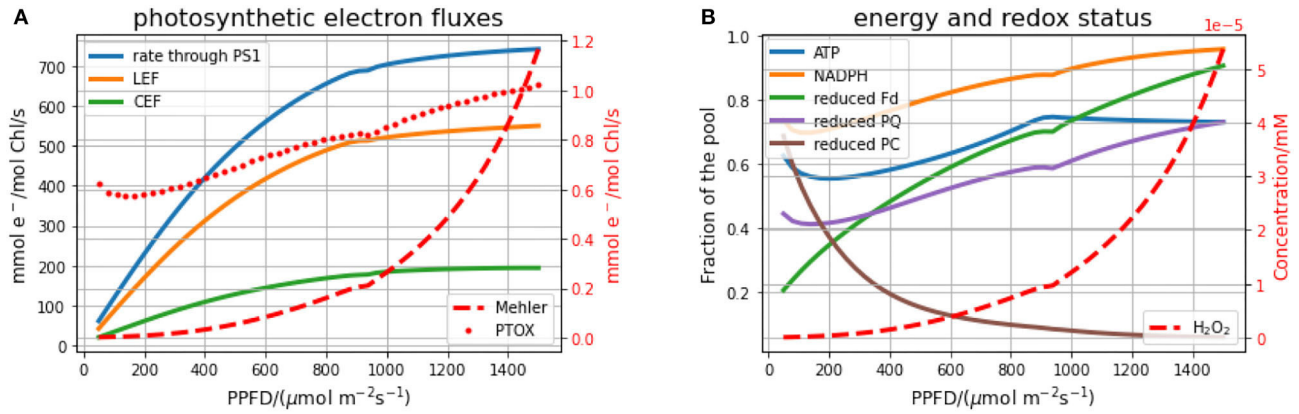


FIGURE 3 | Stationary fluxes and stationary energy and redox status of the electron transport chain for different intensities of constant light. **(A)** displays the electron flux through PSI, linear (LEF) and cyclic electron flow (CEF), the Mehler reaction and the plastid terminal oxidase (PTOX). **(B)** displays the energy equivalents (ATP, NADPH) and redox states of the electron chain (PQ, PC, Fd), as well as the stationary H_2O_2 concentration resulting from the Mehler reaction.

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 μmol 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 μmol 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.

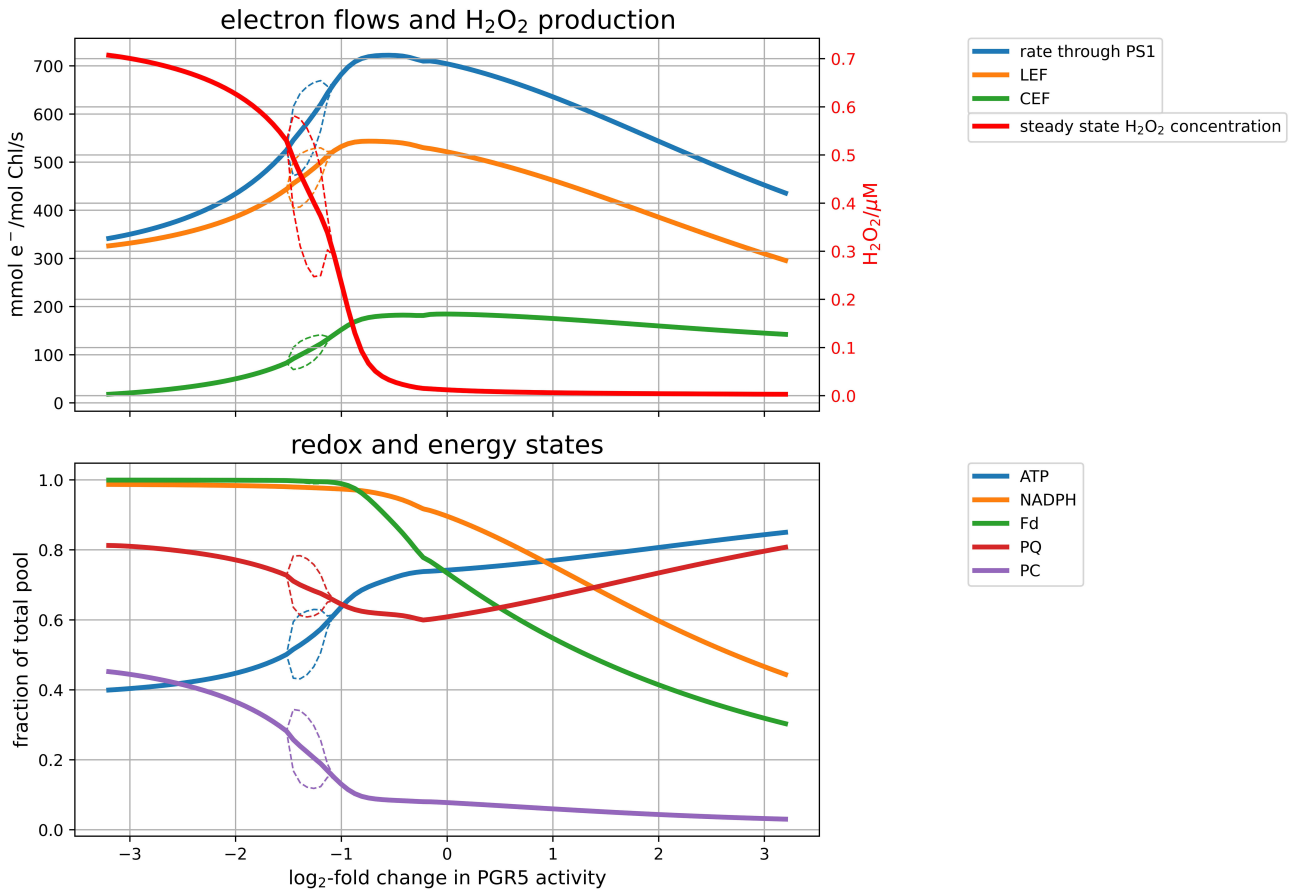


FIGURE 4 | Response of the model under high light (PPFD 1,000 μmol m^{-2}s^{-1}) to changes in CEF activity. Altered CEF activity was simulated by changing the rate constant for PGR5, the enzyme transferring electrons from Fd to PQ. The top panel displays electron flows through the PETC and the H_2O_2 concentrations resulting from the flux through the Mehler reaction and the ASC-GSH cycle. The bottom panel shows the energy (ATP) and redox state (NADPH, Fd, PQ, PC) of the system. In both panels, the solid lines indicate stationary values. The thin dashed lines indicate a parameter range, in which limit cycle oscillations were observed, denoting the minimum and maximum values of the oscillating variable. Outside these parameter regions, the solid line indicates stationary values, within the bubble averages over oscillations.

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 μmol m^{-2}s^{-2}) and low (PPFD 100 μmol m^{-2}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 golbal 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

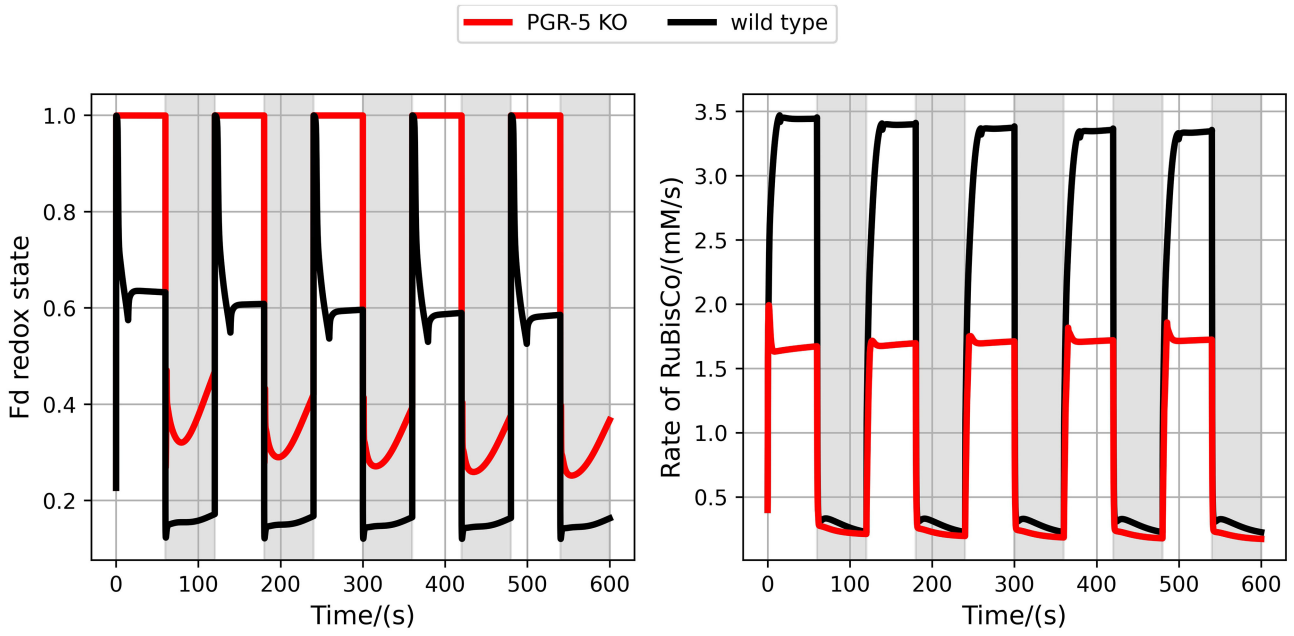


FIGURE 5 | Simulating fluctuating light for the wildtype and the PGR5 knockout (KO) mutant. Shown are the Fd redox state **(left)** and the RuBisCO activity **(right)** for the wildtype (black) and the PGR5 knockout (red). Light intensities were changed every 60s between high light (PPFD 600 μmol m^{-2}s^{-1}, white regions) and low light (PPFD 40 μmol m^{-2}s^{-1}, grey regions).

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 H_2O_2 concentrations in response to changed light intensities and PGR5 activities. Whereas under low light conditions (of less than approximately 500 μmol m^{-2}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 H_2O_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-Schussele et al., 2020). Although it is only one of the many known pathways (Maurino and Flügge, 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).

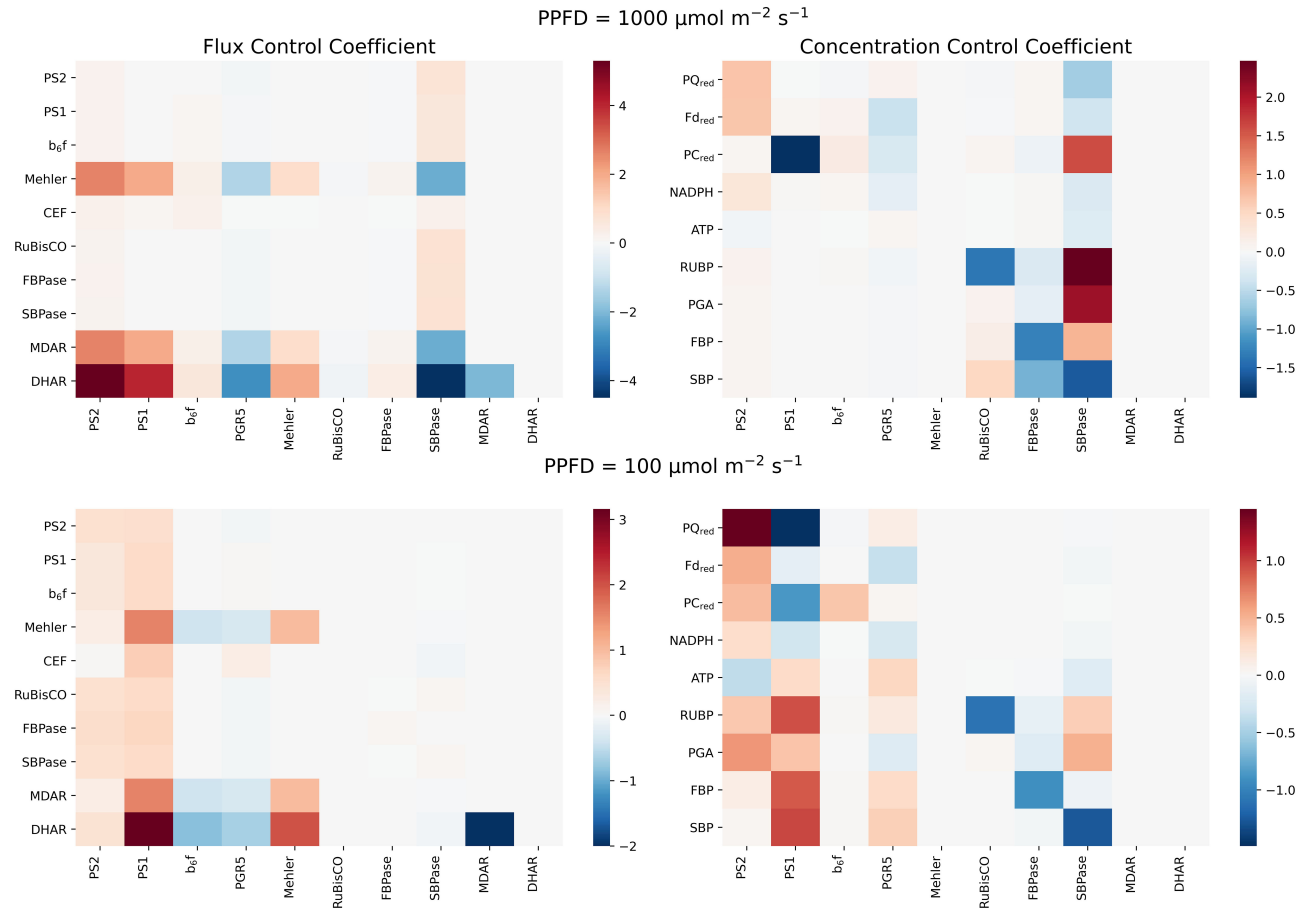


FIGURE 6 | Results of the MCA. Flux **(left)** and concentration **(right)** control coefficients of representative reactions of photosynthesis in high and low light conditions. The top panels show the distribution of control under high light (PPFD 1,000 μmol m^{-2}s^{-1}), the bottom panels under low light (PPFD 100 μmol m^{-2}s^{-1}). For all panels, on the x-axis we marked the perturbed parameters. The parameters are perturbed by $\pm$ 1% and the resulting effect on steady state values is monitored for the quantities on the y-axis. It can be clearly seen how the control of photosynthesis shifts from the supply side in low light **(bottom)** to the demand side under high light, exhibiting striking control of SBPase **(top)**.

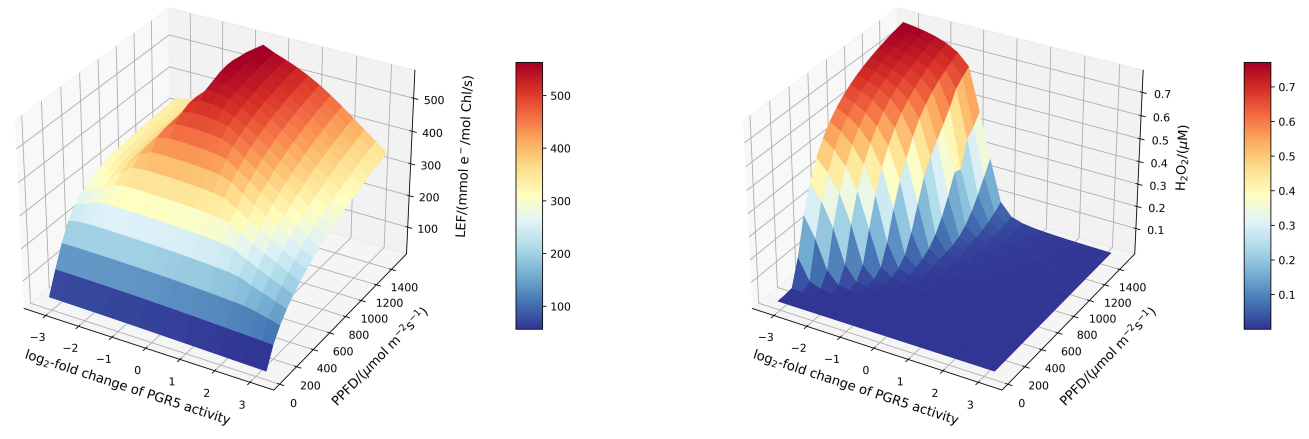


FIGURE 7 | Surface plots of the stationary linear electron flux **(left)** and H_2O_2 concentration **(right)** in response to altered PGR5 (CEF) activity and light intensity. Increased CEF activity reduces stationary H_2O_2 concentrations. For higher light intensities, more CEF activity is required to maintain low H_2O_2 levels. This indicates that CEF activity should be regulated for maximal efficiency under various light regimes.

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_2 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 $^1O_2$ 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_2 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

Alric, 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., Aseeva, 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-Liszkay, 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 1. *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.bbabio.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 O2-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., Głowacka, 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 Ebenhöh, 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 Garcia-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., Macia, 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., Ebenhöh, 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.bbabio.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 perioxide 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.

3.7 Shifts in carbon partitioning by photosynthetic activity increase terpenoid synthesis in glandular trichomes

The paper "Shifts in carbon partitioning by photosynthetic activity increase terpenoid synthesis in glandular trichomes" investigates the bioenergetics of photosynthetic glandular trichomes, especially regarding the production of terpenes. In it we provide the first reconstruction of specialised metabolism in Type-VI photosynthetic glandular trichomes of *Solanum lycopersicum* and predict that increasing light intensities results in a shift of carbon partitioning from catabolic to anabolic reactions driven by the energy availability of the cell. Moreover, we show the benefit of shifting between isoprenoid pathways under different light regimes, leading to a production of different classes of terpenes. Our computational predictions were confirmed *in vivo*, demonstrating a significant increase in production of monoterpenoids while the sesquiterpenes remained unchanged under higher light intensities. I was involved in this project during every step of the research process, as well as writing the manuscript.

Publication	Shifts in carbon partitioning by photosynthetic activity increase terpenoid synthesis in glandular trichomes
Author position	Shared first
Status	peer-reviewed and published
doi	10.1111/tpj.16352

the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2023) **115,** 1716–1728 doi: 10.1111/tpj.16352

Shifts in carbon partitioning by photosynthetic activity increase terpenoid synthesis in glandular trichomes

Nima P. Saadat[1,2,*], Marvin van Aalst[1,*], Alejandro Brand[3], Oliver Ebenhöh[1,2], Alain Tissier[2] and Anna B. Matuszyńska[2,4,*]

[1]*Institute of Theoretical and Quantitative Biology, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany,*
[2]*Cluster of Excellence on Plant Sciences, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany,*
[3]*Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle, Germany, and*
[4]*Computational Life Science, Department of Biology, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany*

Received 16 December 2022; accepted 8 June 2023; published online 20 June 2023.
*For correspondence (e-mail nima.saadat@hhu.de; marvin.van.aalst@hhu.de; anna.matuszynska@rwth-aachen.de).

SUMMARY

Several commercially important secondary metabolites are produced and accumulated in high amounts by glandular trichomes, giving the prospect of using them as metabolic cell factories. Due to extremely high metabolic fluxes through glandular trichomes, previous research focused on how such flows are achieved. The question regarding their bioenergetics became even more interesting with the discovery of photosynthetic activity in some glandular trichomes. Despite recent advances, how primary metabolism contributes to the high metabolic fluxes in glandular trichomes is still not fully elucidated. Using computational methods and available multi-omics data, we first developed a quantitative framework to investigate the possible role of photosynthetic energy supply in terpenoid production and next tested experimentally the simulation-driven hypothesis. With this work, we provide the first reconstruction of specialised metabolism in Type-VI photosynthetic glandular trichomes of *Solanum lycopersicum*. Our model predicted that increasing light intensities results in a shift of carbon partitioning from catabolic to anabolic reactions driven by the energy availability of the cell. Moreover, we show the benefit of shifting between isoprenoid pathways under different light regimes, leading to a production of different classes of terpenes. Our computational predictions were confirmed *in vivo*, demonstrating a significant increase in production of monoterpenoids while the sesquiterpenes remained unchanged under higher light intensities. The outcomes of this research provide quantitative measures to assess the beneficial role of chloroplast in glandular trichomes for enhanced production of secondary metabolites and can guide the design of new experiments that aim at modulating terpenoid production.

Keywords: bioenergetics, glandular trichomes, photosynthesis, stoichiometric model, secondary metabolites, terpenes, tomatos.

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) (Werker, 2000). While 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, the production of some metabolites can reach up to 20% of the leaf dry weight (Fobes et al., 1985), GTs are often referred to as true **metabolic cell factories** (Huchelmann et al., 2017). Products of GTs include terpenoids, phenylpropanoids, flavonoids, fatty acid derivatives and acyl sugars (Glas et al., 2012) exhibiting antifungal, insecticide or pesticide properties. Therefore, 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 (Balcke et al., 2017). 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 (Kortbeek et al., 2016). Although these pathways are thought to be largely independent, some exchange of precursors may occur (Paetzolda et al., 2010), and such crosstalk requires further investigation. For instance, is there some crosstalk of plastidial and cytosolic pathways providing the 5-carbon precursors, as suggested in the latest work in peppermint (Koley et al., 2020)? 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* [Bergau et al., 2015]). Considering that in the case of many plants where the seeds are green during embryogenesis, the light can influence the fatty acid synthesis, and potentially power refixation of CO_2 (Goffman et al., 2005; Ruuska et al., 2004), we took the challenge to understand whether a similar **effect of light** can be predicted in trichomes. Till now it is still unclear what the advantages and disadvantages of photosynthetic GTs are in contrast to non-photosynthetic GTs. Moreover, there is limited research on GTs ability to absorb light. Even research focused on light intensity-mediated trichome production does not focus on their photosynthetic activity (Escobar-Bravo et al., 2018). 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 (Brand & Tissier, 2022).

To shed light on the advantages of photosynthetic GT for terpenoid synthesis and secondary metabolism, investigations of the system's bioenergetics and reaction flux distributions are needed. Mathematical, computational models provide a coherent framework to study metabolism. Constraint-based stoichiometric models (Maarleveld et al., 2013) 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 (Heinrich & Schuster, 1996). 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* (Chatterjee et al., 2017), *Arabidopsis thaliana* (Poolman et al., 2009) and *Solanum lycopersicum L.* (Yuan et al., 2016) among many others. Flux balance analysis (FBA) (Orth et al., 2011; Sweetlove & Ratcliffe, 2011), 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 optimised in a linear programming 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 type VI of a *Solanum lycopersicum* LA4024 (see schematic representation on Figure 1) using previously published transcriptome and metabolome data (Balcke et al., 2017). With a general, mathematical framework, we investigated the effect of having photosynthetically active machinery inside of a trichome and systematically tested the model on how GTs achieve high metabolic productivity proposed by Balcke et al., (2017). 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. It complements the previous work by Balcke et al., (2017) by not only confirming their hypothesis that the light-dependent reactions of photosynthesis support the secondary metabolite pathways but explaining how this support is achieved. Finally, our predictions were tested *in vivo* and we provide an experimental validation by showing that under high light conditions, production of the most abundant MEP-derived terpenes (2-carene and β-phellandrene/D-limonene) increases, while the most abundant

MEV-derived terpenes (β-caryophyllene and α-humulene) remain unchanged.

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) (Lewis et al., 2010). Figure 2 displays that with increasing light absorption, the rate of terpenoid synthesis in photosynthetic GTs increases up until approximately 50 $\frac{\mu molPhotons}{s \cdot m^2}$. This increase in terpenoid synthesis rate with increasing absorbed light is particularly interesting due to the fact that the model cannot utilise atmospheric carbon dioxide, and sucrose is the only carbon

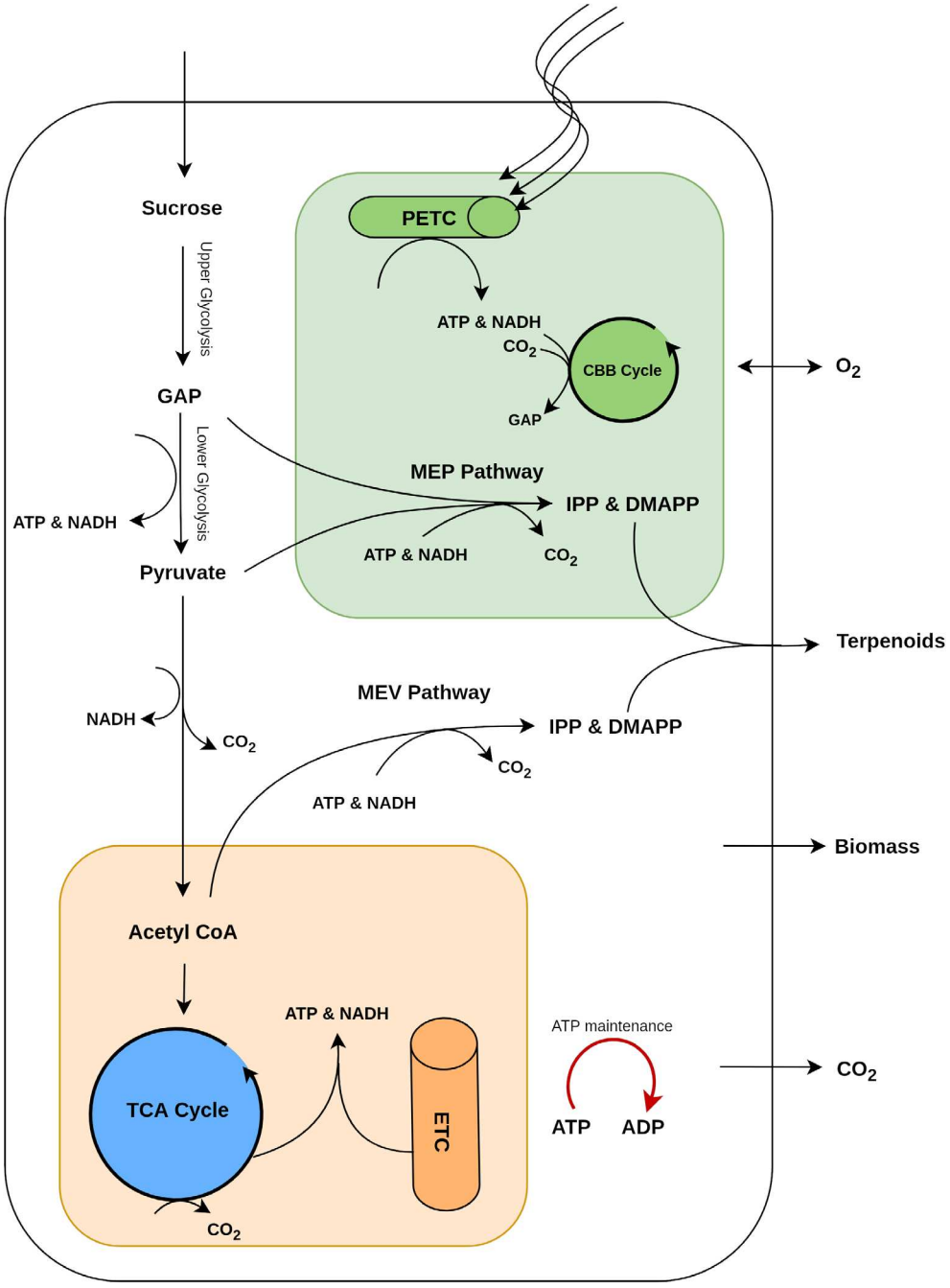


Figure 1. Schematic overview of the key processes included in a constraint-based model of photosynthetic glandular trichome (GT) metabolism. While the model is built using transcriptome and metabolome data and includes a large number of reactions, only pathways and metabolites of importance to the results are highlighted in the presented model scheme. These include CBB Cycle, Calvin-Benson-Bassham Cycle; DMAPP, dimethylallyl diphosphate; IPP, isopentenyl diphosphate; MEP, methyl-erythritolphosphate pathwy; MEV, mevalonate pathway; PETC, photosynthetic electron transfer chain; TCA Cycle, tricarboxylic acid Cycle.

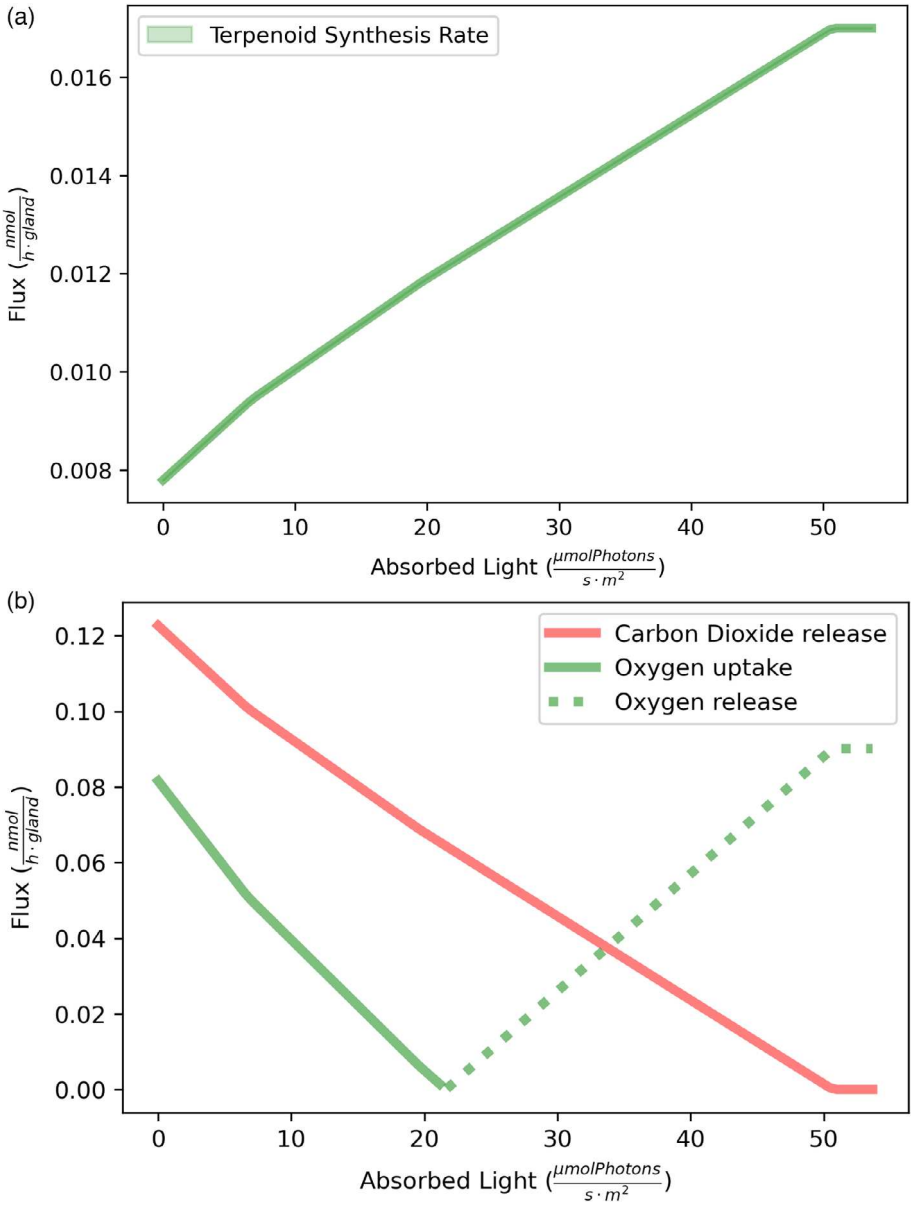


Figure 2. Impact of rates of absorbed light on the predicted fluxes through the photosynthetic glandular trichome.
(a) Terpenoid synthesis flux over different rates of absorbed light. The rate terpenoid synthesis increases with higher amounts of absorbed light.
(b) Oxygen and carbon dioxide exchange fluxes over different rates of absorbed light. Increased light absorption decreases carbon dioxide release and oxygen absorption. In higher rates of light absorption, oxygen is secreted.

source. This means that there is a change in metabolic fluxes that enables this increase in terpenoid synthesis rate. To further investigate what changes in the metabolism of photosynthetic GTs in increasing light intensities, we inspect the respective changes in the exchange fluxes of the model. Figure 3 shows the exchange fluxes of carbon dioxide and oxygen in our pFBA model simulations over increasing light absorptions. Noticeably, the release of carbon dioxide systematically decreases up until approximately 50 $\frac{\mu molPhotons}{s \cdot m^2}$. Interestingly, the consumption of oxygen decreases to zero at approximately 21 $\frac{\mu molPhotons}{s \cdot m^2}$. From this light intensity on, oxygen release begins and increases until 50 $\frac{\mu molPhotons}{s \cdot m^2}$. 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 release at 21 $\frac{\mu molPhotons}{s \cdot m^2}$ 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.

The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd.,
The Plant Journal, (2023), **115**, 1716–1728

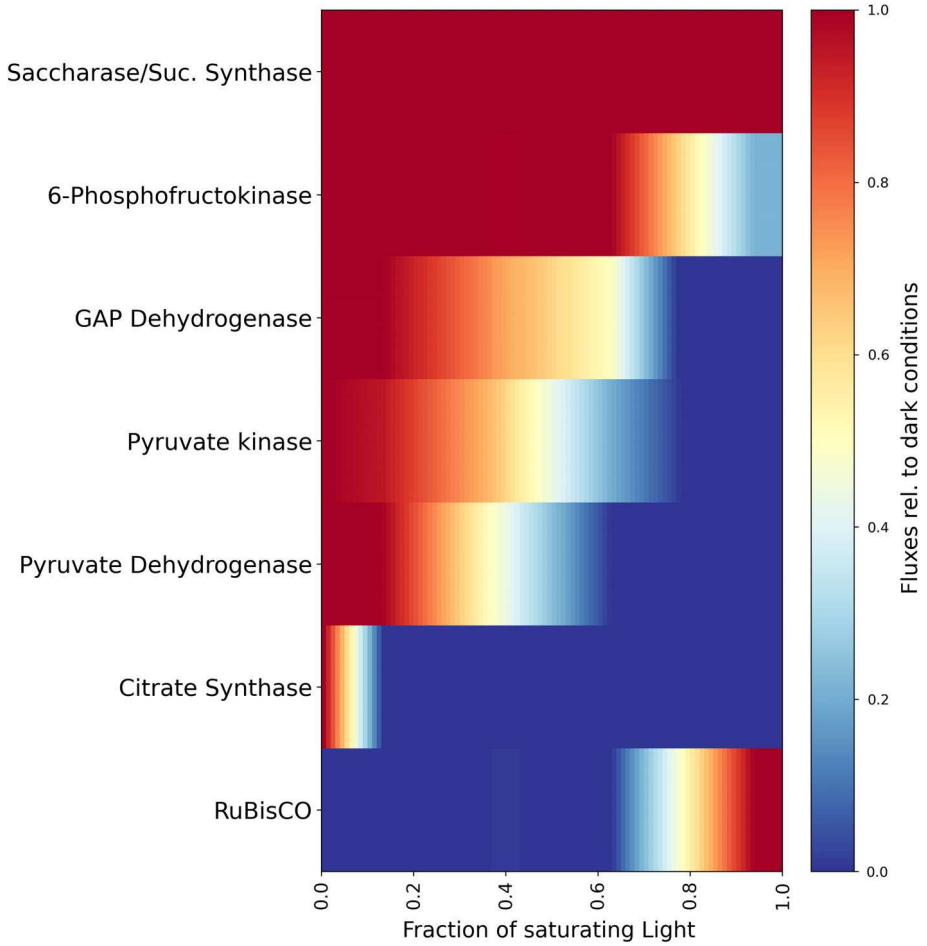


Figure 3. The relative fluxes of six selected catabolic reactions and one carbon fixation reaction calculated for increasing fractions of saturating light. The fluxes are normalised to the respective fluxes under completely dark conditions. Reactions of upper glycolysis remain unaffected by increased light absorption, while the fluxes of reactions of lower glycolysis and the TCA cycle decrease with increased absorbed light. Additionally, RuBisCO flux is only present when light absorption is high.

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. Figure 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-phosphofructokinase, 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. There is no reason to think that trichomes would not have the same regulatory mechanisms as mesophyll cells, leading to decreased glycolysis in higher light conditions. Increased photosynthesis is accompanied by increased photorespiration, which eventually supplies reducing equivalents in the mitochondria, which can then be used to fuel the respiratory electron transport chain (Balcke et al., 2023). The increase in terpenoid synthesis flux observed in Figure 2. and the decrease in catabolic fluxes in Figure 3 strongly

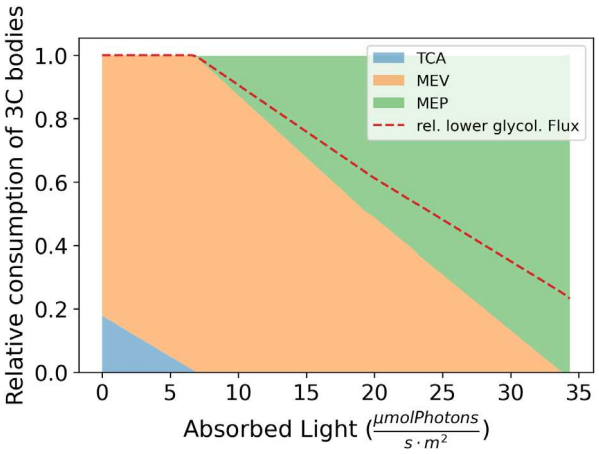


Figure 4. Predicted relative consumption of GAP, pyruvate and acetyl-CoA (here described as 3C bodies) by different pathways over increasing fractions of saturating light. The fraction of the lower glycolysis flux relative to dark conditions is displayed as a dashed line. Increasing rates of absorbed light decrease the fraction of 3C bodies consumed by catabolic pathways, like lower glycolysis and TCA cycle. Furthermore, the fraction of 3C bodies consumed for terpenoid synthesis switches from the MEV to MEP pathway in high rates of light absorption.

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 GTs. 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 only very high levels of photosynthetic energy supply allows 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, also known as the isoprenoid pathway, which is the primary terpenoid synthesis pathway in non-photosynthetic GTs. GAP and pyruvate are metabolites within the lower glycolysis pathway and also initial substrates of the MEP, which is a terpenoid synthesis pathway present in 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. Figure 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 Figure 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, for example, 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 that the distribution of terpenoid synthesis between the MEV and the MEP pathways is highly dependent on the light conditions and resulting energy levels of the photosynthetic GTs.

The general conclusions of our model simulations have then been tested experimentally. The impact of light

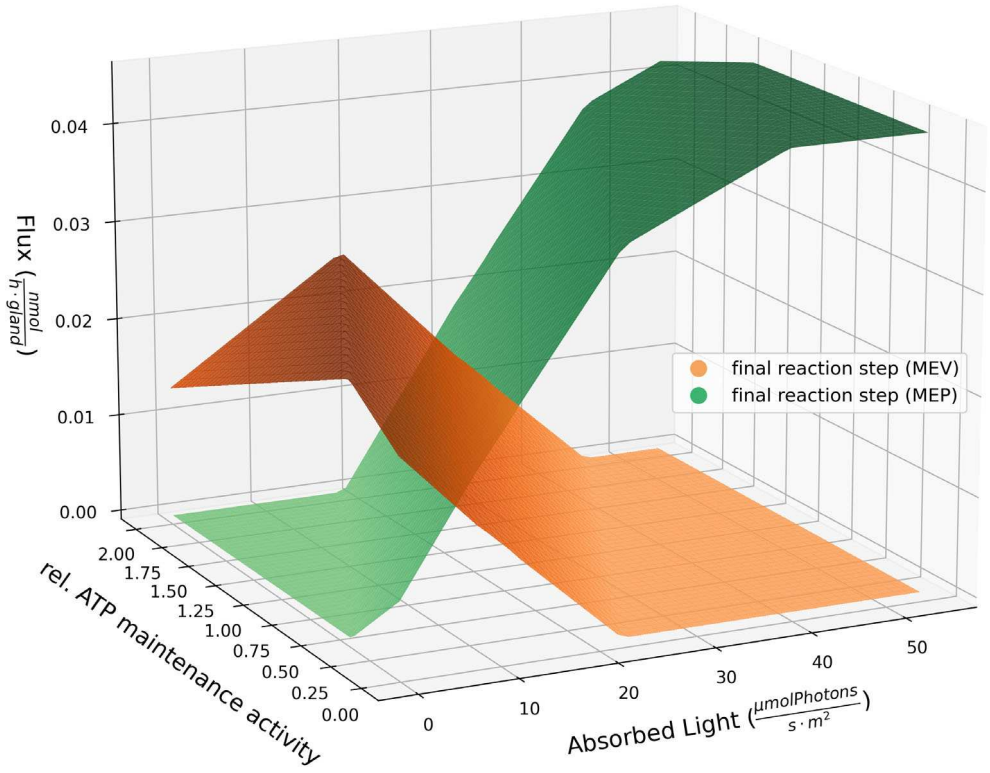


Figure 5. Fluxes of the final reaction steps of the MEV and MEP pathway over increasing relative ATP maintenance activities, as well as increasing rates of light absorption. In energetically favourable conditions, like high light and low ATP maintenance (function representing the additional energy requirement for the maintenance of cells), terpenoid synthesis is carried out by the MEP pathway. In opposite conditions, meaning low light and high ATP maintenance, the MEV pathway is performing terpenoid synthesis.

intensity on the shift from the MEV to MEP precursor pathway has been tested by quantifying sesquiterpenes, produced by MEV, and monoterpenes, produced by MEP. Three-week-old tomato (*Solanum lycopersicum* cv Moneymaker) plants were exposed to nearly threefold higher light intensity (HL) for seven days and the productivity of the main volatile terpenoids produced in the type VI GTs was estimated by GC–MS along different leaf developmental stages. The results showed a significant increase in monoterpenoids while the sesquiterpenes remained unchanged, and such increment is linked to the age of the leaves (Figure 6). These findings suggest that the photosynthetic light reactions support the productivity of the specialised reactions occurring in the plastids through the MEP pathway.

Finally, the high rate of terpenoid synthesis in high-light conditions is partly 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 quanta of **absorbed light** and systematically changed the activities of RuBisCO (Figure 7). Interestingly, the overall rate of carbon refixation is increasing the rate of terpenoid synthesis by almost 20%, while the shift in carbon partitioning between catabolism and anabolism increases it by almost 200%. This shows that the impact of energy-dependent shift in carbon partitioning and isoprenoid synthesis pathways is a lot higher than the RuBisCO-dependent refixation of carbon dioxide.

DISCUSSION

In photosynthetic GTs, synthetic pathways of terpenoids and other secondary metabolites are found in the cytosol of the cells and the chloroplasts. The additional terpenoid synthesis pathway in photosynthetic GTs has been subject to many speculations, for example, terpenoid production in chloroplasts is specialised for the production of particular secondary metabolites (Besser et al., 2009). In our work, we built a simplified, yet data-driven constraint-based model of photosynthetic glandular trichome metabolism, and used it to show that one of the two different synthesis pathways is more advantageous for terpenoid production than the other in different energy availabilities. Previously published multi-omics data (Balcke et al., 2017) supports our hypothesis that the energy and reducing power (ATP and NADPH) from photosynthesis are primarily used to power the secondary

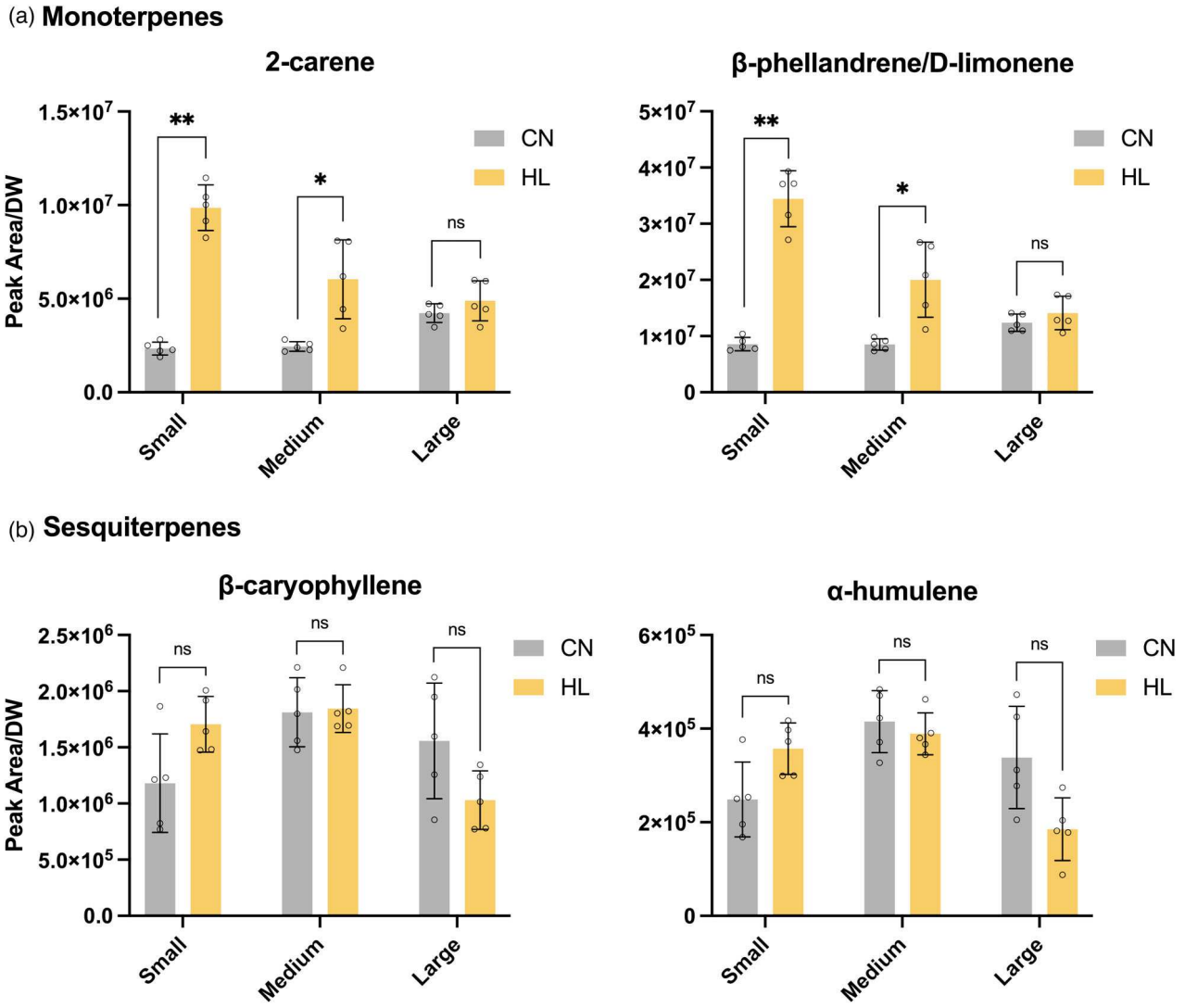


Figure 6. Effect of high light in volatile terpene content of leaves of tomato. Estimation of (a) monoterpenes and (b) sesquiterpenes by gas chromatography–mass spectrometry (GC–MS) on young (Small, from bottom-to-top leaf number 6), expanded (Medium, leaf number 5) and fully develop (Large, leaf number 4) leaves, in control (CN) and high light (HL) conditions. Chromatogram peak areas were normalised by leaf dry weight (DW). Error bars indicate SD ($n = 5$ biological replicates; $^{*}P < 0{,}05$, $^{**}P < 0{,}01$; using t-test).

metabolism. Our model provides a mechanistic explanation of how this is achieved.

We show that with lower energy availability, the cytosolic MEV pathway is more advantageous for terpenoid synthesis because the catabolic pathways, producing the critical 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 shortcut of catabolic reactions reduces the loss of carbon as carbon dioxide and increases the flux of carbon through anabolic processes. The general conclusions derived from the theoretical analyses have been strengthened with the experimental evidence that under higher light intensities production of monoterpenoids is significantly increased, while the production of sesquiterpenes remains unchanged (Figure 6).

We show that in higher light conditions, energy levels of the photosynthetic GTs are so advantageous, that excess energy can be spent to perform carbon refixation using the CBB cycle. In the supplementary material, we calculated that the terpenoid yield per sucrose is twice as high in high light compared to low light conditions. This

illustrates that the benefit of including chloroplasts in GTs is not only the ability to shift carbon partitioning from catabolic to anabolic processes but also to further maximise 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 Figure 7. This dual behaviour, that is, in the absence versus in presence of photosynthesis, is reminiscent of the situation in photosynthetic leaves, where the TCA cycle is inactive during the day and active during the night (Tcherkez & Atkin, 2021).

Interestingly our model shows that even without CBB cycle activity, the TCA cycle may be reversed in high energy availability and function as a reductive TCA cycle. This reductive TCA cycle could theoretically take over the function of the CBB cycle, using energy to fix carbon dioxide which was produced in catabolic and anabolic reactions, thus increasing carbon use efficiency. This is a very interesting observation, as, from a bioenergetic point of view, such a scenario is possible. Considering the previous results where Phosphoenolpyruvate carboxylase (PEPC) expression was significantly increased in trichomes compared to leaves (Balcke et al., 2017), we consider PEPC as a plausible candidate to mediate the carbon fixation. However, we decided to adjust the key reactions of the TCA cycle for this scenario as irreversible 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 sulphur bacteria and different thermophilic prokaryotes and archaea (Beh et al., 1993; Wahlund & Tabita, 1997). 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 (Schuurink & Tissier, 2020). 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 a solar panel 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. Interestingly, such an increase in the efficiency of carbon use through RubisCO, but without the CBB cycle, has been observed in other plant cells. Schwender *et al.* (Schwender et al., 2004) showed that Rubisco without the Calvin cycle improves the carbon efficiency of developing embryos of *Brassica napus L.* (oilseed rape) during the formation of oil.

Photosynthetic carbon refixation indicates that photorespiration may be present in photosynthetic GTs. Although photorespiratory genes were very low expressed in the transcriptome data (Balcke et al., 2017), and photorespiration is not included in our model, we show that in high light there is oxygen evolution in photosynthetic GTs. Therefore, new experimental data obtained in high light intensities and gas exchange rates is required to investigate putative photorespiratory activities. Furthermore, it remains unclear if and how high the 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. For example, what is the composition of terpenoids under different light intensities, or even light colours? A recent study using different basil cultivars showed that light spectra affect the concentrations and volatile emissions of important compounds (Kivimäenpä et al., 2022). Such light modulation requires further investigation with a use of more detailed models of secondary metabolism in photosynthetic GTs that take the light spectrum into consideration. For transparency, we include the light spectra used for our experimental validation in Figure S1. As most of the processes discussed here are heavily dependent on enzyme kinetics and saturation, constraint-based models like the one presented may not be the best method for answering these new emerging questions. Mechanistic models, for example 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. Further interdisciplinary studies combining experiments and robust theoretical simulations can provide a quantitative understanding of whether, how, and by how much the production of specific terpenoids could be increased, and with this work, we provide the stepping stone for such analyses. The outcomes of this research provide quantitative measures to assess the beneficial role of chloroplast in GTs and can further guide the design of new experiments aiming at enhanced terpenoid production.

METHODS

Choice of the model organism

In this study, we have chosen to investigate type VI GT in the tomato genus. The tomato genus displays seven types of trichomes: II, III and V (non-glandular) and I, IV, VI and VII (glandular trichomes), with type VI being the most abundant one in the *Solanum lycopersicum* species. *S. lycopersicum* serves as an excellent model organism for glandular trichome study due to the availability of (i) high-quality complete genome sequence (Tomato Genome Consortium, 2012), (ii) excellent genetic resources (Falara

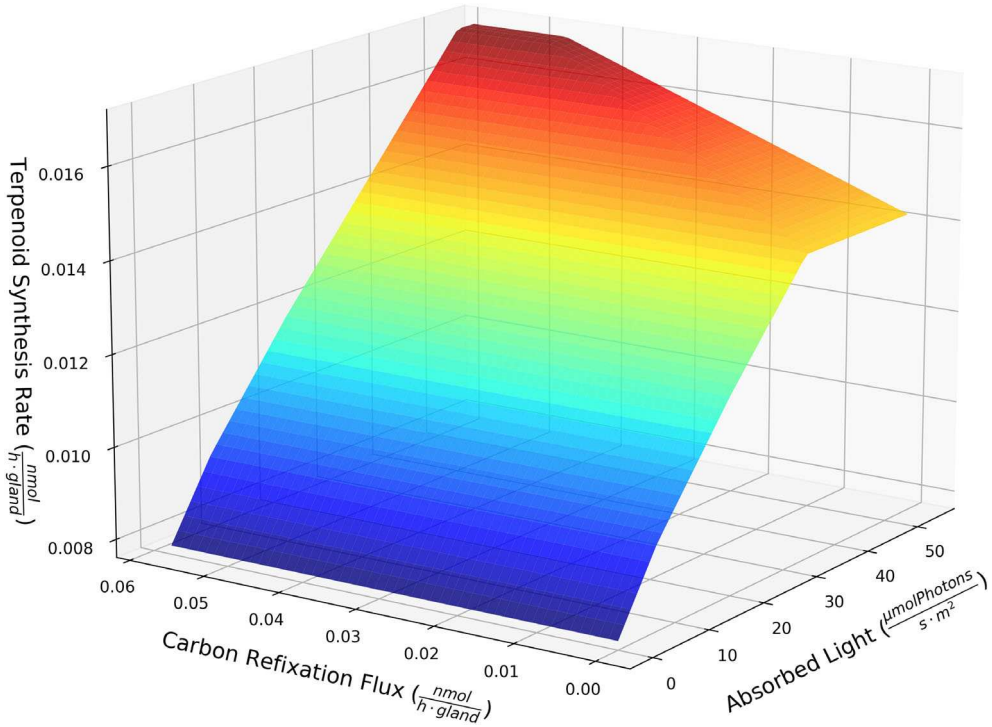


Figure 7. Predicted flux of the terpenoid synthesis under changing carbon refixation rates, as well as increasing rates of light absorption. Under simulated saturating light, the rate of terpenoid synthesis is two-fold higher than in darkness. If additionally the carbon refixation flux is increased to the maximal chosen value, the increase in terpenoid synthesis is less than 20%. This result shows how the energy-dependent shift in carbon partitioning and isoprenoid synthesis pathway is higher than the RuBisCO dependent refixation of carbon dioxide.

et al., 2011), (iii) comparative multi-omics data (Balcke et al., 2017), (iv) several mathematical models available, including whole genome metabolic network reconstruction (Yuan et al., 2016) and (v) in contrast to other well-studied organisms like peppermint (Rios-Estepa et al., 2008, 2010), possession of photosynthetic GT.

Modelling environment

Our model is implemented in Python, using our in-house developed package moped, "an integrative hub for reproducible construction, modification, curation and analysis of metabolic models" (Saadat et al., 2022). With moped all decision processes and modelling steps are well documented in a transparent and repeatable fashion. 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. The summary of the construction steps is provided in the section below.

Model construction and assumptions

Although a genome-scale model of tomato metabolism is available (iHY3410 model (Yuan et al., 2016)), we decided to use a bottom-up approach and perform the reconstruction ourselves, because 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 (Balcke et al., 2017), the LycoCyc database (tomato metabolic pathway database, version 3.3 (Fernandez-Pozo et al., 2015), available from Solanaceae Genomics Network, http://www.sgn.cornell.edu) and biochemical knowledge in plants from scientific publications. All reactions and metabolites found in transcriptomics and metabolomics data have been added to the model from the LycoCyc database using the moped metabolic modelling package, ensuring GPR rules for all added reactions in the network.

We used Meneco, a tool for metabolic network completion (Prigent et al., 2017) to subsequently fill gaps in our network with annotated reactions from the LycoCyc database, so our model is capable of synthesising all compounds found within the metabolomics data (Balcke et al., 2017), all terpenoids found in photosynthetic GTs of tomato (Besser et al., 2009) as well as all amino acids, nucleotide bases and lipid precursors from sucrose, light, orthophosphate, ammonia, sulfate, protons and water. All reactions in our model have been checked for mass and charge balance and are able to carry steady-state fluxes. Our model shows the ability to synthesise biomass precursors and terpenoids on a realistic scale. The model has been examined for inconsistencies in energy metabolism by analysing the model behaviour according to changes in ATP demand. Increasing ATP demand leads to plausible changes in key reactions of the model, such as a decrease in objective function flux. A detailed description of every implemented step in model construction, as well as the code for reproducing the entire model reconstruction process, can be found in our *model_construction.ipynb* notebook in the supplementary material.

Our model is a data-driven, yet simplified, constraint-based model which is ensured not to include 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, as

represented on Figure 1) is able to represent photosynthetic GT metabolic profiles. 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 (Gardeström & Igamberdiev, 2016). 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. The resulting model consists of 1307 reactions and 1371 metabolites and thanks to the integration of the multiomics data, its behaviour has been ensured to match reported experimental observations (Balcke et al., 2017). There are nine exchange reactions, allowing the free exchange of inorganic metabolites such as oxygen, as well as light absorption and sucrose uptake. To ensure the highest quality and consistency our model has been thoroughly inspected using the MEMOTE standarised testing suite (Lieven et al., 2020).

Optimisation and objective function

In most constraint-based models and their analyses, the maximisation objective is the production of biomass (Gottstein et al., 2016). 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, cell wall and fatty acid precursors. We used a simplified biomass function inspired by plant biomass functions from Seaver et al., (2015), similar to standard biomass functions used successfully for FBA in plants (Arnold & Nikoloski, 2014; Zager & Lange, 2018). Our aim was to capture the necessity for growth and self-repair, while setting the objective function to maximise the production of terpenoids. 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 (Cheung et al., 2013).

Calculation of flux units and light intensity units

There is limited research on GTs ability to absorb light (e.g. Conneely et al., 2021), and we have not found any dedicated research focused on the optical properties of GTs. Overall, while there is some evidence to suggest that GTs may have optical properties allowing for some light absorption, more research is needed to fully understand their ability to absorb light. Therefore we have decided to estimate the maximal absorption rate based on the reported maximal production fluxes. Light is therefore represented as **photons absorbed by the photosystems** used for photosynthesis in contrast to the incident light that will be several-fold higher.

Although it is known that due to diel cycles of photosynthesis different metabolic flux patterns in the light and the dark are observed and require different treatments for the optimisation problem (Cheung et al., 2014), our FBA on the trichome model is performed under continuous light. The units of light absorption are represented in $\frac{\mu mol\text{Photons}}{s\cdot m^2}$ and the detailed calculations are provided in the Supporting Information. As it has been reported that carbon dioxide exchange is 100 times lower in photosynthetic GTs than in leaves we decided not to include a carbon dioxide influx, however, carbon dioxide is produced in the system and can flow out of the model (Balcke et al., 2017).

A suggested terpenoid production rate of GTs has been provided by Turner *et al.* (Turner & Croteau, 2004) at 0.017 $\frac{nmol}{\text{h}\cdot\text{gland}}$. Assuming that this rate can be applied to the maximal terpenoid production rate of photosynthetic GTs of tomatoes, we transform our calculated fluxes to the corresponding units by $\text{Flux}\cdot\frac{0.017\frac{nmol}{\text{h}\cdot\text{gland}}}{\text{max.Terpflux}}$. Next, in order to convert the fluxes of photons into units of light intensities, we calculated the light-absorbing surface of the GT type VI. Although the head of GT type VI is made up of four secretory cells, we simplify the whole surface of the head as a sphere. Based on the measured values of the diameters of GTs from (Kowalski et al., 2019) and bright-field microscopic image from (Bergau et al., 2015) we took the estimate of 50 μm as the diameter. This number can be substituted with a different value and the light conversion function will be adapted. Under these assumptions, the surface area can be estimated as

$$A = 4 * \pi\left(\frac{50\,\mu m}{2}\right)^2 = 8000\,\mu m^2 = 8\cdot 10^{-9} m^2.$$

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

$$\frac{0.017 * \text{Max.LightFlux}}{\text{Max.Terp.Flux}} = \frac{0.017 * 480}{8.85} = 0.92\frac{nmol\text{Photons}}{\text{h}\cdot\text{gland}}.$$

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

$$\frac{1 nmol\text{Photons}}{\text{h}\cdot\text{gland}} = \frac{1}{28800}\frac{mol\ \text{Photons}}{s\cdot m^2} = 56\frac{\mu mol\text{Photons}}{s\cdot m^2}.$$

For our maximal light flux, this corresponds to $0.92\cdot 56\frac{\mu mol\text{Photons}}{s\cdot m^2} = 51.52\frac{\mu mol\text{Photons}}{s\cdot m^2}$ as the saturating light intensity, providing the light flux conversion factor of $\text{LightFlux}\cdot\frac{51.52\frac{\mu mol\text{Photons}}{s\cdot m^2}}{\text{sat.Lightflux}}$.

Experimental setup

The high light experiment was conducted using two LED-light panels (Rhenac GreenTech AG, Hennef, Germany) placed inside a phytochamber and separated by a black curtain. Tomato (*Solanum lycopersicum* cv Moneymaker) plants were germinated on soil under control conditions (CN): 16 h light, 422 μ mol $m^{-2}sec^{-1}$ at 25°C and 8 h dark at 20°C, 70% humidity day and night. After 22 days of growth, half of the plants were transferred to high light (HL) conditions for further 7 days, where light intensity was adjusted to 1289 μ mol m^{-2} sec^{-1}. Light spectra were recorded using a Specbos 1211UV (JETI, Jena, Germany) photometer (Figure S1). Volatile terpenoids of leaflets of three different developmental stages were collected by surface extraction using leaf discs of 1 cm diameter and vortexed for 30 sec in n-hexane. Mono- and sesquiterpenes were detected using a Trace GC Ultra gas chromatograph coupled with an ATAS Optic 3 injector and an ISQ mass spectrometer (Thermo Scientific) with electron impact ionisation. The chromatographic separation was performed on a ZB-5 ms capillary column (30 m × 0.32 mm, Phenomenex). The flow rate of helium was 1 ml min−1, and the injection temperature rose from 60 to 250°C at 10°C sec^{-1} during 30 sec. The GC oven temperature ramp was 50°C for 1 min, 50–150°C at 7°C min^{-1} and 150–300°C at 25°C min^{-1} for 2 min. Mass spectrometry was performed at 70 eV in full scan mode with

m/*z* from 50 to 450. Data analysis was done with the Xcalibur software (Thermo Scientific).

ACKNOWLEDGEMENT

This work was funded by the Deutsche Forschungsgemeinschaft Research Grantproject ID 420069095 (AB, AT, AM), Deutsche Forschungsgemeinschaft under Germany´s Excellence Strategy – EXC-2048/1 – project ID 390686111 (NS, OE, AM), and EU's Horizon 2020 research and innovation programme under the Grant Agreement 862 087 (MvA). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Light spectra of the two light conditions used in the experiment. Wavelength recorded from 230 to 1000 nm. CN, control; HL, high light.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://gitlab.com/qtb–hhu/models/glandular–trichomes.

REFERENCES

Arnold, A. & **Nikoloski, Z.** (2014) Bottom-up metabolic reconstruction of Arabidopsis and its application to determining the metabolic costs of enzyme production. *Plant Physiology*, **165**, 1380–1391. Available from: https://arxiv.org/abs/1209.3737

Balcke, G.U., Bennewitz, S., Bergau, N., Athmer, B., Henning, A., Majovsky, P. ***et al.*** (2017) Multi-omics of tomato glandular Trichomes reveals distinct features of central carbon metabolism supporting high productivity of specialized metabolites. *The Plant Cell*, **29**, 960–983.

Balcke, G.U., Vahabi, K., Giese, J., Finkemeier, I. & **Tissier, A.** (2023) The arabidopsis concert of metabolic acclimation to high light stress. *bioRxiv*. Available from: https://www.biorxiv.org/content/early/2023/02/15/2023.02.14.528433

Beh, M., Strauss, G., Huber, R., Stetter, K.-O. & **Fuchs, G.** (1993) Enzymes of the reductive citric acid cycle in the autotrophic eubacterium aquifex pyrophilus and in the archaebacterium thermoproteus neutrophilus. *Archives of Microbiology*, **160**, 306–311.

Bergau, N., Bennewitz, S., Syrowatka, F., Hause, G. & **Tissier, A.** (2015) The development of type VI glandular trichomes in the cultivated tomato Solanum lycopersicum and a related wild species S. habrochaites. *BMC Plant Biology*, **15**, 1–15.

Besser, K., Harper, A., Welsby, N., Schauvinhold, I., Slocombe, S., Li, Y. ***et al.*** (2009) Divergent regulation of terpenoid metabolism in the trichomes of wild and cultivated tomato species. *Plant Physiology*, **149**, 499–514.

Brand, A. & **Tissier, A.** (2022) Control of resource allocation between primary and specialized metabolism in glandular trichomes. *Current Opinion in Plant Biology*, **66**, 102172.

Chatterjee, A., Huma, B., Shaw, R. & **Kundu, S.** (2017) Reconstruction of Oryza sativa indica genome scale metabolic model and its responses to varying RuBisCO activity, light intensity, and enzymatic cost conditions. *Frontiers in Plant Science*, **8**, 1–18.

Cheung, C.M., Williams, T.C., Poolman, M.G., Fell, D.A., Ratcliffe, R.G. & **Sweetlove, L.J.** (2013) A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. *The Plant Journal*, **75**, 1050–1061.

Cheung, C.Y.M., Poolman, M.G., Fell, D.A., Ratcliffe, R.G. & **Sweetlove, L.J.** (2014) A diel flux balance model captures interactions between light and dark metabolism during day-night cycles in C3 and Crassulacean acid metabolism leaves. *Plant Physiology*, **165**, 917–929.

Conneely, L.J., Mauleon, R., Mieog, J., Barkla, B.J. & **Kretzschmar, T.** (2021) Characterization of the Cannabis sativa glandular trichome proteome. *PLoS One*, **16**, 1–26.

Escobar-Bravo, R., Ruijgrok, J., Kim, H.K., Grosser, K., Van Dam, N.M., Klinkhamer, P.G.L. ***et al.*** (2018) Light intensity-mediated induction of trichome-associated allelochemicals increases resistance against thrips in tomato. *Plant & Cell Physiology*, **59**, 2462–2475.

Falara, V., Akhtar, T.A., Nguyen, T.T.H., Spyropoulou, E.A., Bleeker, P.M., Schauvinhold, I. ***et al.*** (2011) The tomato terpene synthase gene family. *Plant Physiology*, **157**, 770–789.

Fernandez-Pozo, N., Menda, N., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R. ***et al.*** (2015) The sol genomics network (SGN) from genotype to phenotype to breeding. *Nucleic Acids Research*, **43** (Database issue), D1036–D1041.

Fobes, J.F., Mudd, J.B. & **Marsden, M.P.F.** (1985) Epicuticular lipid accumulation on the leaves of Lycopersicon pennellii (Corr.) D'Arcy and Lycopersicon esculentum mill. *Plant Physiology*, **77**, 567–570. Available from: https://doi.org/10.1104/pp.77.3.567

Gardeström, P. & **Igamberdiev, A.U.** (2016) The origin of cytosolic ATP in photosynthetic cells. *Physiologia Plantarum*, **157**, 367–379.

Glas, J.J., Schimmel, B.C.J., Alba, J.M., Escobar-Bravo, R., Schuurink, R.C. & **Kant, M.R.** (2012) Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores. *International Journal of Molecular Sciences*, **13**, 17077–17103.

Goffman, F.D., Alonso, A.P., Schwender, J., Shachar-Hill, Y. & **Ohlrogge, J.B.** (2005) Light enables a very high efficiency of carbon storage in developing embryos of rapeseed. *Plant Physiology*, **138**, 2269–2279.

Gottstein, W., Olivier, B.G., Bruggeman, F.J. & **Teusink, B.** (2016) Constraint-based stoichiometric modelling from single organisms to microbial communities. *Journal of the Royal Society Interface*, **13**, 20160627.

Heinrich, R. & **Schuster, S.** (1996) The regulation of cellular systems. *Book*, **416**, 138–198.

Huchelmann, A., Boutry, M. & **Hachez, C.** (2017) Plant glandular Trichomes: natural cell factories of high biotechnological interest. *Plant Physiology*, **175**, 6–22.

Kivimäenpä, M., Mofikoya, A., Abd El-Raheem, A.M., Riikonen, J., Julkunen-Tiitto, R. & **Holopainen, J.K.** (2022) Alteration in light spectra causes opposite responses in volatile phenylpropanoids and terpenoids compared with phenolic acids in sweet basil (ocimum basilicum) leaves. *Journal of Agricultural and Food Chemistry*, **70**, 12287–12296.

Koley, S., Grafahrend-Belau, E., Raorane, M.L. & **Junker, B.H.** (2020) The mevalonate pathway contributes to monoterpene production in peppermint. *bioRxiv*.

Kortbeek, R.W.J., Xu, J., Ramirez, A. ***et al.*** (2016) Engineering of tomato glandular Trichomes for the production of specialized metabolites. *Methods in Enzymology*, **576**, 305–331.

Kowalski, R., Kowalska, G., Jankowska, M., Nawrocka, A., Kałwa, K., Pankiewicz, U. ***et al.*** (2019) Secretory structures and essential oil composition of selected industrial species of lamiaceae. *Acta Scientarum Polonorum Hortorum Cultus*, **8**, 2–69.

Lewis, N.E., Hixson, K.K., Conrad, T.M., Lerman, J.A., Charusanti, P., Polpitiya, A.D. ***et al.*** (2010) Omic data from evolved e. coli are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*, **6**, 390.

Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Ataman, M., Babaei, P. ***et al.*** (2020) MEMOTE for standardized genome-scale metabolic model testing. *Nature Biotechnology*, **38**, 272–276.

Maarleveld, T.R., Khandelwal, R.A., Olivier, B.G., Teusink, B. & **Bruggeman, F.J.** (2013) Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnology Journal*, **8**, 997–1008.

Orth, J.D., Thiele, I. & **Palsson, B.Ø.** (2011) What is flux balance analysis? *Nature Biotechnology*, **28**, 245–248.

Paetzolda, H., Garms, S., Bartramb, S. ***et al.*** (2010) The Isogene 1-Deoxy-D-Xylulose 5-phosphate synthase 2 controls isoprenoid profiles, precursor

pathway allocation, and density of tomato Trichomes. *Molecular Plant*, **3**, 904–916.

Poolman, M.G., **Miguet, L.**, **Sweetlove, L.J.** & **Fell, D.A.** (2009) A genome-scale metabolic model of Arabidopsis thaliana and some of its properties. *Plant Physiology*, **151**, 1570–1581.

Prigent, S., **Frioux, C.**, **Dittami, S.M.**, **Thiele, S.**, **Larhlimi, A.**, **Collet, G.** ***et al.*** (2017) Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. *PLoS Computational Biology*, **13**, e1005276.

Rios-Estepa, R., **Lange, I.**, **Lee, J.M.** & **Lange, B.M.** (2010) Mathematical modeling-guided evaluation of biochemical, developmental, environmental, and genotypic determinants of essential oil composition and yield in peppermint leaves. *Plant Physiology*, **152**, 2105–2119.

Rios-Estepa, R., **Turner, G.W.**, **Lee, J.M.**, **Croteau, R.B.** & **Lange, B.M.** (2008) A systems biology approach identifies the biochemical mechanisms regulating monoterpenoid essential oil composition in peppermint. *The Proceedings of the National Academy of Sciences USA*, **105**, 2818–2823.

Ruuska, S.A., **Schwender, J.** & **Ohlrogge, J.B.** (2004) The capacity of green oilseeds to utilize photosynthesis to drive biosynthetic processes. *Plant Physiology*, **136**, 2700–2709.

Saadat, N.P., **Van Aalst, M.** & **Ebenhöh, O.** (2022) Network reconstruction and modelling made reproducible with moped. *Metabolites*, **12**, 275. https://doi.org/10.3390/metabo12040275

Schuurink, R. & **Tissier, A.** (2020) Glandular trichomes: micro-organs with model status? *New Phytologist*, **225**, 2251–2266.

Schwender, J., **Goffman, F.D.**, **Ohlrogge, J.B.** & **Shachar-Hill, Y.** (2004) Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature*, **432**, 779–782.

Seaver, S.M., **Bradbury, L.M.**, **Frelin, O.**, **Zarecki, R.**, **Ruppin, E.**, **Hanson, A.D.** ***et al.*** (2015) Improved evidence-based genome-scale metabolic models for maize leaf, embryo, and endosperm. *Frontiers in Plant Science*, **6**, 142.

Sweetlove, L.J. & **Ratcliffe, R.G.** (2011) Flux-balance modeling of plant metabolism. *Frontiers in Plant Science*, **2**, 38.

Tcherkez, G. & **Atkin, O.K.** (2021) Unravelling mechanisms and impacts of day respiration in plant leaves: an introduction to a virtual issue. *New Phytologist*, **230**, 5–10.

Tomato Genome Consortium. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.

Turner, G.W. & **Croteau, R.** (2004) Organization of monoterpene biosynthesis in mentha. Immunocytochemical localizations of geranyl diphosphate synthase, limonene-6-hydroxylase, isopiperitenol dehydrogenase, and pulegone reductase. *Plant Physiology*, **136**, 4215–4227.

Wahlund, T.M. & **Tabita, F.R.** (1997) 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**, 4859–4867.

Werker, E. (2000) Trichome diversity and development. *Advances in Botanical Research*, **31**, 1–35.

Yuan, H., **Cheung, C.Y.M.**, **Poolman, M.G.**, **Hilbers, P.A.J.** & **Van Riel, N.A.W.** (2016) A genome-scale metabolic network reconstruction of tomato (Solanum lycopersicum L.) and its application to photorespiratory metabolism. *The Plant Journal*, **85**, 289–304.

Zager, J.J. & **Lange, B.M.** (2018) Assessing flux distribution associated with metabolic specialization of glandular Trichomes. *Trends in Plant Science*, **23**, 1–10. Available from: https://doi.org/10.1016/j.tplants.2018.04.003

4 Conclusion

The goal of this thesis was to understand how the central carbon metabolism and energy management of plants work and how they can be optimised for various purposes, including higher yields for crop enhancement and plant-based industrial terpene production. For this, I investigated the Calvin-Benson-Bassham cycle and its tight coupling to the photosynthetic electron transport chain (sections 3.5 and 3.6), the effects of photorespiration and alternative photorespiratory pathways on carbon fixation (sections 3.1 and 3.2) as well as the effects of different bioenergetic cell states on the shift of metabolic routes for the production of secondary metabolites (section 3.7).

We now increase the scope of observation to also include the photosynthetic electron transport chain (section 3.5), alternative electron flows, and the production of ROS (section 3.6). The results of the studies on metabolic control on rubisco clearly show that the notion of a single rate-limiting step in carbon fixation is wrong. Depending on the exact distribution of the enzyme concentration in the PETC and CBB cycle, we find both PETC dominant and CBB dominant control states, as well as mixed states in between. The same holds if single reactions are examined in either the PETC or CBB. For example, the control in the CBB cycle can be mostly on rubisco, sbpase or a mix between them. Since enzyme concentrations between plant species can vary drastically, we strongly advise to look beyond the behaviour of model organisms and cherish the diversity of species and their various survival strategies.

The results of studies on ROS and alternative electron flows highlights the mechanisms that excess energy obtained through photosynthesis can be dissipated and how the activity of the CBB cycle controls alternative electron flows. A particularly important outcome of this study concerns attempts to make carbon fixation or photorespiration more efficient in the sense of energy use. Although this idea sounds intuitively beneficial and should lead to higher yields under low light conditions, care needs to be taken in stressful conditions. If the CBB cycle and photorespiration are significantly more energy efficient, then less energy can be consumed and thus more excess energy from photosynthesis needs to be dissipated. The results of our model show an almost exponential increase in hydrogen peroxide (H_2O_2) concentration under high light conditions, which would even further increase in the case of more energy-efficient carbon fixation.

The idea of contextual switches between limiting carbon fixation rate and energy supply has been further developed in the work on photorespiration (sections 3.1 and 3.2). There we found that at high light and atmospheric CO_2 concentrations, neither the amount of rubisco or any other CBB cycle enzyme, nor the amount of energy supplied, but the transport rate of carbon dioxide into the chloroplast was the limiting factor for additional carbon fixation. This had profound implications

for the design and performance of different alternative photorespiratory pathways under different conditions. The alternative pathways that were studied can be broadly categorised by either fixing an additional carbon dioxide, instead of loosing it by the activity of glycine decarboxylase in the mitochondrium, or relocating the carbon dioxide release into the chloroplast and then doing either a single or two decarboxylation steps. The model results agree with experimental studies that the relocation of the carbon dioxide release is beneficial, as not all of the carbon dioxide release by glycine decarboxylase will end up in the chloroplast. Therefore, this relocation of the release locally increases the carbon dioxide concentration, which at the same time reduces the rubisco oxygenase activity. At their core, these pathways thus use extra energy to inhibit rubisco oxygenase as much as possible, which is still a net positive effect. A much more promising approach is the fixation of an additional carbon during photorespiration, as in the TaCo pathway variant. This again highlights that given sufficient energy supply, the main issue plants face is the transport of carbon dioxide into the chloroplast. This result is not unsurprising, as plants have evolved stomata precisely for the reason of increased gas exchange, which comes at the cost of water loss. Further evidence for this comes from the development of C4 and CAM photosynthesis, which use spatial and temporal mechanisms for carbon prefixation and concentration respectively, in order to adapt to arid conditions.

Lastly, we again increase the scope of observation and include secondary metabolism as well (section 3.7). In this work on terpene metabolism in photosynthetic glandular trichomes we could show that even in cells that have a continuous and sufficient supply of sucrose, production of terpenoids and other secondary metabolites can be increased by the presence of both the PETC and the CBB cycle, even though there is only very little gas exchange between the photosynthetic glandular trichome and the surrounding air. The main mechanisms we could identify were light-driven carbon partitioning and refixation of CO_2 released by catabolic processes. The light-driven carbon partitioning clearly showed that the presence of the PETC allowed the photosynthetic glandular trichome to use energy supplied by the light-reactions instead of relying on catabolising sucrose for energy, providing direct applicability to bioproduction of terpenoids. This energy-driven shift from catabolic to anabolic reactions again shows the interconnectedness of plant metabolism.

One recurring result of this work was that plant metabolism is tightly interlinked and minor changes to one part of it can have far-reaching effects on other parts of it. This highlights the necessity to avoid overly reductionist approaches to understanding plant metabolism and instead to try to systematically collate the knowledge gained over time, in order to build a complete picture. On particular helpful tool in this regard are mathematical models, as they not only allow to distil and combine knowledge from different studies and experiments, but do require rigour in doing so. This is in stark contrast to hand-waving arguments that are still way to common in biological research. But even the state of mathematical modelling in biology currently leaves a lot to be desired. Too often pathways are looked at in isolation and their interaction with the remaining metabolism is either ignored or assumed to be of a simple linear nature. While convenient, it has become clear in this thesis that more often than not the remaining metabolism in fact shows non-linear response to changes in the pathway of interest.

This is often due to the inherit challenges of the modelling approach chosen. While FBA-type models allow for a holistic overview of metabolism, they completely fail to account for non-linear kinetic behaviour or state-dependent constraints. On the other hand, ODE-type models allow for a very fine-grained description of single pathways, but are increasingly difficult to work with at scale. While trying to understand how different photorespiratory bypasses affect carbon fixation, we thus employed both these techniques to get a better overall picture, see section 3.2. This allowed us to first use the stoichiometric genome-scale model to assess which parts of the plant metabolism will be affected by the introduction of a photorespiratory bypass and then to use the kinetic model to provide a detailed picture of the affected processes. This combined approach yielded much better results than either approach would have had in isolation, as the iterative nature of comparing the respective results did allow us to continuously update and improve both our models. However, we found that there were certain differences in the model predictions which were fundamental to the modelling approaches. In particular, the FBA model predicted that the carbon-fixing bypasses would be more efficient at low-light, or energy-limited, conditions, while the ODE model predicted that they would be more efficient at high-light conditions. This difference could ultimately be explained, as in the ODE model at high-light conditions the carbon influx would become limiting. Thus the carbon-fixing aspect of the bypasses was predicted as the major advantage. In contrast, in the FBA model at low-light the energy-efficiency of the bypasses was more important. The ODE model on the other hand predicted that in low-light conditions the plastidic carbon-dioxide pool was increased, which cause less photorespiration overall. Those non-linear effects can arbitrarily be added to FBA models, but they are never their intrinsic results. On the other hand, the stoichiometric model clearly showed energetic differences of different compartmentalisation related to energy cost of metabolite transport between those compartments, which weren't part of the ODE model. It is thus always important to model any given biological system from multiple perspectives, as otherwise it is too easy not to do justice to the inherit complexity of it.

To conclude, this thesis provides several models and approaches to understanding primary and secondary plant metabolism. I used these models to answer questions about both fundamental research as well as industrial applicability. This work will hopefully guide further research, crop development and above all - be useful.

Bibliography

[1] Achim V Trebst, Harry Y Tsujimoto, and Daniel I Arnon. “Separation of light and dark phases in the photosynthesis of isolated chloroplasts”. In: *Nature* 182.4632 (1958), pp. 351–355.

[2] Bob B Buchanan. “The carbon (formerly dark) reactions of photosynthesis”. In: *Photosynthesis research* 128 (2016), pp. 215–217.

[3] Robert E Blankenship. *Molecular mechanisms of photosynthesis*. John Wiley & Sons, 2021.

[4] Bernard E Pfeil, Benoıt Schoefs, and Cornelia Spetea. “Function and evolution of channels and transporters in photosynthetic membranes”. In: *Cellular and Molecular Life Sciences* 71 (2014), pp. 979–998.

[5] Robert R Wise and J Kenneth Hoober. *The structure and function of plastids*. Vol. 23. Springer Science & Business Media, 2007.

[6] Nima P Saadat et al. “Computational analysis of alternative photosynthetic electron flows linked with oxidative stress”. In: *Frontiers in plant science* 12 (2021), p. 750580.

[7] John Whitmarsh and Govindjee. “The photosynthetic process”. In: *Concepts in photobiology: photosynthesis and photomorphogenesis*. Springer, 1999, pp. 11–51.

[8] Daniel I Arnon. “The light reactions of photosynthesis”. In: *Proceedings of the National Academy of Sciences* 68.11 (1971), pp. 2883–2892.

[9] Alexander N Tikhonov. “The cytochrome b 6 f complex: biophysical aspects of its functioning in chloroplasts”. In: *Membrane protein complexes: Structure and function* (2018), pp. 287–328.

[10] Laura S Green et al. “Ferredoxin and ferredoxin-NADP reductase from photosynthetic and nonphotosynthetic tissues of tomato”. In: *Plant physiology* 96.4 (1991), pp. 1207–1213.

[11] Yuri Munekage et al. “Cyclic electron flow around photosystem I is essential for photosynthesis”. In: *Nature* 429.6991 (2004), pp. 579–582.

[12] Alexander Hahn et al. “Structure, mechanism, and regulation of the chloroplast ATP synthase”. In: *Science* 360.6389 (2018), eaat4318.

[13] Zhenhui Kang, Tong Qin, and Zhiping Zhao. “Thioredoxins and thioredoxin reductase in chloroplasts: A review”. In: *Gene* 706 (2019), pp. 32–42.

[14] Bob B Buchanan. “The ferredoxin/thioredoxin system: a key element in the regulatory function of light in photosynthesis”. In: *Bioscience* 34.6 (1984), pp. 378–383.

[15] Peter Geigenberger et al. “The unprecedented versatility of the plant thioredoxin system”. In: *Trends in plant science* 22.3 (2017), pp. 249–262.

[16] 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”. In: *Plant and Cell physiology* 43.9 (2002), pp. 1017–1026.

[17] Sheikh Mansoor et al. “Reactive oxygen species in plants: from source to sink”. In: *Antioxidants* 11.2 (2022), p. 225.

[18] Kozi Asada. “The water–water cycle as alternative photon and electron sinks”. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 355.1402 (2000), pp. 1419–1431.

[19] Wojciech J Nawrocki et al. “The plastid terminal oxidase: its elusive function points to multiple contributions to plastid physiology”. In: *Annual review of plant biology* 66 (2015), pp. 49–74.

[20] Edelmira Valero et al. “Modeling the ascorbate-glutathione cycle in chloroplasts under light/dark conditions”. In: *BMC systems biology* 10 (2015), pp. 1–20.

[21] E-A MOHAMED et al. “Overexpression of bacterial catalase in tomato leaf chloroplasts enhances photo-oxidative stress tolerance”. In: *Plant, Cell & Environment* 26.12 (2003), pp. 2037–2046.

[22] Mónika Lingvay et al. “Photobleaching of chlorophyll in light-harvesting complex II increases in lipid environment”. In: *Frontiers in plant science* 11 (2020), p. 849.

[23] Anna Matuszyńska et al. “A mathematical model of non-photochemical quenching to study short-term light memory in plants”. In: *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1857.12 (2016), pp. 1860–1869.

[24] Petar H Lambrev et al. “Kinetic and spectral resolution of multiple nonphotochemical quenching components in Arabidopsis leaves”. In: *Plant physiology* 152.3 (2010), pp. 1611–1624.

[25] Manuela Nilkens et al. “Identification of a slowly inducible zeaxanthin-dependent component of non-photochemical quenching of chlorophyll fluorescence generated under steady-state conditions in Arabidopsis”. In: *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1797.4 (2010), pp. 466–475.

[26] Danielle A Way and Robert W Pearcy. “Sunflecks in trees and forests: from photosynthetic physiology to global change biology”. In: *Tree Physiology* 32.9 (2012), pp. 1066–1081.

[27] John F Allen et al. “Chloroplast protein phosphorylation couples plastoquinone redox state to distribution of excitation energy between photosystems”. In: *Nature* 291.5810 (1981), pp. 25–29.

[28] Xia Zhao et al. “Non-photochemical quenching plays a key role in light acclimation of rice plants differing in leaf color”. In: *Frontiers in plant science* 7 (2017), p. 1968.

[29] Stephan Eberhard, Giovanni Finazzi, and Francis-André Wollman. “The dynamics of photosynthesis”. In: *Annual review of genetics* 42 (2008), pp. 463–515.

[30] Peter Jahns and Alfred R Holzwarth. “The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II”. In: *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1817.1 (2012), pp. 182–193.

[31] Paul Falkowski et al. “The global carbon cycle: a test of our knowledge of earth as a system”. In: *science* 290.5490 (2000), pp. 291–296.

[32] Peter M Cox et al. “Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model”. In: *Nature* 408.6809 (2000), pp. 184–187.

[33] John A Raven. “Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments”. In: *Aquatic Microbial Ecology* 56.2-3 (2009), pp. 177–192.

[34] Yinon M Bar-On and Ron Milo. “The global mass and average rate of rubisco”. In: *Proceedings of the National Academy of Sciences* 116.10 (2019), pp. 4738–4743.

[35] Melvin Calvin. “Chemical and photochemical reactions of thioctic acid and related disulfides”. In: (1954).

[36] William L Ogren and George Bowes. “Ribulose diphosphate carboxylase regulates soybean photorespiration”. In: *Nature New Biology* 230.13 (1971), pp. 159–160.

[37] Archie R Portis Jr. “Regulation of ribulose 1, 5-bisphosphate carboxylase/oxygenase activity”. In: *Annual review of plant biology* 43.1 (1992), pp. 415–437.

[38] Archie R Portis Jr. “The regulation of Rubisco by Rubisco activase”. In: *Journal of Experimental Botany* 46.special_issue (1995), pp. 1285–1291.

[39] Archie R Portis Jr, RM Lilley, and T John Andrews. “Subsaturating ribulose-1, 5-bisphosphate concentration promotes inactivation of ribulose-1, 5-bisphosphate carboxylase/oxygenase (rubisco)(studies using continuous substrate addition in the presence and absence of rubisco activase)”. In: *Plant Physiology* 109.4 (1995), pp. 1441–1451.

[40] Hadi Farazdaghi. “The single-process biochemical reaction of Rubisco: A unified theory and model with the effects of irradiance, CO2 and rate-limiting step on the kinetics of C3 and C4 photosynthesis from gas exchange”. In: *Biosystems* 103.2 (2011), pp. 265–284.

[41] Gosta Pettersson and Ulf Ryde-Pettersson. “A mathematical model of the Calvin photosynthesis cycle”. In: *European Journal of Biochemistry* 175.3 (1988), pp. 661–672. ISSN: 14321033. DOI: 10.1111/j.1432-1033.1988.tb14242.x.

[42] Moritz E Beber et al. “eQuilibrator 3.0: a database solution for thermodynamic constant estimation”. In: *Nucleic acids research* 50.D1 (2022), pp. D603–d609.

[43] Anne Arnold and Zoran Nikoloski. “A quantitative comparison of Calvin–Benson cycle models”. In: *Trends in plant science* 16.12 (2011), pp. 676–683.

[44] Graham D Farquhar, S von von Caemmerer, and Joseph A Berry. “A biochemical model of photosynthetic CO 2 assimilation in leaves of C 3 species”. In: *planta* 149 (1980), pp. 78–90.

[45] BE Medlyn et al. “Temperature response of parameters of a biochemically based model of photosynthesis. II. A review of experimental data”. In: *Plant, Cell & Environment* 25.9 (2002), pp. 1167–1179.

[46] Hans R Schultz. “Extension of a Farquhar model for limitations of leaf photosynthesis induced by light environment, phenology and leaf age in grapevines (Vitis vinifera L. cvv. White Riesling and Zinfandel)”. In: *Functional Plant Biology* 30.6 (2003), pp. 673–687.

[47] Thomas D Sharkey et al. “Fitting photosynthetic carbon dioxide response curves for C3 leaves”. In: *Plant, cell & environment* 30.9 (2007), pp. 1035–1040.

[48] Leonid E Fridlyand and Renate Scheibe. “Regulation of the Calvin cycle for CO2 fixation as an example for general control mechanisms in metabolic cycles”. In: *Biosystems* 51.2 (1999), pp. 79–93.

[49] 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”. In: *Nonlinear Analysis: Real World Applications* 10.3 (2009), pp. 1490–1499.

[50] Mark G. Poolman, David A. Fell, and Simon Thomas. “Modelling photosynthesis and its control”. In: *Journal of Experimental Botany* 51.suppl_1 (Feb. 2000), pp. 319–328. ISSN: 1460-2431. DOI: 10.1093/jexbot/51.suppl_1.319. URL: https://academic.oup.com/jxb/article-lookup/doi/10.1093/jexbot/51.suppl%7B%5C_%7D1.319.

[51] Christoph Giersch, Dirk Lämmel, and Graham Farquhar. “Control analysis of photosynthetic CO 2 fixation”. In: *Photosynthesis research* 24 (1990), pp. 151–165.

[52] Gaëlle Damour and Laurent Urban. “Models of photosynthesis need to and can be upgraded to include the effects of drought, phenological changes, sink activity and carbohydrate accumulation on the light exposure/photosynthetic capacity relationship:[Communication orale]”. In: 2007.

[53] Brian D Hahn. “A mathematical model of the Calvin cycle: analysis of the steady state”. In: *Annals of botany* 57.5 (1986), pp. 639–653.

[54] Agu Laisk, Hillar Eichelmann, and Vello Oja. “C 3 photosynthesis in silico”. In: *Photosynthesis research* 90 (2006), pp. 45–66.

[55] 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”. In: *Plant physiology* 145.2 (2007), pp. 513–526.

[56] Ian E Woodrow and Keith A Mott. “Modelling C3 photosynthesis: a sensitivity analysis of the photosynthetic carbon-reduction cycle”. In: *Planta* 191.4 (1993), pp. 421–432.

[57] A Laisk et al. “A mathematical model of the carbon metabolism in photosynthesis. Difficulties in explaining oscillations by fructose 2, 6-bisphosphate regulation”. In: *Proceedings of the Royal Society of London. B. Biological Sciences* 237.1289 (1989), pp. 389–415.

[58] Anna Matuszyńska, Nima P. Saadat, and Oliver Ebenhöh. “Balancing energy supply during photosynthesis – a theoretical perspective”. In: *Physiologia Plantarum* 166.1 (May 2019), pp. 392–402. ISSN: 13993054. DOI: 10.1111/ppl.12962.

[59] Oliver Ebenhöh et al. “Short-term acclimation of the photosynthetic electron transfer chain to changing light: a mathematical model”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1640 (2014), p. 20130223.

[60] Sarah Courbier. “Breaking Dawn: The twilight of starch degradation in the light”. In: *Plant Physiology* 189.4 (May 2022), pp. 1887–1889. ISSN: 0032-0889. DOI: 10.1093/plphys/kiac202. eprint: https://academic.oup.com/plphys/article-pdf/189/4/1887/45228177/kiac202.pdf. URL: https://doi.org/10.1093/plphys/kiac202.

[61] Hirofumi Ishihara et al. “Rising rates of starch degradation during daytime and trehalose 6-phosphate optimize carbon availability”. In: *Plant Physiology* 189.4 (2022), pp. 1976–2000.

[62] Alexandra Pokhilko et al. “Adjustment of carbon fluxes to light conditions regulates the daily turnover of starch in plants: a computational model”. In: *Molecular BioSystems* 10.3 (2014), pp. 613–627.

[63] Alan M McClain and Thomas D Sharkey. “Triose phosphate utilization and beyond: from photosynthesis to end product synthesis”. In: *Journal of experimental botany* 70.6 (2019), pp. 1755–1766.

[64] Stefan Timm et al. “Glycine decarboxylase controls photosynthesis and plant growth”. In: *FEBS letters* 586.20 (2012), pp. 3692–3697.

[65] Stefan Timm et al. “Mitochondrial dihydrolipoyl dehydrogenase activity shapes photosynthesis and photorespiration of Arabidopsis thaliana”. In: *The Plant Cell* 27.7 (2015), pp. 1968–1984.

[66] Stefan Timm et al. “A cytosolic pathway for the conversion of hydroxypyruvate to glycerate during photorespiration in Arabidopsis”. In: *The Plant Cell* 20.10 (2008), pp. 2848–2859.

[67] Louise E Anderson. “Chloroplast and cytoplasmic enzymes II. Pea leaf triose phosphate isomerases”. In: *Biochimica et Biophysica Acta (BBA)-Enzymology* 235.1 (1971), pp. 237–244.

[68] Franziska Flügel et al. “The photorespiratory metabolite 2-phosphoglycolate regulates photosynthesis and starch accumulation in Arabidopsis”. In: *The Plant Cell* 29.10 (2017), pp. 2537–2551.

[69] Patricia E López-Calcagno et al. “Overexpressing the H-protein of the glycine cleavage system increases biomass yield in glasshouse and field-grown transgenic tobacco plants”. In: *Plant Biotechnology Journal* 17.1 (2019), pp. 141–151.

[70] Chris J Chastain and William L Ogren. “Glyoxylate inhibition of ribulosebisphosphate carboxylase/oxygenase activation state in vivo”. In: *Plant and Cell Physiology* 30.7 (1989), pp. 937–944.

[71] William J Campbell and William L Ogren. “Glyoxylate inhibition of ribulosebisphosphate carboxylase/oxygenase activation in intact, lysed, and reconstituted chloroplasts”. In: *Photosynthesis Research* 23 (1990), pp. 257–268.

[72] Rainer E Häusler et al. “Control of photosynthesis in barley mutants with reduced activities of glutamine synthetase and glutamate synthase: III. Aspects of glyoxylate metabolism and effects of glyoxylate on the activation state of ribulose-1, 5-bisphosphate carboxylase-oxygenase”. In: *Planta* 200.4 (1996), pp. 388–396.

[73] Berkley J Walker et al. “The costs of photorespiration to food production now and in the future”. In: *Annual review of plant biology* 67 (2016), pp. 107–129.

[74] Yonatan Savir et al. “Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape”. In: *Proceedings of the National Academy of Sciences* 107.8 (2010), pp. 3475–3480.

[75] Camille Bathellier et al. “Rubisco is not really so bad”. In: *Plant, cell & environment* 41.4 (2018), pp. 705–716.

[76] Avi I Flamholz et al. “Revisiting trade-offs between Rubisco kinetic parameters”. In: *Biochemistry* 58.31 (2019), pp. 3365–3376.

[77] CB Osmond. “Crassulacean acid metabolism: a curiosity in context”. In: *Annual review of plant physiology* 29.1 (1978), pp. 379–414.

[78] John C Cushman. “Crassulacean acid metabolism. A plastic photosynthetic adaptation to arid environments”. In: *Plant physiology* 127.4 (2001), pp. 1439–1448.

[79] Rowan F Sage. “The evolution of C4 photosynthesis”. In: *New phytologist* 161.2 (2004), pp. 341–370.

[80] A Arnold and Z Nikoloski. “Comprehensive classification and perspective for modelling photorespiratory metabolism”. In: *Plant Biology* 15.4 (2013), pp. 667–675.

[81] Brian D Hahn. “A mathematical model of leaf carbon metabolism”. In: *Annals of Botany* 54.3 (1984), pp. 325–339.

[82] Brian D Hahn. “A mathematical model of photorespiration and photosynthesis”. In: *Annals of Botany* 60.2 (1987), pp. 157–169.

[83] Akiho Yokota, Hiroshi Komura, and Shozaburo Kitaoka. “Refixation of photorespired CO2 during photosynthesis in Euglena gracilis Z”. In: *Agricultural and biological chemistry* 49.11 (1985), pp. 3309–3310.

[84] Honglong Zhao, Yi Xiao, and Xin-Guang Zhu. “Kinetic modeling of photorespiration”. In: *Photorespiration: Methods and Protocols* (2017), pp. 203–216.

[85] Chandra Bellasio. “A generalised dynamic model of leaf-level C3 photosynthesis combining light and dark reactions with stomatal behaviour”. In: *Photosynthesis research* 141.1 (2019), pp. 99–118.

[86] John PA Ioannidis. “Why most published research findings are false”. In: *PLoS medicine* 2.8 (2005), e124.

[87] Carmen M Reinhart and Kenneth S Rogoff. “Growth in a Time of Debt”. In: *American economic review* 100.2 (2010), pp. 573–578.

[88] Thomas Herndon, Michael Ash, and Robert Pollin. “Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff”. In: *Cambridge journal of economics* 38.2 (2014), pp. 257–279.

[89] Marc N Branch. “The “reproducibility crisis:” Might the methods used frequently in behavior-analysis research help?” In: *Perspectives on Behavior Science* 42.1 (2019), pp. 77–89.

[90] Ron Caspi et al. “The MetaCyc database of metabolic pathways and enzymes-a 2019 update”. In: *Nucleic acids research* 48.D1 (2020), pp. D445–d453.

[91] Ali Ebrahim et al. “COBRApy: constraints-based reconstruction and analysis for python”. In: *BMC systems biology* 7 (2013), pp. 1–6.