Evolution of C4 photosynthesis predicted by constraint-based modelling

Mary-Ann Blätke[1*] **and Andrea Bräutigam**[1,2]

***For correspondence:**
blaetke@ipk-gatersleben.de (MB)

[1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstraße 3, 06466 Stadt Seeland OT Gatersleben, Germany); [2]Computational Biology, Faculty of Biology, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany

Abstract

Evolutionary constraint-based modelling is a powerful tool for the analysis of evolutionary trajectories in bacteria with unknown usefulness in eukaryotes. We hypothesised that the agriculturally important trait of C4 photosynthesis is accessible to modelling because it evolved from an ancestral ground state of C3 photosynthesis. All genes required for C4 are present in C3 but not arranged in a pathway. Plants performing C4 metabolism are capable of highly efficient carbon fixation and consequentially are adapted to adverse growth conditions and/or are capable of high growth rates. We hypothesised that constraint-based modelling will identify the selective pressures under which C4 photosynthesis evolves. Indeed, a carefully curated two-celled model predicts C4 under two different selective pressures. The model attributed differences between extant metabolic flavours of C4 to light availability and distribution. Likely kinetic constraints on reactions were identified for flux solutions which did not match known physiological fluxes. The data shows evolutionary constraint-based modelling in eukaryotes predicts molecular evolution with precision.

Introduction

Identifying specific evolutionary trajectories and modelling the outcome of adaptive strategies at the molecular levels is a major challenge in evolutionary systems biology ***Papp et al.*** (***2011***). The evolution of novel metabolic pathways from existing parts may be predicted using constraint-based modelling ***Orth et al.*** (***2010***). In constraint-based models, selective pressures are coded via the objective functions for which the model is optimised. The factors which constrain evolution are integrated into the models via changes in model inputs or outputs and via flux constraints. We hypothesised that the evolution of C4 photosynthesis is accessible to constraint-based modelling. Because all enzymes and transporter involved are present in the ancestral state ***Aubry et al.*** (***2011***). C4 photosynthesis ***Hatch*** (***1987***) is a carbon concentration mechanism (CCM) in plants which allows colonisation of marginal habitats ***Sage et al.*** (***2012***) and high biomass production in annuals such as crops ***Sage*** (***2004***); ***Edwards et al.*** (***2010***). A model build for the ancestral phenotype should predict the novel phenotype if given the correct selective pressures coded as the optimisation functions.

C4 photosynthesis evolved independently in at least 67 independent origins in the plant kingdom ***Scheben et al.*** (***2017***). It is frequently assumed that evolution of C4 photosynthesis requires water limitation (i.e. ***Bräutigam and Gowik*** (***2016***); ***Heckmann et al.*** (***2013***); ***Mallmann et al.*** (***2014***)). However, ecophysiological research showed that C4 can likely evolve in wet habitats ***Osborne***

and Freckleton (***2009***); ***Lundgren and Christin*** (***2016***). Raising the question under which particular selective pressure or pressures the trait is selected for. Ultimately, C4 photosynthesis acts as a biochemical pump which enriches the CO_2 concentration at the site of Rubisco to overcome a major limitation of carbon fixation ***Sage*** (***2004***). Enrichment is beneficial because Rubisco, the carbon fixation enzyme, can react productively with CO_2 and form two molecules of 3-PGA but it also reacts with O_2 and produces 2-phosphoglycolate which requires detoxification by photorespiration ***Ogren and Bowes*** (***1971***). The ratio between both reactions is determined by the enzyme specificity towards CO_2, by the temperature, and the concentrations of both reactants which in turn is modulated by stresses such as drought and pathogen load. Evolution of Rubisco itself is constrained since any increase in specificity is paid for by a reduction in speed ***Spreitzer and Salvucci*** (***2002***). Lower speeds are prohibitive since Rubisco is a comparatively slow enzyme and can comprise up to 50 % of the total leaf protein ***Ellis*** (***1979***). In the C4 cycle, phosphoenolpyruvate carboxylase affixes CO_2 to a C3 acid, phosphoenolpyruvate (PEP), forming a C4 acid, oxaloacetate (OAA). After stabilisation of the resulting C4 acid by reduction to malate or transamination to aspartate, it is transferred to the site of Rubisco and decarboxylated by one of three possible decarboxylation enzymes, NADP-dependent malic enzyme (NADP-ME), NAD-dependent malic enzyme (NAD-ME), or PEP carboxykinase (PEP-CK) ***Hatch*** (***1987***); ***Schlüter et al.*** (***2016***). Species such as corn (*Zea mays*) ***Pick et al.*** (***2011***) and great millet (*Sorghum bicolor*) ***Döring et al.*** (***2016***) use NADP-ME, species like common millet (*Panicum miliaceum*) ***Hatch*** (***1987***) and African spinach (*Gynandropsis gynandra*) ***Feodorova et al.*** (***2010***); ***Voznesenskaya et al.*** (***2007***) use NAD-ME and species such as guinea grass (*Panicum maximum*) ***Bräutigam et al.*** (***2014***) use mainly PEP-CK with the evolutionary constraints leading to one or the other enzyme unknown. Mixed forms are only known to occur between a malic enzyme and PEP-CK but not between both malic enzymes ***Wang et al.*** (***2014***). After decarboxylation, the C3 acid diffuses back to the site of phosphoenolpyruvate carboxylase (PEPC) and is recycled for another C4 cycle by pyruvate phosphate dikinase (PPDK) ***Hatch*** (***1987***); ***Schlüter et al.*** (***2016***). In its most typical form, this C4 cycle is distributed between different cell types in a leaf in an arrangement called Kranz anatomy ***Haberlandt*** (***1904***). Initial carbon fixation by PEPC occurs in the mesophyll cell, the outer layer of photosynthetic tissue. The secondary fixation by Rubisco after decarboxylation occurs in an inner layer of photosynthetic tissue, the bundle sheath which in turn surrounds the veins. Both cells are connected by plasmodesmata which are pores with limited transfer specificity between cells. A model which may test possible carbon fixation pathways at the molecular level thus requires two cell architectures connected by transport processes ***Bräutigam and Weber*** (***2010***).

Stoichiometric constraint-based models of genome-scale or close to it are well suited to study evolution (summarised in ***Papp et al.*** (***2011***)). Evolution of different metabolic modes from a ground state, the metabolism of *Escherichia coli*, such as glycerol usage ***Lewis et al.*** (***2010***) or endosymbiotic metabolism ***Pál et al.*** (***2006***) have been successfully predicted. Metabolic maps of plant metabolism are of higher complexity compared to bacteria since they require information about intracellular compartmentation and intracellular transport and may require multicellular approaches. In plants, aspects of complex metabolic pathways, such as the energetics of CAM photosynthesis ***Cheung et al.*** (***2014***), and fluxes in C3 and C4 metabolism ***Boyle and Morgan*** (***2009***); ***de Oliveira Dal'Molin et al.*** (***2011***); ***de Oliveira Dal'Molin et al.*** (***2010a***); ***Arnold and Nikoloski*** (***2014***); ***Saha et al.*** (***2011***) have been elucidated with genome scale models. The C4 cycle is not an emergent property of these current C4 models as the cycle is forced by constraints ***de Oliveira Dal'Molin et al.*** (***2011***); ***Mallmann et al.*** (***2014***). In the C4GEM model, the fluxes representing the C4 cycle are *a priori* constrained to the cell types ***de Oliveira Dal'Molin et al.*** (***2011***), and in the Mallmann model, the C4 fluxes are induced by activating flux through PEPC ***Mallmann et al.*** (***2014***). Models in which specific *a priori* constraints activated C4 were successfully used to study metabolism under conditions of photosynthesis, photorespiration, and respiration ***Saha et al.*** (***2011***) and to study N-assimilation under varying conditions ***Simons et al.*** (***2013***). However, they are incapable of testing under which conditions the pathway may emerge.

Schematic models suggest that the C4 cycle evolves from its ancestral metabolic state C3 photosynthesis along a sequence of stages (summarised in ***Sage*** (***2004***); ***Bräutigam and Gowik*** (***2016***)). In the presence of tight vein spacing and of photosynthetically active bundle sheath cells (i.e. Kranz anatomy), a key intermediate in which the process of photorespiration is divided between cell types is thought to evolve ***Monson*** (***1999***); ***Sage et al.*** (***2012***); ***Heckmann et al.*** (***2013***); ***Bauwe*** (***2010***). The metabolic fluxes in this intermediate suggest an immediate path towards C4 photosynthesis ***Mallmann et al.*** (***2014***); ***Bräutigam and Gowik*** (***2016***). ***Heckmann et al.*** (***2013***) built a kinetic model in which the complex C4 cycle was represented by a single enzyme, PEPC. Assuming carbon assimilation as a proxy for fitness, the model showed that the evolution from a C3 progenitor species with Kranz-type anatomy towards C4 photosynthesis occurs in modular, individually adaptive steps on a Mount Fuji fitness landscape. Neither of the current models matches the molecular complexity of the natural systems and rely on artful complexity reduction to enable the modelling process. Hence current models fail to capture the molecular variations observed in different species which represent some of the over sixty independent origins ***Sage*** (***2004***)

In this study, we establish a generic two-celled, constraint-based model starting from ***Arnold and Nikoloski*** (***2014***). C4 photosynthesis emerges as the optimal solution under two different evolutionary objectives, which are resource limitation and minimised photorespiration. We test which input and metabolic constraints favour the emergence of particular C4 metabolic pathways at the molecular level. In the process, we demonstrate that evolution is predictable at the molecular level in a eukaryotic system and define the selective pressures and limitations guiding the "choice" of metabolic flux.

Results

Curation of *Arabidopsis* core model

Flux balance analysis requires five types of information, the metabolic map of the organism, the input, the output, a set of constraints, i.e. limitations on input, directionality of reactions, forced flux through reactions, and optimisation criteria for the algorithm which approximate the selective pressures the metabolism evolved under. For reconstruction of the C3 metabolic map we curated the *Arabidopsis* core model ***Arnold and Nikoloski*** (***2014***) manually (***Table 1***) to represent the metabolism of a mesophyll cell in a mature photosynthetically active leaf of a C3 plant , further on called *one-cell* model (provided in ***Figure 1–source data 1***).

For the inputs, we considered a photoautotrophic growth scenario with a fixed CO_2 uptake of about 20 µmol/(m^2 s) ***Lacher*** (***2003***). Light, sulphates, and phosphate are freely available. Due to the observation that nitrate is the main source (80 %) of nitrogen in leaves in many species ***Macduff and Bakken*** (***2003***), we set nitrate as the sole nitrogen source. If both ammonia and nitrate are allowed, the model will inevitably predict the physiologically incorrect sole use of ammonia since fewer reactions and less energy are required to convert it into glutamate, the universal amino group currency in plants. Water and oxygen can be freely exchanged with the environment in both directions.

To compute the output, we assume a mature fully differentiated and photosynthetically active leaf, which is optimised for the synthesis and export of sucrose and amino acids to the phloem under minimal metabolic effort. Following the examples of models in bacteria, many plant models use a biomass function which assumes that the leaf is required to build itself ***de Oliveira Dal'Molin et al.*** (***2010a***); ***Arnold and Nikoloski*** (***2014***); ***Saha et al.*** (***2011***) using photoautotrophic i.e. ***Arnold and Nikoloski*** (***2014***) or heterotrophic i.e. ***Cheung et al.*** (***2014***) energy and molecule supply. In plants, however, leaves transition from a sink phase in which they build themselves from metabolites delivered by the phloem to a source phase in which they produce metabolites for other organs including sink leaves ***Turgeon*** (***1989***). The composition of *Arabidopsis* phloem exudate ***Wilkinson and Douglas*** (***2003***) was used to constrain the relative proportions of the 18 amino acids and the

ratio of sucrose : total amino acids ($2.2:1$). To account for daily carbon storage as starch for export during the night, we assume that half of the assimilated carbon is stored in the *one-cell* model. We explicitly account for maintenance costs by the use of a generic ATPase and use the measured ATP costs for protein degradation and synthesis of a mature *Arabidopsis* leaf ***Li et al.*** (***2017***) as a constraint. We initially assume a low photorespiratory flux according to the ambient CO_2 and O_2 partial pressures considering no heat, drought, salt or osmotic stress which may alter the ratio towards higher flux towards the oxygenation reaction.

To develop a largely unconstrained model and detect possible errors in the metabolic map, we initially kept the model unconstrained with regard to fixed fluxes, flux ratios, and reaction directions. Different model iterations were run in (re-)design, simulate, validate cycles against known physiology with errors sequentially eliminated and a minimal set of constraints required for a C3 model recapitulating extant plant metabolism determined.

The initial FBA resulted in carbon fixation by enzymes such as the malic enzymes which, in reality, are constrained by the kinetics of the enzymes towards decarboxylation. All decarboxylation reactions were made unidirectional towards decarboxylation to prevent erroneous carbon fixation in the flux distribution. The next iteration of FBA predicted loops through nitrate reductases which ultimately converted NADH to NADPH. We traced this loop to an error in the initial model, in which malate dehydrogenases in the cytosol and mitochondrion were NADP-dependent instead of NAD-dependent. After correction of the co-factor in the *one-cell* model, the loops through nitrate reductases were no longer observed. Another iteration predicted excessive flux through the mitochondrial membrane where multiple metabolites were exchanged and identified missing transport processes as the likely reason. Based on ***Linka and Weber*** (***2010***), we added known fluxes across the mitochondrial and plastidic envelope membranes which remedied the excessive fluxes in the solution. The chloroplastic ADP/ATP carrier protein is constrained to zero flux since its mutant is only affected during the night but not if light is available ***Reiser*** (***2004***).

The obtained flux distribution contained excessive fluxes through multiple transport proteins across internal membranes which ultimately transferred protons between the organelles and the cytosol. Since for most if not all transport proteins the precise protonation state of metabolites during transport is unknown and hence cannot be correctly integrated into the model, we allowed protons to appear and disappear as needed in all compartments. This provision precludes any and all conclusions about the energetics of membrane transport. ATP generation occurred in a distorted way distributed across different organelles which were traced to the H^+ consumption of the ATPases in mitochondria and chloroplasts. The stoichiometry was altered to to $3:1$ (chloroplast) and $4:1$ (mitochondria) ***Petersen et al.*** (***2012***); ***Turina et al.*** (***2016***). We assume no flux for the chloroplastic NADPH dehydrogenase and plastoquinol oxidase because ***Josse et al.*** (***2000***); ***Yamamoto et al.*** (***2011***) have shown that their effect on the photosynthesis is considerably low.

In preparation for modelling the C4 cycle, we ensured that all reactions known to occur in C4 (i.e. malate/pyruvate exchange, likely via DiT2 in maize ***Weissmann et al.*** (***2016***), possibly promiscuous amino transferases ***Duff et al.*** (***2012***)) are present in the *one-cell* model, since ***Aubry et al.*** (***2011***) showed that all genes encoding enzymes and transporters underlying the C4 metabolism are already present in the genome of C3 plants. We integrated cyclic electron flow ***Shikanai*** (***2016***) and alternative oxidases in the mitochondria ***Vishwakarma et al.*** (***2015***), since both have been hypothesised to be important during the evolution and/or execution of the C4 cycle.

The *one-cell*model comprises in total 413 metabolites and 572 reactions, whereof 139 are internal transporters, 90 are export and 8 import reactions (see also below), which are involved in 59 subsystems. ***Figure 1*** provides an overview of the primary subsystems according to ***Arnold and Nikoloski*** (***2014***).

One-cell model reproduces physiological results

Comparing the fluxes for nutrient import and export of the *one-cell* model, see ***Figure 1***(B), with literature data, shows that the determined photosynthetic photon flux density (PPFD) is of about

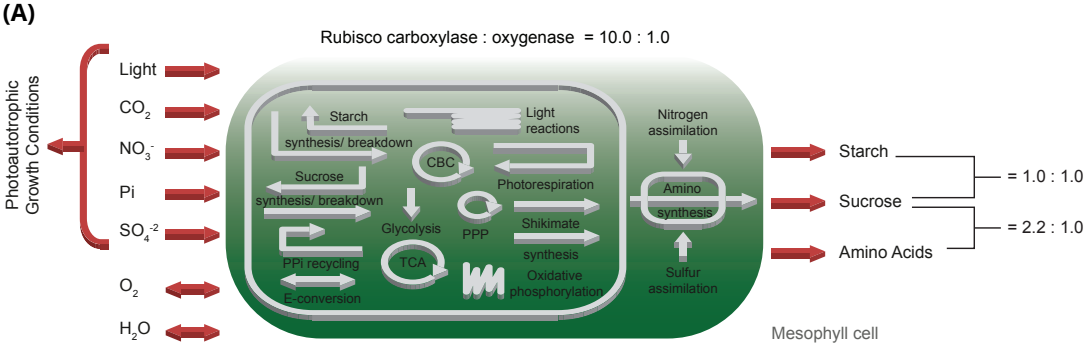
Table 1. Curation of the *Arabidopsis* core model from ***Arnold and Nikoloski*** (***2014***)

***Arabidopsis* core model**	**Observation**	***one-cell* model**	**Reference**
NADP-dependent malate dehydrogenases in all compartments	cycles through nitrate reductase to interconvert NAD and NADP	NAD-dependent malate dehydrogenases in all compartments, NADP-dependent malate dehydrogenase only in chloroplast	***Swarbreck et al.*** (***2007***)
Cyclic electron flow	absence of cyclic electron flow	added	***Shikanai*** (***2016***)
Alternative oxidase	missing alternative routes for electrons to pass the electron transport chain to reduce oxygen	added alternative oxidase reactions to the chloroplast and mitochondria	***Vishwakarma et al.*** (***2015***)
Alanine transferase	No alanine transferase in cytosol Alanine transferase	added	***Liepman*** (***2003***)
Transport chloroplast	no maltose transporter by MEX1 no glucose transporter by MEX1 and pGlcT MEX1 no unidirectional transport of ATP, ADP, AMP by BT-like no Mal/OAA, Mal/Pyr, and Mal/Glu exchange by DiTs no folate transporter by FBT and FOLT1	added added added added added	***Linka and Weber*** (***2010***)
Transport Mitochondria	no Mal/OAA, Cit/iCit, Mal/KG exchange by DTC no H+ importer by UCPs import no OAA/Pi exchange by DIC1-3 no ATP/Pi exchange by APCs no NAD/ADP and NAD/AMP exchange by NDT2 no ThPP/ATP exchange by TPCs no Asp/Glu by AGCs no uncoupled Ala exchange	added added added added added added added added	***Linka and Weber*** (***2010***)
Transport peroxisome	missing NAD/NADH, NAD/ADP, NAD/AMP exchange by PXN no ATP/ADP and ATP/AMP exchange by PNCs	added added	***Linka and Weber*** (***2010***)
H^+ sinks/sources	H^+ sinks/source reaction for the cytosol and futile transport cycles introduced by H^+-coupled transport reactions	H^+ sinks/source reaction added for each compartment	
ATPase stoichiometry	False H^+/ATP ratios for the plastidal and mitochondrial ATP synthase	H^+/ATP ratio set to 3 : 1 (chloroplast) and 4:1 (mitochondria)	***Petersen et al.*** (***2012***); ***Turina et al.*** (***2016***)
Alanine/aspartate transferase	no direct conversion of alanine and aspartate	added to cytosol, chloroplast and mitochondria	***Schultz and Coruzzi*** (***1995***); ***Duff et al.*** (***2012***)

193.7 µE. The maximum uptake rate of CO_2 is used by the *one-cell* model to produce the maximum amount of phloem sap, as well as 0.8 µmol/(m^2 s) of NO_3^- and 18.2 µmol/(m^2 s) of H_2O. According to the assumed ratio of sucrose and amino acids in the phloem sap, the flux of sucrose predicted by the model is 0.8 µmol/(m^2 s) and of amino acids 0.3 µmol/(m^2 s). The rate of oxygen supply by the network is of about 20.9 µmol/(m^2 s).

The CO_2 uptake rate and the phloem sap output have a positive linear relationship, see ***Figure 1–Figure Supplement 1***(A). The same is true for the correlation of the PPFD and phloem sap output in the range of 100 µE–200 µE, see ***Figure 1–Figure Supplement 1***(B). Above 200 µE, the CO_2 uptake rate acts as a limiting factor restricting the increase of phloem sap production. If either the PPFD or the CO_2 uptake rate is zero, the phloem sap cannot be produced compare ***Figure 1–Figure Supplement 1***(A) and (B). Most of the metabolic processes use ATP/ADP as main energy equivalent (60.1 %), followed by NADP/NADPH (37.5 %) and NAD/NADH (2.4 %), see ***Figure 1–Figure Supplement 2***(D).

(A)



(B)

Molecular Species	Flux	Physiological Range	Reference
(i) Inputs			
Photons	193.7 µE	100 µE–400 µE	***Bailey et al.*** (***2001***)
CO_2	20 µmol/(m^2 s)	20 µmol/(m^2 s)	***Lacher*** (***2003***)
NO_3^-	0.5 µmol/(m^2 s)	0.11 µmol/(m^2 s)–0.18 µmol/(m^2 s)	***Kiba et al.*** (***2012***)
H_2O	18.2 µmol/(m^2 s)	-	
(ii) Outputs			
O_2	20.9 µmol/(m^2 s)	16.5 µmol/(m^2 s)	***Sun et al.*** (***1999***)
Amino Acids	0.3 µmol/(m^2 s)	-	
Surcose/Starch	0.8 µmol/(m^2 s)	-	

Note: CO_2 has one carbon per molecule while Sucrose has 12. Starch is configured to have the same number of carbons compared to sucrose while amino acids on average have 5.5 carbons.

Figure 1. (A)Schematic representation of the primary subsystems in the *one-cell* model and the used input/output constraints; adapted from ***Arnold and Nikoloski*** (***2014***). (B) Characteristic input/output fluxes of *one-cell* model in comparison to physiological observations.

Figure 1–source data 1. SBML code of the *one-cell* model

Figure 1–source data 2. Jupyter notebook - Predicted Fluxes of C3 metabolism

Figure 1–Figure supplement 1. Effect of CO_2 and PPFD variation.

Figure 1–source data 3. Jupyter notebook- Effect of the CO_2 uptake rate on C3 metabolism

Figure 1–source data 4. Jupyter notebook - Effect of the PPFD on C3 metabolism

Figure 1–Figure supplement 2. Energy Flux Distribution in the *one-cell* Model.

Figure 1–source data 5. Jupyter notebook - Input, output fluxes and flux of energy in C3 metabolism

Nearly all ATP is produced by the light reactions (97.2 %) and consumed by the reductive pentose phosphate cycle (94.1 %), see ***Figure 1–Figure Supplement 2***(A). The oxidative phosphorylation produces only (1 %) of ATP. In proportion, the maintenance cost for protein synthesis and degradation makeup 28 % of the respiratory ATP produced by the oxidative phosphorylation (***Figure 1–Figure Supplement 2***(E)). Similarly, nearly all NADPH is produced by the light reaction (98.9 %), which is consumed by the reductive pentose-phosphate cycle (98.3 %) as well (***Figure 1–Figure Supplement 2***(B)). The canonical glycolysis and photorespiration produce around equal amounts of NADH, 45 % and 47.7 %, significantly less NADH is produced through the pyruvate dehydrogenase activity 6.85 %. Nitrate assimilation (45 %), glutamate biosynthesis (47.7 %), glyoxylate cycle (21.6 %) and alternative respiration (11.8 %) consume the produced NADH (***Figure 1–Figure Supplement 2***(C)). As such, the model operates within parameters expected for a C3 plant.

C4 as an emergent property

To rebuild the characteristic physiology of C4 leaves, we duplicated the *one-cell* model and connected the two network copies by bi-directional transport of cytosolic metabolites including amino acids, sugars, single phosphorylated sugars, mono-/di-/tri-carboxylic acids, glyceric acids, glycolate, glycerate, glyceraldehyde-3-phosphate, di-hydroxyacetone-phosphate and CO_2, see Methods and Materials for details. Since constraint-based modelling is limited to static model analysis, we intro-

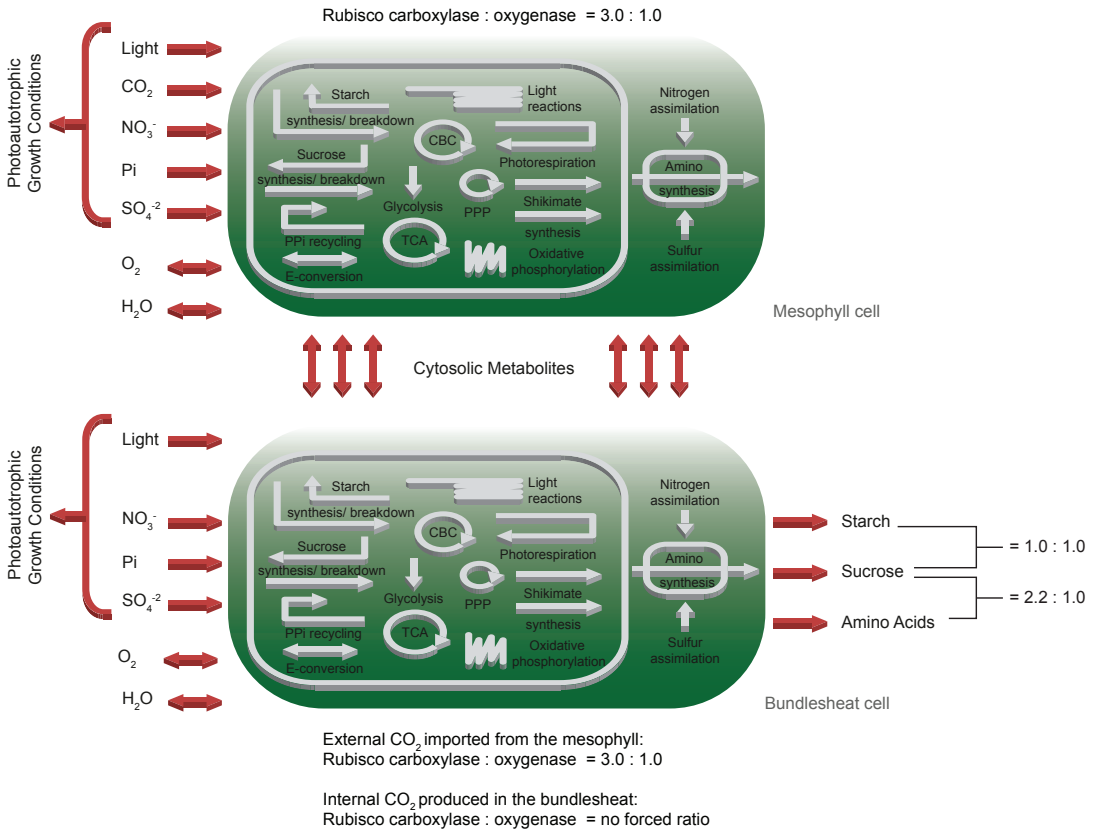


Figure 2. Schematic representation of the primary subsystems in the *two-cell* model and the used input/output constraints; adapted from ***Arnold and Nikoloski*** (***2014***)

duced two Rubisco populations in the bundle sheath network to approximate CO_2 concentration-dependent changes in the oxygenation : carboxylation ratio of Rubisco (v_{RBO}/v_{RBC}) itself. We kept the native constrained Rubisco population that is forced to undertake oxygenation reactions and added a CCM-dependent Rubisco population which can only carboxylate ribulose 1,5-bisphosphate. The CCM-dependent Rubisco population is only able to use CO_2 produced by the bundle sheath network but not environmental CO_2 released by the mesophyll. C4 plants have a higher CO_2 consumption and thus, an increased CO_2 uptake of 40 µmol/(m^2 s) was used ***Leakey*** (***2006***). All other constraints and the objective of the *one-cell* model are maintained in the *two-cell* model, see ***Figure 2***.

Initially, we optimised for the classical objective function of minimal total flux through the metabolic network. If the model produced a C4 solution, it would explain how C4 can evolve without water limitation. This objective function approximates resource use as each reaction in the network requires an enzyme or transporter which in turn relies on its synthesis which depends on the resources. The optimisation was carried out under different fluxes through photorespiration. If the photorespiratory flux is high, the two cell model predicts a metabolic mode which mirrors observations in C4 plants ***Figure 3*** and ***Figure 3–Figure Supplement 1***(A). If the photorespiratory flux is low, the two cell model predicts C3-type photosynthesis identical to the *one-cell* model, see ***Figure 3*** and ***Figure 3–Figure Supplement 1***(A). At intermediate photorespiratory flux, the model predicts increased and distributed flux through photorespiration with increased decarboxylation by glycine decarboxylase ***Figure 3*** and ***Figure 3–Figure Supplement 1***(A).

Enzymes and transport processes characteristic of C4 photosynthesis were plotted to illustrate these changes ***Figure 3*** and ***Figure 3–Figure Supplement 1***(A). At low photorespiratory levels, flux through PEPC is barely detectable (***Figure 3***(A)). If photorespiration increases to moderate levels,

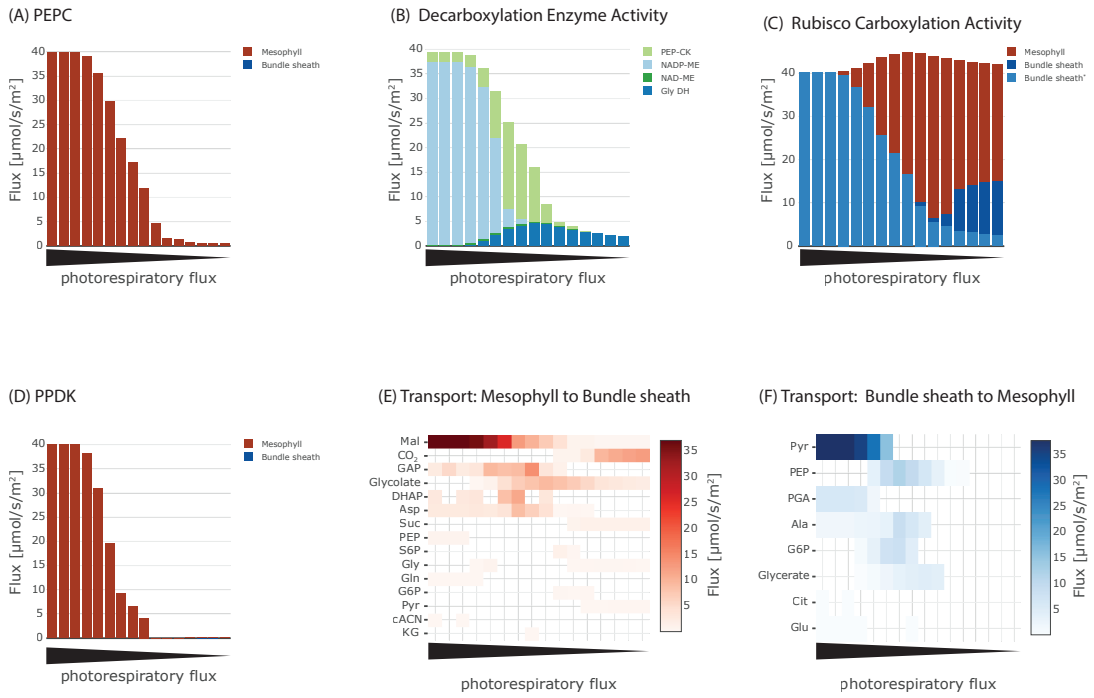


Figure 3. Effect of oxygenation : carboxylation ratio on the major steps in C4 cycle, including (A) activity of phosphoenolpyruvate carboxylase (PEPC), (B) metabolite transport to the bundle sheath, (C) activity of Rubisco, (D) activity of the decarboxylation enzymes, (E) metabolite transport to the mesophyll, and (F) activity of pyruvate phosphate dikinase (PPDK).

Figure 3–source data 1. Jupyter notebook - Analysing the effect of oxygenation : carboxylation ratio on the emergence of the C4 cycle

Figure 3–Figure supplement 1. Flux maps illustrating the effect of the oxygenation : carboxylation ratio of Rubisco on the C3-C4 trajectory.

flux through PEPC can be predicted and increases to 40 µmol/(m^2 s), i.e. all CO_2 is funnelled through PEPC, for high photorespiratory fluxes. Concomitant with flux through PEPC, the activity of the decarboxylation enzymes changes (***Figure 3***(B)). At low to intermediate levels of photorespiratory flux, glycine decarboxylase complex activity is predicted to shuttle CO_2 to the bundle sheath at up to 4.7 µmol/(m^2 s). Decarboxylation of C4 acids is initially mostly mediated by PEP-CK and is largely taken over by NADP-ME at high fluxes through photorespiration. Flux through NAD-ME is very low under all photorespiration levels. The decarboxylation enzymes dictate flux through the different Rubiscos in the model (***Figure 3***(C)). At low photorespiratory flux, both the Rubiscos in mesophyll and bundle sheath are active. Only very little flux occurs through the CCM-dependent Rubisco, which is a result of the glycine decarboxylase (***Figure 3***(B)). With increasing photorespiratory flux, this flux through glycine decarboxylase increases (***Figure 3***(B)) and therefore total Rubisco activity exceeds the carbon intake flux (***Figure 3***(C)). Carbon fixation switches to the CCM-dependent Rubisco with increasing flux through PEPC (***Figure 3***(A)) and the classic C4 cycle decarboxylation enzymes (***Figure 3***(B)). Flux through PPDK mostly reflects flux through PEPC (***Figure 3***(D)). The transport fluxes between the cells change with changing photosynthetic mode (***Figure 3***(E and F)).

At low rates of photorespiration when PEPC is barely active, the only flux towards the bundle sheath is CO_2 diffusion (***Figure 3***(E)) with no fluxes towards the mesophyll (***Figure 3***(F)). In the intermediate phase which is characterised by the bump in Rubisco activity and maximal decarboxylation by glycine decarboxylase complex, glycolate and glycerate are predicted to be transported (***Figure 3***(E) and (F)). At the same, a low-level C4 cycle dependent on the transport of aspartate, malate, PEP and alanine operates (***Figure 3***(E) and (F)). Once the C4 cycle supports carbon assimilation,

the major fluxes between mesophyll and bundle sheath are carried by malate and pyruvate (***Figure 3***(E) and (F)). Flux through PPDK (***Figure 3***(D)) is lower than flux through PEPC (***Figure 3***(A)) at the intermediate stage since the returning C3 acid is PEP in this intermediate stage (***Figure 3***(F)). For low to moderate photorespiration, the costs of photorespiration are not contradicting the minimal metabolic effort requirement, and C3-like photosynthesis as predicted by the *one-cell* model occurs. Most of the CO_2 is directly transferred to the Calvin Benson cycle using the constrained Rubisco population in the mesophyll. Only a few quantities of CO_2 are pushed over to the Calvin Benson cycle in the bundle sheath, see ***Figure 3–Figure Supplement 1***(A). In the *two-cell* model, the C3-C4 intermediate state emerges for moderate photorespiratory fluxes. The characteristic photorespiratory CO_2 pump is established via a glycolate/glycerate shuttle using glycine dehydrogenase in the bundle sheath to supply the CCM-dependent Rubisco population with internally produced CO_2, see ***Figure 3–Figure Supplement 1***(B). Increasing the proportion of photorespiration even further induces the emergence of the C4 cycle. The activity of the C4 cycle enzymes phosphoenolpyruvate kinase and pyruvate phosphate dikinase increases, and a malate/pyruvate shuttle is switched on with NADP-ME as the major decarboxylation enzymes, see ***Figure 3–Figure Supplement 1***(C). NADP-ME produces internal CO_2 that is transferred to the Calvin-Benson cycle through the CCM-dependent Rubisco populations. The C4-cycle is accompanied by a triose phosphate cycle using PGA and GAP/DHAP (***Figure 3***(E and F)) to balance ATP and NADP across mesophyll and bundle sheath. When the model is optimised for minimal total flux and is executed under high photorespiration conditions, a classical C4 cycle is predicted. Since minimal total flux is a proxy for resource efficiency as a selective pressure, we conclude that a C4 photosynthetic cycle is more resource-effective compared to C3 and that resource efficiency is a sufficient selective pressure for C4 photosynthesis to emerge. This model hence predicts that C4 can be evolved in wet habitats. If carbon limitation is assumed and approximated by using minimisation of photorespiration as part of the objective function, the C4 cycle is also predicted (see below).

Differentiation of C4 modes

Among the known independent evolutionary events leading to C4 photosynthesis, 20 are towards NAD-ME while 21 occurred towards NADP-ME ***Sage*** (***2004***). PEP-CK is dominant or at least co-dominant only in *Panicum maximum* ***Bräutigam et al.*** (***2014***), *Alloteropsis semialata semialata* ***Christin et al.*** (***2012***), and in the *Chloridoideae* ***Sage*** (***2004***). We performed three separate experiments, where only one decarboxylation enzyme can be active at a time to analyse whether the emergence of the C4 cycle is independent of the decarboxylation enzyme. The other decarboxylation enzymes were de-activated by constraining the reaction flux to 0. The flux distributions obtained under the assumption of oxygenation : carboxylation ratio of $1:3$ and minimisation of photorespiration as an additional objective predicts the emergence of a C4 cycle for each known decarboxylation enzyme. While the flux distribution in the mesophyll is identical for three predicted C4 cycles of the decarboxylation enzymes, it is diverse in the bundle sheath due to the different localisation of the decarboxylation and related transport processes, see ***Figure 4***. The flux distribution does not mimic the known variation in transfer acids since all of the decarboxylation enzymes use the malate/pyruvate shuttle. In the case of NAD-ME and PEP-CK, the *two-cell* model also predicts a supplementary flux through the aspartate/alanine shuttle. In the literature, NAD-ME uses aspartate/alanine shuttle solely, while PEP-CK is known to use the aspartate/alanine and malate/pyruvate shuttle in parallel ***Hatch*** (***1987***). The observations from the constraint-based model contradict the known physiology of C4 photosynthesis at least in the case of decarboxylation by NAD-ME, so we tested whether transfer acids other than malate and pyruvate are feasible and explored the near-optimal space. Performing a flux variability analysis (FVA) and allowing the minimal total flux to differ by 1.5 %, predicts that for most metabolites which are transferred between mesophyll and bundle sheath, the variability is similar for all three decarboxylation types. Differences between the decarboxylation types were observed only for NAD-ME and PEP-CK types and only for the transfer acids malate, aspartate, pyruvate and alanine. Minor differences were present for triose

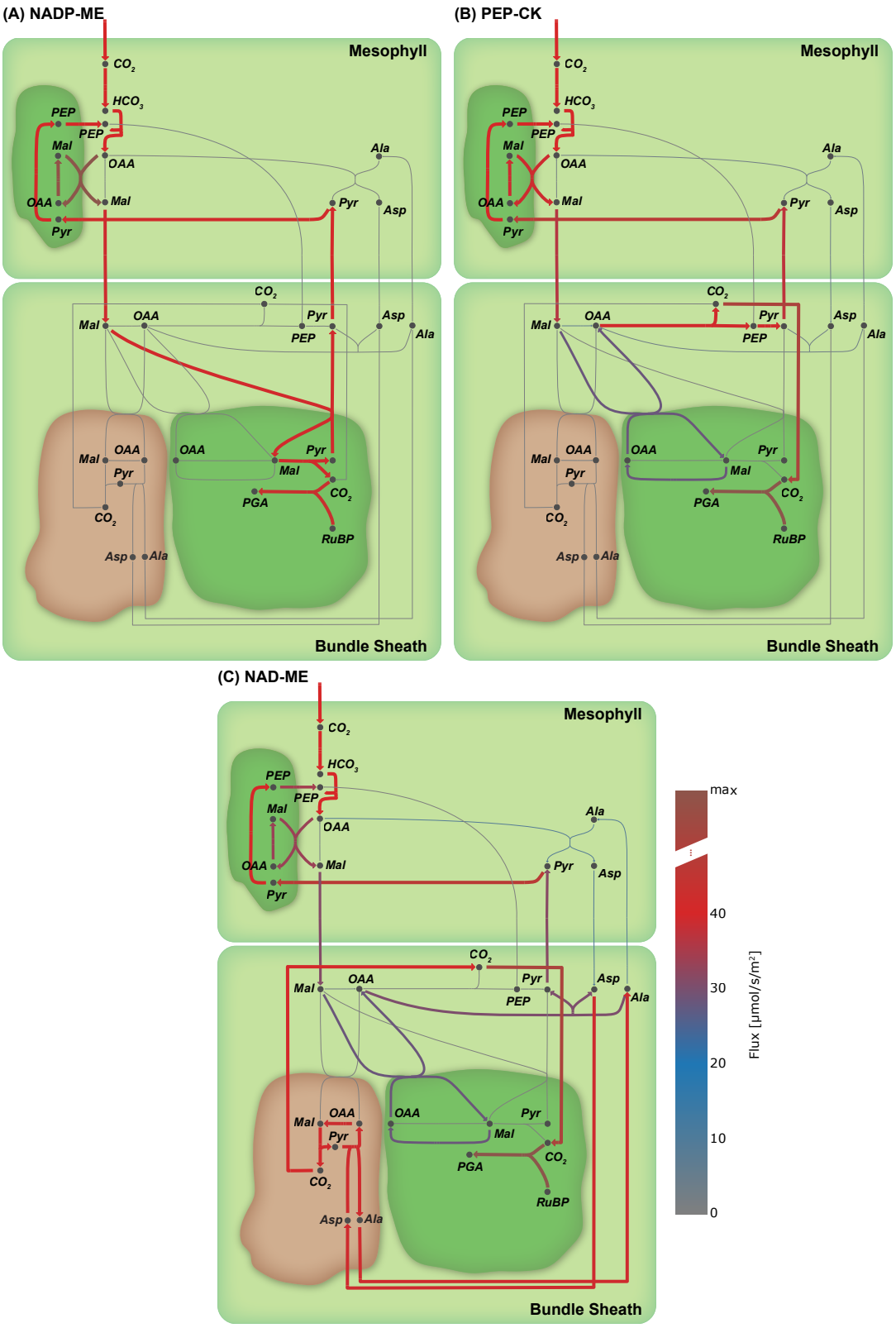


Figure 4. Flux maps illustrating the effect of the C4 mode. (A) NADP-ME, (B) PEP-CK, (C) NAD-ME. (Arc width and colour are set relative to flux values in flux, grey arcs – no flux)

Figure 4–source data 1. Jupyter notebook - Effect of C4 mode on the emergence of the C4 cycle

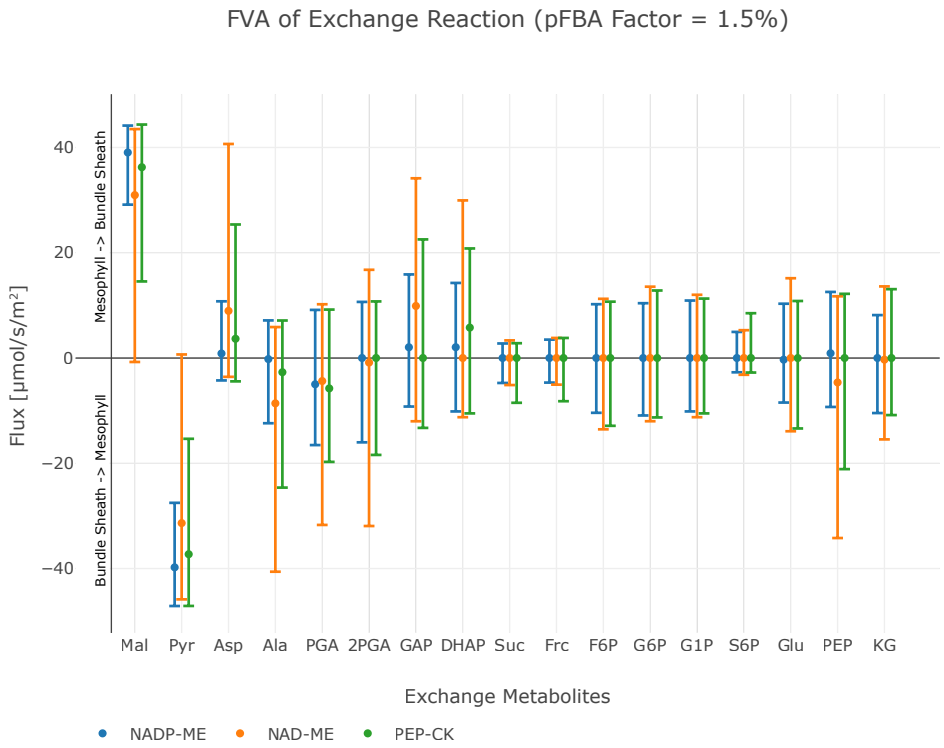


Figure 5. Flux variability analysis of metabolite exchange with 1.5 % deviation of the total flux minimum. The upper bar defines the maximum exchange flux, while the lower bar defines the minimum exchange flux, points indicate the value of the original flux solution under minimal metabolic effort constraint. Positive flux values correspond to the transport direction from mesophyll to bundle sheath, negative values to the transport direction from bundle sheath to mesophyll, see also ***Figure 4–source data 1***.

phosphates and phosphoglycerates as well as for PEP. FVA identifies only minor variation for the NADP-ME type ***Figure 5***. In the case of NAD-ME but not in the case of NADP-ME the activity of the malate/pyruvate shuttle can be taken over by the aspartate/alanine shuttle and partly taken over in case of PEP-CK, see ***Figure 5***. The aspartate/alanine shuttle is thus only a near optimal solution when the model and by proxy evolutionary constraints are resource efficiency and minimal photorespiration.

Light as evolutionary driver of the C4 mode

A model in which C4 emerges without constraints (***Figure 3***) and in which each decarboxylation enzyme represents a possible path for flux (***Figure 4***) can be used to test for conditions under which different decarboxylation fluxes are optimal. By identifying the constraints and objective functions under which different fluxes through the decarboxylation enzymes emerge as the optimal solution, the selective forces which drove them can be approximated. To analyse the effect of other conditions on the particular C4 state, we apply the minimisation of photorespiration as an additional objective, to ensure the emergence of the C4 cycle. Since NAD-ME and PEP-CK type plants use amino acids as transfer acids in nature, nitrogen availability has been tagged as a possible evolutionary constraint that selects for decarboxylation by NAD-ME or PEP-CK. When nitrate uptake was limiting, the optimal solution to the model predicted the same behaviour as was predicted for unlimited nitrate uptake ***Figure 6–Figure Supplement 1***. Flux through NADP-ME and supplementary flux through PEP-CK dropped proportionally, since restricting nitrogen limits the export of all metabolites from the system ***Figure 6–Figure Supplement 1***. The likely reason is

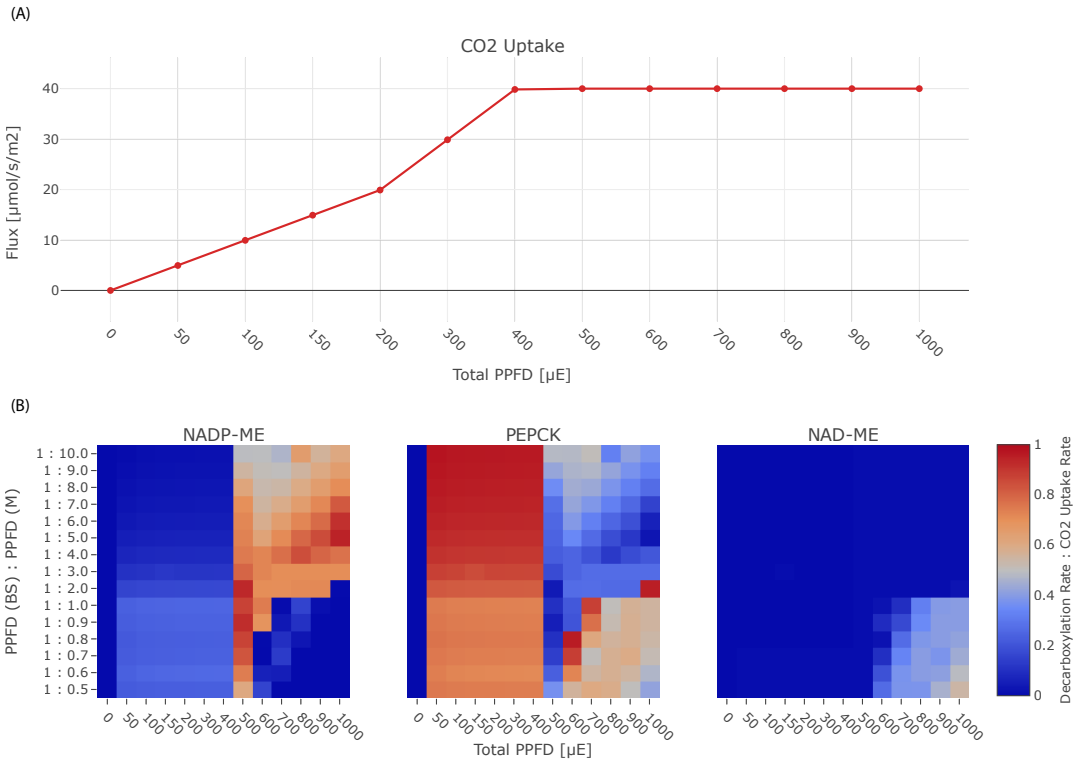


Figure 6. Effect of light on the C4 mode. (A) CO_2 uptake rate in dependence of the total PPFD, (B) Heat-maps illustrating the activity of the decarboxylation enzymes PEP-CK, NADP-ME, and NAD-ME relative to the CO_2 uptake rate in dependence of the total PPFD and the photon distribution among mesophyll and bundle sheath.

Figure 6–source data 1. Jupyter notebook - Effect of light on the C4 mode

Figure 6–Figure supplement 1. Effect of other relevant factors on the C4 mode.

Figure 6–source data 2. Jupyter notebook - Effect of NO_3^- limitation on the C4 mode

Figure 6–source data 3. Jupyter notebook - Effect of H_2O limitation on the C4 mode

Figure 6–source data 4. Jupyter notebook - Effect of CO_2 limitation on the C4 mode

Figure 6–source data 5. Jupyter notebook - Effect of malate : aspartate transport ratio on the C4 mode

a fixed output function for an amino acid to sucrose ratio. Similarly, limiting water or CO_2 uptake into the model resulted in the same optimal solution as unlimited uptake with the difference of proportionally lower flux overall, see ***Figure 6–Figure Supplement 1***.

Given that C4 plants sometimes optimise light availability to the bundle sheath ***Bellasio and Lundgren*** (***2016***) we next explored light availability and light distribution. In the experiment, we varied the total PPFD between 0 µE to 1000 µE and photon distribution in the range $0.1 \leq PPFD_B$ / $PPFD_M \leq 2$, see ***Figure 6***. Under light limitation, if the total PPFD is lower than 400 µE, the CO_2 uptake rate is reduced, leading to a decreased activity of the decarboxylation enzymes (***Figure 6***(A)). PEP-CK is used in the optimal solutions active under light-limiting conditions (***Figure 6***(B). Under limiting light conditions, photon distribution with a larger proportion in the bundle sheath shifts decarboxylation towards NADP-ME but only to up to 26 %. Under non-limiting conditions, the distribution of light availability determines the optimal decarboxylation enzyme. NADP-ME is the preferred decarboxylation enzyme with supplemental contributions by PEP-CK if light availability is near the threshold of 400 µE or if at least twice as many photons are absorbed by the mesophyll. Excess light availability and a larger proportion of photons reaching the bundle sheath leads to optimal solutions which favour PEP-CK as the decarboxylation enzyme. In the case of very high light availability and an abrupt shift towards the bundle sheath, NAD-ME becomes the optimal solution (***Figure 6***(B)). NAD-ME is the least favourable enzyme overall, only low activity is predicted

under extreme light conditions, where the bundle sheath absorbs equal or more photons than the mesophyll ***Figure 6***(B). PEP-CK complements the activity of NADP-ME and NAD-ME to 100 % in many conditions, meaning the *two-cell* model also predicts the co-existence of PEP-CK/NADP-ME and PEP-CK/NAD-ME mode, while the flux distribution indicates no parallel use of NAD-ME and NADP-ME, compare ***Figure 6***(B).

Finally, we assumed that intercellular transport capacity for charged metabolites might be different between species. Assuming a fixed transport ratio between aspartate and malate (***Figure 6–Figure Supplement 1***D) introduces a shift in the C4 state. Higher proportions of malate exchange foster the use of NADP-ME (***Figure 6–Figure Supplement 1***D). In contrast, higher portions of aspartate exchange foster the use of PEP-CK (***Figure 6–Figure Supplement 1***D). Therefore, the aspartate : malate transport ratio, which is most likely connected with the abundance of transporters and their kinetics, might have also affected the choice of the decarboxylation enzyme during C4 evolution.

Discussion

Evolutionary constraint-based modelling can suggest the molecular outcomes of past evolutionary events if models are parametrised with objective functions representing possible selective pressures. In the case of C4 photosynthesis, more than sixty independent evolutionary origins represent metabolic types characterised by their decarboxylation enzyme. The selective pressure which drives evolution towards one or the other flux are unknown and were tested using constraint-based modelling.

One-cell model reflects C3 plant physiology

To analyse evolution towards C4 photosynthesis based on C3 metabolism, a constraint-based model of C3 metabolism is required (***Figure 1***). Design, simulation, validation cycles used current knowledge about plant physiology to identify possible errors in the metabolic map required for modelling. Even after error correction ***Table 1***, a significant problem remained, namely excessive fluxes to balance protons in all compartments. This observation leads to the realisation that the biochemical knowledge about transport reactions does not extend to the protonation state of the substrates which affects all eukaryotic constraint-based modelling efforts. In plants, predominantly export and vacuolar transport reactions are directly or indirectly coupled with proton gradients to energise transport ***Bush*** (***1993***); ***Neuhaus*** (***2007***). For chloroplasts and mitochondria, proton-coupled transport reactions have been described but may couple different metabolite transporters together rather than energising them ***Furumoto et al.*** (***2011***). Introducing proton sinks in all compartments solves the immediate modelling problem. However, intracellular transport reactions and their energetic costs are no longer correctly assessed by the model. Despite this band-aid fix which will be required for all eukaryotic constraint-based models which include proton-coupled transport reactions, the curated *one-cell* model correctly predicts energy usage and its distribution (***Figure 1–Figure Supplement 2*** and ***Li et al.*** (***2017***)). This indicates that in models which exclude vacuolar transport and energised export reactions, energy calculations remain likely within the correct order of magnitude. Overall, our *one-cell* model operates within parameters expected for a C3 plant: The predicted PPFD lies within the range of light intensities used for normal growth condition of *Arabidopsis thaliana*, which varies between 100 µE–200 µE, see ***Figure 1***(B). The gross rate of O_2 evolution for a PPFD of 200 µE is estimated to be 16.5 µmol/(m^2 s) in the literature ***Sun et al.*** (***1999***), which is in close proximity to the predicted flux of the *one-cell* model, see ***Figure 1***(B). For the amount of respiratory ATP that is used for maintenance, ***Li et al.*** (***2017***) predicted an even lower proportion of energy 16 %, see ***Figure 1–Figure Supplement 2***. The model's flux map is in accordance with known C3 plant physiology ***Heldt and Piechulla*** (***2015***), and its input and output parameters match expected values (***Figure 2***(B)).

Two-cell model predicts C4 plant physiology

Most evolutionary concepts about C4 photosynthesis assume that selective pressure drives pathway evolution due to photorespiration and carbon limitation ***Heckmann et al.*** (***2013***). Most extant C4 species occupy dry and arid niches ***Edwards et al.*** (***2010***), even more the period of C4 plant evolution was accompanied with an increase oxygen concentration in the atmosphere ***Sage*** (***2004***). Therefore, it is frequently assumed that carbon limitation by excessive photorespiration drives the evolution of C4 photosynthesis. Yet, in most habitats plants are limited by nutrients other than carbon ***Ågren et al.*** (***2012***); ***Körner*** (***2015***). Ecophysiological analyses also show that C4 can emerge in non-arid habitats ***Liu and Osborne*** (***2014***); ***Lundgren and Christin*** (***2016***); ***Osborne and Freckleton*** (***2009***). To resolve this apparent contradiction, we tested whether resource limitation may also lead to the evolution of a C4 cycle. We optimised the model approximating resource limitation via an objective function for total minimal flux at different photorespiratory levels. Indeed, with increasing photorespiration, the optimisation for resource efficiency leads to the emergence of the C4 cycle as the optimal solution. Balancing the resource cost of photorespiration against the resource cost of the C4 cycle, the model predicts that N limitation may have facilitated C4 evolution given high levels of photorespiration. Extant C4 species have higher C:N ratios reflecting the N savings the operational C4 cycle enables ***Sage et al.*** (***1987***). The photorespiratory pump using glycine decarboxylase based CO2 enrichment also emerges from the model, showing that C2 photosynthesis is also predicted under simple resource limitation. Indeed N savings have been reported from C2 plants compared with their C3 sister lineages ***Schlüter et al.*** (***2016***). Simply minimising photorespiration as the objective function also yields C4 photosynthesis as the optimal solution. Hence, two alternatively or parallelly acting selective pressure towards C4 photosynthesis are identified by the model. In both cases, the model correctly predicts the C4 cycle of carboxylation and decarboxylation and the C2 photorespiratory pump as observed in extant plants.

The optimal solutions for the metabolic flux patterns predict an intermediate stage in which CO_2 transport via photorespiratory intermediates glycolate and glycerate (***Figure 3***(E) and (F)) and decarboxylation by glycine decarboxylase complex (***Figure 3***(B)) is essential. All of the models of C4 evolution (***Monson*** (***1999***); ***Bauwe*** (***2010***); ***Sage et al.*** (***2012***); ***Heckmann et al.*** (***2013***); ***Williams et al.*** (***2013***)) predict that the establishment of a photorespiratory CO_2 pump is an essential intermediate step towards the C4 cycle. The photorespiratory CO_2 pump, also known as C2 photosynthesis, relocates the photorespiratory CO_2 release to the bundle sheath cells. Plants using the photorespiratory CO_2 pump are often termed C3-C4 intermediates owing to their physiological properties ***Sage et al.*** (***2012***). Displaying the flux solution in ***Figure 3*** on a metabolic map in ***Figure 3–Figure Supplement 1*** clearly illustrates that increasing photorespiratory flux through Rubisco drives the two-cell metabolic model from C3 to C4 metabolism by passing the C3-C4 intermediate state. On the C3-C4 trajectory, the activity of Rubisco is shifted from the mesophyll to the bundle sheath, as well as from the constrained to the CCM-dependent Rubisco population as a consequence of the increased costs of photorespiration under increased $p_{O_2} : p_{CO_2}$ ratio, see ***Equation 5***. The increase of the oxygenation rate in the photorespiration constraint drives the reprogramming of the metabolism to avoid oxygenation by establishing the C4 cycle. Therefore, our analysis recovers the evolutionary C3-C4 trajectory and confirms the emergence of a photorespiratory CO_2 pump as an essential step during the C4 evolution also under optimisation for resources ***Heckmann et al.*** (***2013***). The model may also provide a reason for why some plant species have halted their evolution in this intermediary phase ***Scheben et al.*** (***2017***). Under the conditions of resource limitations and intermediate photorespiration, the model predicts intermediacy as the optimal solution. In a very narrow corridor of conditions, no further changes are required to reach optimality.

Two-cell model realises different C4 states

Since the C4 cycle is an emergent property of the model, different input and reaction constraints can be tested for their influence on the molecular nature of the C4 cycle. This approach may identify the selective pressure and boundaries limiting evolution. Initial optimisation without additional

constraints or input limitations predict a C4 cycle based on decarboxylation by NADP-ME (***Figure 3*** and ***Figure 3–Figure Supplement 1***(A)). This prediction recapitulates intuition; the NADP-ME based C4 cycle is considered the "most straight forward" incarnation of C4 photosynthesis, it is always explained first in textbooks and is a major focus of research. The NADP-ME based cycle thus represents the optimal solution when resource limitation or photorespiration are considered. Once NADP-ME is no longer available via constraint, PEP-CK and NAD-ME become optimal solutions albeit with a prediction of malate and pyruvate as the transfer acids (***Figure 6***). The FVA identified aspartate and alanine as slightly less optimal solutions (***Figure 5***). Since *in vivo* this slightly less optimal solution has evolved in all NAD-ME origins tested to date, kinetic rather than stoichiometric reasons suggest themselves for the use of aspartate and alanine ***Bräutigam et al.*** (***2018***).

Light is a potential evolutionary driver for the different C4 states

Since all extant C3 species and therefore also the ancestors of all C4 species contain all decarboxylation enzymes ***Aubry et al.*** (***2011***), it is unlikely that unavailability of an enzyme is the reason for the evolution of different decarboxylation enzymes in different origins ***Sage*** (***2004***). Stochastic processes during evolution, i.e. up-regulation of particular enzyme concentrations via changes in expression and therefore elements *cis* to the gene ***Bräutigam and Gowik*** (***2016***), may have played a role in determining which C4 cycle evolved. Alternatively, environmental determinants may have contributed to the evolution of different C4 cycles. Physiological experiments have pointed to a connection between nitrogen use efficiency and type of decarboxylation enzyme ***Pinto et al.*** (***2015***). Hence the variation in nitrogen input to the model was tested for their influence on optimal solutions with regard to decarboxylation enzymes. Input limitation of nitrogen, water as a metabolite, and CO_2 limited the output of the system but did not change the optimal solution concerning decarboxylation ***Figure 6–Figure Supplement 1*** making it an unlikely candidate as the cause. Differences in nitrogen use is possibly a consequence of decarboxylation type.

In some grasses, light penetrable cells overlay the vascular bundle leading to different light availability (summarised in ***Bellasio and Lundgren*** (***2016***); ***Karabourniotis et al.*** (***2000***)) and hence light availability and distribution were tested (***Figure 6***(B)). Changes in light input and distribution of light input between mesophyll and bundle sheath indeed altered the optimal solutions (***Figure 6***(B)). The changes in the solution can be traced to the energy status of the plant cells. For very high light intensities, the alternative oxidases in the mitochondria are used to dissipate the energy and hence a path towards NAD-ME is paved. Under light limitation, the C4 cycle requires high efficiency and hence PEP-CK which, at least in part allows energy conservation by using PEP rather than pyruvate as the returning C4 acid, is favoured. Interestingly, the sensitivity of different species towards environmental changes in light is influenced by the decarboxylation enzyme present ***Sonawane et al.*** (***2018***). NADP-ME species are less compromised compared to NAD-ME species by shade possibly reflecting an evolutionary remnant as NAD-ME is predicted to emerge only in high light conditions. PEP-CK is more energy efficient compared to malic enzyme based decarboxylation which requires PEP recycling by PPDK at the cost of two molecules of ATP ***Figure 3***(D). Notably, two C4 plants known to rely on PEP-CK *P. maximum* and *A. semialata* (African accessions) are shade plants which grow in the understory ***Lundgren and Christin*** (***2016***). PEP-CK can be co-active with NADP-ME and NAD-ME (***Figure 6***(B)). This co-use of PEP-CK with a malic enzyme has been shown in C4 plants ***Pick et al.*** (***2011***); ***Wingler et al.*** (***1999***) and explained as an adaptation to different energy availability and changes in light conditions ***Pick et al.*** (***2011***); ***Bellasio and Griffiths*** (***2013***). Dominant use of PEP-CK in the absence of malic enzyme activity as suggested (***Figure 3***(B), ***Figure 3–Figure Supplement 1*** and ***Figure 4***)) is rare *in vivo* ***Ueno and Sentoku*** (***2006***). But observed in *P. maximum* and in *A. semialata*.

Conclusion

Constraint-based modelling of photosynthetically active plant cells revealed a major knowledge gap impeding constraint-based modelling namely the unknown protonation state of most transport

substrates during intracellular transport processes. Single cell models suggested C3 photosynthetic metabolism as the optimal solution. After model duplication and the introduction of a CCM-dependent Rubisco, a molecularly correct C4 cycle emerged as the optimal solution under resource limitation and photorespiration reduction as objective functions which points to resource limitation as an additional driver of C4 evolution. Light and light distribution was the environmental variable governing the choice of decarboxylation enzymes. Modelling in compartmented eukaryotic cells correctly predicts the evolutionary trajectories leading to extant C4 photosynthetic plant species.

Methods and Materials

Flux Balance Analysis

Flux balance analysis (FBA) is a constraint-based modelling approach ***Orth et al.*** (***2010***) to investigate the steady-state behaviour of a metabolic network defined by its stoichiometric matrix S. By employing linear programming, FBA allows computing an optimised flux distribution that minimises and/or maximises the synthesis and/or consumption rate of one specific metabolite or a combination of various metabolites. Next to the steady-state assumption and stoichiometric matrix S, FBA relies on the definition of the reaction directionality and reversibility, denoted by the lower bound v_{min} and upper bound v_{max} , as well as the definition of an objective function z. The objective function z defines a flux distribution v, with respect to an objective c.

$$\begin{aligned} \text{min/max} \quad & z_{_{FBA}} = c^T v \\ \text{s.t.} \quad & \\ & S \cdot v = 0 \\ & v_{min} \leq v \leq v_{max} \end{aligned} \tag{1}$$

The degeneracy problem, the possible existence of alternate optimal solutions, is one of the major issues of constraint-based optimisation, such as FBA ***Mahadevan and Schilling*** (***2003***). To avoid this problem, we use the parsimonious version of FBA (pFBA) ***Lewis et al.*** (***2010***). This approach incorporates the flux parsimony as a constraint to find the solution with the minimum absolute flux value among the alternative optima, which is in agreement with the assumption that the cell is evolutionary optimised to allocate a minimum amount of resources to achieve its objective.

$$\begin{aligned} \text{min/max} \quad & z_{_{pFBA}} = \sum |v_i| \\ \text{s.t.} \quad & \\ & S \cdot v = 0 \\ & v_{min} \leq v \leq v_{max} \\ & c^T v = z_{_{FBA}} \end{aligned} \tag{2}$$

All FBA experiments in this study employ pFBA and are performed using the cobrapy module in a python 2.7 environment run on a personal computer (macOS Sierra, 4 GHz Intel Core i7, 32 GB 1867 MHz DDR3). All FBA experiments are available as jupyter notebooks in the supplementary material.

Generic Model for C3 Metabolism

Metabolic Model. The generic model representing the metabolism of a mesophyll cell of a mature photosynthetically active C3 leaf, further on called *one-cell* model, is based on the *Arabidopsis* core model ***Arnold and Nikoloski*** (***2014***). The model is compartmentalised into cytosol (c), chloroplast (h), mitochondria (m), and peroxisome (p). Each reaction in the *Arabidopsis* core model ***Arnold and Nikoloski*** (***2014***) was compared with the corresponding entry in AraCyc ***Mueller et al.*** (***2003***). Based on the given information, we corrected co-factors, gene associations, enzyme

commission numbers and reversibility (information from BRENDA **Schomburg et al.** (**2002**) were included). The gene associations and their GO terms **Ashburner et al.** (**2000**) of the cellular components were used to correct the location of reactions. Major additions to the model are the cyclic electron flow **Shikanai** (**2016**), alternative oxidases in mitochondria and chloroplast **Vishwakarma et al.** (**2015**), as well as several transport processes between the compartments and the cytosol **Linka and Weber** (**2010**). NAD-dependent dehydrogenase to oxidise malate is present in all compartments **Gietl** (**1992**); **Berkemeyer et al.** (**1998**), which excludes the interconversion of NAD and NADP by cycles through the nitrate reductase present in the *Arabidopsis* core model. Correctly defining the protonation state of the metabolites in the various cellular compartments is a general drawback of metabolic models due to the lack of knowledge in that area. This issue mainly affects biochemical reactions and transport reactions involving protons. We added a sink/source reaction for protons in the form:

$$\leftrightarrow H_\{x\} \quad x = c, h, m, p \tag{3}$$

to all compartments to prevent futile fluxes of protons and other metabolites coupled through the proton transport. The curated *one-cell* model is provided in **Figure 1–source data 1**.

Table 2. Flux Boundary Constraints of Im-/Export Reactions

Input (Reaction ID)	Flux [µmol/(m^2 s)]	
	Lower bound	Upper bound
Photons (Im_hnu)	0	inf
CO_2 (Im_CO2)	0	20
NO_3^- (Im_NO3)	0	inf
NH_4^+ (Im_NH4)	0	0
SO_4^{2-} (Im_SO4)	0	inf
H_2S (Im_H2S)	0	inf
Pi	0	inf
H_2O (Im_H2O)	-inf	inf
O_2 (Im_O2)	-inf	inf
Amino Acids (Ex_AA)	0	inf
Surcose (Ex_Suc)	0	inf
Starch (Ex_starch)	0	inf
Other export reactions	0	0

-inf/inf is approximated by -10^6 / 10^6

Import. As in **Arnold and Nikoloski** (**2014**), we assume photoautotrophic growth conditions. Only the import of light, water, CO_2, inorganic phosphate (Pi), nitrate/ammonium, and sulphates/hydrogen sulphide is allowed, compare **Table 2**. More specifically, we do only allow for nitrate uptake, since it is the main source (80 %) of nitrogen in leaves **Macduff and Bakken** (**2003**). The CO_2 uptake is limited to 20 µmol/(m^2 s) **Lacher** (**2003**). Therefore, the carbon input constrains the model.

Export. In contrast to **Arnold and Nikoloski** (**2014**), we focus on mature, fully differentiated and photosynthetic active leaves supporting the growth of the plant through the export of nutrients in the phloem sap, mainly sucrose and amino acids. An output reaction for sucrose *Ex_Suc* is already included in the model. An additional export reaction *Ex_AA* represents the relative proportion of 18 amino acids in the phloem sap of *Arabidopsis* as stoichiometric coefficients in accordance to experimentally measured data from **Wilkinson and Douglas** (**2003**). The ratio of exported sucrose : total amino acid is estimated to be $2.2 : 1$ **Wilkinson and Douglas** (**2003**). This ratio is included as a

flux ratio constraint of the reactions *Ex_Suc* and *Ex_AA*. Furthermore, it is known that the export of sucrose and the formation of starch is approximately the same ***Stitt and Zeeman*** (***2012***), which is reflected by the flux ratio constraint $v_{Ex_Suc} : v_{Ex_starch} = 1 : 1$. The model allows for the export of water and oxygen. The flux of all other export reactions is set to 0, see ***Table 2*** for a summary.

Additional Constraints. We explicitly include the maintenance costs in our model to cover the amounts of ATP that is used to degradation and re-synthesis proteins for each compartment. ***Li et al.*** (***2017***) specifies the ATP costs for protein degradation and synthesis of each compartment of a mature *Arabidopsis* leaf. Based on the given data, we were able to calculate the flux rates to constrain the maintenance reactions in each compartment ***Table 3***.

Table 3. Maintenance Costs by Compartment

Compartment	Flux [µmol/(m^2 s)]
cytosol	0.0427
chloroplast	0.1527
mitochondria	0.0091
peroxisome	0.0076

The *one-cell* model contains maintenance reactions only for the cytsol (*NGAM_c*), chloroplast (*NGAM_h*) and mitochondria (*NGAM_m*) in the form:

$$ATP_\{x\} + H2O_\{x\} \rightarrow ADP_\{x\} + H_\{x\} + Pi_\{x\} \quad x = c, h, m \tag{4}$$

An equivalent maintenance reaction cannot be formulated for the peroxisome since in the *one-cell* model ATP/ADP are not included as peroxisomal metabolites. The flux through the maintenance reactions is fixed to the determined maintenance costs given in ***Table 3***. The peroxisomal maintenance costs are added to the cytosolic maintenance costs.

The CO_2 and O_2 partial pressures determine the ratio of the oxygenation : carboxylation rate of Rubisco (given by reactions *RBO_h* and *RBC_h*) and can be described by the mathematical expression:

$$\frac{v_{RBO_h}}{v_{RBC_h}} = \frac{1}{S_R} \cdot \frac{p_{O_2}}{p_{CO_2}} \tag{5}$$

, where S_R specifies the ability of Rubisco to bind CO_2 over O_2. In the case of a mature leave and ambient CO_2 and O_2 partial pressures in temperate regions with adequate water supply, the ratio v_{RBO_h}/v_{RBC_h} is fixed and is predicted to be 10 %, which is encoded by an additional flux ratio constraint.

We assume no flux for the chloroplastic NADPH dehydrogenase (*iCitDHNADP_h*) and plastoquinol oxidase (*AOX4_h*) because ***Josse et al.*** (***2000***) and ***Yamamoto et al.*** (***2011***) have shown that their effect on the photosynthesis is considerably low.

Objective. In accordance with the assumption of mature, fully differentiated and photosynthetic active leaf, the model's objective is to maximise the phloem sap output defined by reactions *Ex_Suc* and *Ex_AA*. Additionally, we assume that the involved plant cells put only a minimal metabolic effort, in the form of energy and resources, into the production of phloem sap as possible. This assumption is in correspondence with minimising the nitrogen investment by reducing the number of enzymes that are active in a metabolic network. Therefore, we perform a parsimonious FBA to minimise the total flux.

For enhanced compliance with the recent standards of the systems biology community, the *one-cell* model is encoded in SBML level 3. Meta-information on subsystems, publications, cross-references are provided as evidence code in the form of MIRIAM URI's. FBA related information,

gene association rules, charge and formula of a species element are encoded using the Flux Balance Constraints package developed for SBML level 3. All fluxes in the model are consistently defined as μmol/(m^2 s).

Generic Model for C4 Metabolism

Metabolic Model. The generic model of C4 metabolism, short *two-cell* model, comprises two copies of the *one-cell* model to represent one mesophyll and one bundle sheath cell. Reactions and metabolites belonging to the metabolic network of the mesophyll are indicated with the prefix *[M]*, whereas the prefix for the bundle sheath is *[B]*. The separate mesophyll and bundle sheath networks are connected via reversible transport reactions of the cytosolic metabolites indicated with the prefix *[MB]*, ***Figure 2***. The C4 evolution not only confined Rubisco to the bundle sheath cells, the CO_2 concentrating mechanism steadily supplies Rubisco with CO_2 in such a way that the oxygenation rate is negligible. Therefore, the bundle sheath network is equipped with two Rubisco populations. The native Rubisco population binds external CO_2 and adheres to forced oxygenation : carboxylation ratios, where the optimised evolutionary population binds only internal CO_2 and the carboxylation occurs independently of the oxygenation. External CO_2 is defined as *[B]_CO2_ex_{c,h}* supplied by the mesophyll network. Internal CO_2 given by *[B]_CO2_{c,h,m}* originates from reactions in the bundle sheath network producing CO_2. External CO_2in the bundle sheath network is only allowed to move to the chloroplast *[B]_Tr_CO2h_Ex* and to react with Rubisco *[B]_RBC_h_Ex*. The differentiation of two Rubisco populations binding either external or internal CO_2 approximates the concentration-dependent shift of the oxygenation : carboxylation ratio.

Imports. As for the *one-cell* model, we assume photoautotrophic growth conditions, see ***Table 2***. During C4 evolution the CO_2 assimilation became more efficient allowing higher CO_2 assimilation rates. *Zea mays* achieves up to 40 μmol/(m^2 s) ([M]_Im_CO_2) ***Rozema*** (***1993***). We assume that the CO_2 uptake from the environment by the bundle sheath has to be bridged by the mesophyll. Therefore, the input flux of [B]_Im_$CO_2$is set to zero.

Exports. The outputs of the *one-cell* model are transferred to the mesophyll and bundle sheath network, as well as the corresponding flux ratios, see ***Table 2***.

Additional Constraints. The ATP costs for cell maintenance in the *genC3* model are assigned to both cell types in the *two-cell* model. Due to declining CO_2 concentrations over evolutionary time and/or adverse conditions which close the stromata, the oxygenation : decarboxylation ratio of the native Rubisco population in the bundle sheath and the mesophyll is increased and can be predicted as $1:3$, the corresponding flux ratios are adapted accordingly. Furthermore, we assume that the total photon uptake in the mesophyll and bundle sheath is in the range of 0 μE to 1000 μE. Since they are more central in the leaf, the photon uptake by the bundle sheath must be equal or less compared to the mesophyll. The mesophyll and bundle sheath networks are connected by a range of cytosolic transport metabolites including amino acids, sugars (glucose, fructose, sucrose, trehalose, ribose), single phosphorylated sugar (glucose-6-phosphate, glucose-1-phosphate, fructose-6-phosphate, sucrose-6-phosphate), mono-/di-/tri-carboxylic acids (phosphoenolpyruvate, pyruvate, citrate, cis-aconitate, isocitrate, α -ketoglutarate, succinate, fumarate, malate), glyceric acids (2-Phosphoglycerate, 3-Phosphoglycerate), glycolate, glycerate, glyceraldehyde-3-phosphate, di-hydroxyacetone-phosphate and CO_2. Nucleotides, NAD/NADH, NADP/NADPH, pyrophosphate, inorganic phosphate are not considered as transport metabolites. Oxaloacetate has been excluded as transport metabolite since concentrations of oxaloacetate are very low *in vivo* and it is reasonably unstable in aqueous solutions. Other small molecules that can be imported by the bundle sheath from the environment, as well as protons and HCO_3^-, are not exchanged between the two cell types.

Objective. The maximisation of the phloem sap output through the bundle sheath and the minimisation of the metabolic effort are kept as objectives in the *two-cell* model.

Acknowledgments

We like to thank Udo Gowik (Carl von Ossietzky University Oldenburg, Germany) and Urte Schlüter (Heinrich-Heine-University Düsseldorf, Germany) for critically revising this manuscript.

References

Ågren GI, Wetterstedt JÅM, Billberger MFK. Nutrient limitation on terrestrial plant growth - Modeling the interaction between nitrogen and phosphorus. New Phytologist. 2012; 194(4):953–960. doi: 10.1111/j.1469-8137.2012.04116.x.

Arnold A, Nikoloski Z. Bottom-up metabolic reconstruction of *Arabidopsis* and its application to determining the metabolic costs of enzyme production. Plant Physiology. 2014; 165(3):1380–1391. doi: 10.1104/pp.114.235358.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene Ontology: Tool for the unification of biology. Nature Genetics. 2000; 25(1):25–29. doi: 10.1038/75556.

Aubry S, Brown NJ, Hibberd JM. The role of proteins in C3 plants prior to their recruitment into the C4 pathway. Journal of Experimental Botany. 2011; 62(9):3049–3059. doi: 10.1093/jxb/err012.

Bailey S, Walters RG, Jansson S, Horton P. Acclimation of *Arabidopsis thaliana* to the light environment: The existence of separate low light and high light responses. Planta. 2001; 213(5):794–801. doi: 10.1007/s004250100556.

Bauwe H. Chapter 6 Photorespiration: The bridge to C4 photosynthesis. In: *C4 photosynthesis and related CO2 concentrating mechanisms* Springer Netherlands; 2010.p. 81–108. doi: 10.1007/978-90-481-9407-0_6.

Bellasio C, Griffiths H. The operation of two decarboxylases, transamination, and partitioning of C4 metabolic processes between mesophyll and bundle sheath cells allows light capture to be balanced for the maize C4 pathway. Plant Physiology. 2013; 164(1):466–480. doi: 10.1104/pp.113.228221.

Bellasio C, Lundgren MR. Anatomical constraints to C4 evolution: Light harvesting capacity in the bundle sheath. New Phytologist. 2016; 212(2):485–496. doi: 10.1111/nph.14063.

Berkemeyer M, Scheibe R, Ocheretina O. A novel non-redox-regulated NAD-dependent malate dehydrogenase from chloroplasts of *Arabidopsis thaliana* L. Journal of Biological Chemistry. 1998; 273(43):27927–27933. doi: 10.1074/jbc.273.43.27927.

Boyle NR, Morgan JA. Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. BMC Systems Biology. 2009; 3(1):4. doi: 10.1186/1752-0509-3-4.

Bräutigam A, Gowik U. Photorespiration connects C3 and C4 photosynthesis. Journal of Experimental Botany. 2016; 67(10):2953–2962. doi: 10.1093/jxb/erw056.

Bräutigam A, Schliesky S, Külahoglu C, Osborne CP, Weber APM. Towards an integrative model of C4 photosynthetic subtypes: Insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C4 species. Journal of Experimental Botany. 2014; 65(13):3579–3593. doi: 10.1093/jxb/eru100.

Bräutigam A, Schlüter U, Lundgren MR, Flachbart S, Ebenhöh O, Schönknecht G, Christin PA, Bleuler S, Droz JM, Osborne C, Weber A, Gowik U. Biochemical mechanisms driving rapid fluxes in C4 photosynthesis. . 2018; doi: 10.1101/387431.

Bräutigam A, Weber APM. Chapter 11 Transport Processes: Connecting the reactions of C4 photosynthesis. In: *C4 Photosynthesis and Related CO2 Concentrating Mechanisms* Springer Netherlands; 2010.p. 199–219. doi: 10.1007/978-90-481-9407-0_11.

Bush DR. Proton-coupled sugar and amino acid transporters in plants. Annual Review of Plant Physiology and Plant Molecular Biology. 1993; 44(1):513–542. doi: 10.1146/annurev.pp.44.060193.002501.

Cheung CYM, Poolman MG, Fell DA, Ratcliffe RG, Sweetlove LJ. 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. 2014; 165(2):917–929. doi: 10.1104/pp.113.234468.

Christin PA, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, Hartwell J, Osborne CP. Adaptive evolution of C4 photosynthesis through recurrent lateral gene transfer. Current Biology. 2012; 22(5):445–449. doi: 10.1016/j.cub.2012.01.054.

Döring F, Streubel M, Bräutigam A, Gowik U. Most photorespiratory genes are preferentially expressed in the bundle sheath cells of the C4 grass *Sorghum bicolor*. Journal of Experimental Botany. 2016; 67(10):3053–3064. doi: 10.1093/jxb/erw041.

Duff SMG, Rydel TJ, McClerren AL, Zhang W, Li JY, Sturman EJ, Halls C, Chen S, Zeng J, Peng J, Kretzler CN, Evdokimov A. The enzymology of alanine aminotransferase (AlaAT) isoforms from *Hordeum vulgare* and other organisms, and the HvAlaAT crystal structure. Archives of Biochemistry and Biophysics. 2012; 528(1):90–101. doi: 10.1016/j.abb.2012.06.006.

Edwards EJ, Osborne CP, Strömberg CAE, Smith SA, Bond WJ, Christin PA, Cousins AB, Duvall MR, Fox DL, Freckleton RP, Ghannoum O, Hartwell J, Huang Y, Janis CM, Keeley JE, Kellogg EA, Knapp AK, Leakey ADB, Nelson DM, Saarela JM, et al. The origins of C4 grasslands: Integrating evolutionary and ecosystem science. Science. 2010; 328(5978):587–591. doi: 10.1126/science.1177216.

Ellis RJ. The most abundant protein in the world. Trends in Biochemical Sciences. 1979; 4(11):241–244. doi: 10.1016/0968-0004(79)90212-3.

Feodorova TA, Voznesenskaya EV, Edwards GE, Roalson EH. biogeographic patterns of diversification and the origins of c4 in *Cleome* (Cleomaceae). Systematic Botany. 2010; 35(4):811–826. doi: 10.1600/036364410x539880.

Furumoto T, Yamaguchi T, Ohshima-Ichie Y, Nakamura M, Tsuchida-Iwata Y, Shimamura M, Ohnishi J, Hata S, Gowik U, Westhoff P, Bräutigam A, Weber APM, Izui K. A plastidial sodium-dependent pyruvate transporter. Nature. 2011; 476(7361):472–475. doi: 10.1038/nature10250.

Gietl C. Malate dehydrogenase isoenzymes: Cellular locations and role in the flow of metabolites between the cytoplasm and cell organelles. Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics. 1992; 1100(3):217–234. doi: 10.1016/0167-4838(92)90476-t.

Haberlandt GFJ. Physiologische Pflanzenanatomie. Verlag von W. Engelmann; 1904. doi: 10.5962/bhl.title.20550.

Hatch MD. C4 photosynthesis: A unique elend of modified biochemistry anatomy and ultrastructure. Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics. 1987; 895(2):81–106. doi: 10.1016/s0304-4173(87)80009-5.

Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ. Predicting C4 photosynthesis evolution: Modular, individually adaptive steps on a Mount Fuji Fitness Landscape. Cell. 2013; 153(7):1579–1588. doi: 10.1016/j.cell.2013.04.058.

Heldt HW, Piechulla B. Pflanzenbiochemie. Springer Berlin Heidelberg; 2015. doi: 10.1007/978-3-662-44398-9.

Josse EM, Simkin AJ, Gaffé J, Labouré AM, Kuntz M, Carol P. A plastid terminal oxidase associated with carotenoid desaturation during chromoplast differentiation. Plant Physiology. 2000; 123(4):1427–1436. doi: 10.1104/pp.123.4.1427.

Karabourniotis G, Bornman JF, Nikolopoulos D. A possible optical role of the bundle sheath extensions of the heterobaric leaves of *Vitis vinifera* and *Quercus coccifera*. Plant, Cell and Environment. 2000; 23(4):423–430. doi: 10.1046/j.1365-3040.2000.00558.x.

Kiba T, Feria-Bourrellier AB, Lafouge F, Lezhneva L, Boutet-Mercey S, Orsel M, Bréhaut V, Miller A, Daniel-Vedele F, Sakakibara H, Krapp A. The *Arabidopsis* nitrate transporter NRT2.4 plays a double role in roots and shoots of nitrogen-starved plants. The Plant Cell. 2012; 24(1):245–258. doi: 10.1105/tpc.111.092221.

Körner C. Paradigm shift in plant growth control. Current Opinion in Plant Biology. 2015; 25:107–114. doi: 10.1016/j.pbi.2015.05.003.

Lacher W. Physiological plant ecology: Ecophysiology and stress physiology of functional groups. Springer-Verlag Berlin Heidelberg; 2003. https://www.springer.com/de/book/9783540435167.

Leakey ADB. Photosynthesis, productivity, and yield of maize are not affected by open-air elevation of CO2 concentration in the absence of drought. Plant Physiology. 2006; 140(2):779–790. doi: 10.1104/pp.105.073957.

Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, König R, Smith RD, Palsson BØ. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Molecular Systems Biology. 2010; 6. doi: 10.1038/msb.2010.47.

Li L, Nelson CJ, Trösch J, Castleden I, Huang S, Millar AH. Protein degradation rate in *Arabidopsis thaliana* leaf growth and development. The Plant Cell. 2017; 29(2):207–228. doi: 10.1105/tpc.16.00768.

Liepman AH. Alanine aminotransferase homologs catalyze the glutamate:glyoxylate aminotransferase reaction in peroxisomes of *Arabidopsis*. Plant Physiology. 2003; 131(1):215–227. doi: 10.1104/pp.011460.

Linka N, Weber APM. Intracellular metabolite transporters in plants. Molecular Plant. 2010; 3(1):21–53. doi: 10.1093/mp/ssp108.

Liu H, Osborne CP. Water relations traits of C4 grasses depend on phylogenetic lineage, photosynthetic pathway, and habitat water availability. Journal of Experimental Botany. 2014; 66(3):761–773. doi: 10.1093/jxb/eru430.

Lundgren MR, Christin PA. Despite phylogenetic effects, C3–C4 lineages bridge the ecological gap to C4 photosynthesis. Journal of Experimental Botany. 2016; 68(2):241–254. doi: 10.1093/jxb/erw451.

Macduff JH, Bakken AK. Diurnal variation in uptake and xylem contents of inorganic and assimilated N under continuous and interrupted N supply to *Phleum pratense* and *Festuca pratensis*. Journal of Experimental Botany. 2003; 54(381):431–444. doi: 10.1093/jxb/erg058.

Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metabolic Engineering. 2003; 5(4):264–276. doi: 10.1016/j.ymben.2003.09.002.

Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber AP, Westhoff P, Gowik U. The role of photorespiration during the evolution of C4 photosynthesis in the genus *Flaveria*. eLife. 2014; 3. doi: 10.7554/elife.02478.

Miller MAE, O'Cualain R, Selley J, Knight D, Karim MF, Hubbard SJ, Johnson GN. dynamic acclimation to high light in *Arabidopsis thaliana* involves widespread reengineering of the leaf proteome. Frontiers in Plant Science. 2017; 8. doi: 10.3389/fpls.2017.01239.

Monson RK. The origins of C4 genes and evolutionary pattern in the C4 metabolic phenotype. In: *C4 Plant Biology* Elsevier; 1999.p. 377–410. doi: 10.1016/b978-012614440-6/50012-4.

Mueller LA, Zhang P, Rhee SY. AraCyc: A biochemical pathway database for *Arabidopsis*. Plant Physiology. 2003; 132(2):453–460. doi: 10.1104/pp.102.017236.

Neuhaus HE. Transport of primary metabolites across the plant vacuolar membrane. FEBS Letters. 2007; 581(12):2223–2226. doi: 10.1016/j.febslet.2007.02.003.

Ogren WL, Bowes G. Ribulose diphosphate carboxylase regulates soybean photorespiration. Nature New Biology. 1971; 230(13):159–160. doi: 10.1038/newbio230159a0.

de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. AraGEM a genome-scale reconstruction of the primary metabolic network in *Arabidopsis*. Plant Physiology. 2010; 152(2):579–589. doi: 10.1104/pp.109.148817.

de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. AraGEM a genome-scale reconstruction of the primary metabolic network in *Arabidopsis*. Plant Physiology. 2010; 152(2):579–589. doi: 10.1104/pp.109.148817.

de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. C4GEM a genome-scale metabolic model to study C4 plant metabolism. Plant Physiology. 2010; 154(4):1871–1885. doi: 10.1104/pp.110.166488.

de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Nielsen LK. AlgaGEM – a genome-scale metabolic reconstruction of algae based on the *Chlamydomonas reinhardtii* genome. BMC Genomics. 2011; 12(Suppl 4):S5. doi: 10.1186/1471-2164-12-s4-s5.

Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nature Biotechnology. 2010; 28(3):245–248. doi: 10.1038/nbt.1614.

Osborne CP, Freckleton RP. Ecological selection pressures for C4 photosynthesis in the grasses. Proceedings of the Royal Society B: Biological Sciences. 2009; 276(1663):1753–1760. doi: 10.1098/rspb.2008.1762.

Pál C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD. Chance and necessity in the evolution of minimal metabolic networks. Nature. 2006; 440(7084):667–670. doi: 10.1038/nature04568.

Papp B, Szappanos B, Notebaart RA. Use of genome-scale metabolic models in evolutionary systems biology. In: *Methods in Molecular Biology* Humana Press; 2011.p. 483–497. doi: 10.1007/978-1-61779-173-4_27.

Petersen J, Förster K, Turina P, Gräber P. Comparison of the H^+/ATP ratios of the H^+-ATP synthases from yeast and from chloroplast. Proceedings of the National Academy of Sciences. 2012; 109(28):11150–11155. doi: 10.1073/pnas.1202799109.

Pick TR, Bräutigam A, Schlüter U, Denton AK, Colmsee C, Scholz U, Fahnenstich H, Pieruschka R, Rascher U, Sonnewald U, Weber APM. Systems analysis of a maize leaf developmental gradient redefines the current C4 model and provides candidates for regulation. The Plant Cell. 2011; 23(12):4208–4220. doi: 10.1105/tpc.111.090324.

Pinto H, Powell JR, Sharwood RE, Tissue DT, Ghannoum O. Variations in nitrogen use efficiency reflect the biochemical subtype while variations in water use efficiency reflect the evolutionary lineage of C4 grasses at inter-glacial CO2. Plant, Cell & Environment. 2015; 39(3):514–526. doi: 10.1111/pce.12636.

Poolman MG, Kundu S, Shaw R, Fell DA. Responses to light intensity in a genome-scale model of rice metabolism. Plant Physiology. 2013; 162(2):1060–1072. doi: 10.1104/pp.113.216762.

Reiser J. molecular physiological analysis of the two plastidic ATP/ADP transporters from *Arabidopsis*. Plant Physiology. 2004; 136(3):3524–3536. doi: 10.1104/pp.104.049502.

Rozema J. Plant responses to atmospheric carbon dioxide enrichment: Interactions with some soil and atmospheric conditions. Vegetatio. 1993; 104-105(1):173–190. doi: 10.1007/bf00048152.

Sage RF, Pearcy RW, Seemann JR. The nitrogen use efficiency of C3 and C4 plants : III. leaf nitrogen effects on the activity of carboxylating enzymes in *Chenopodium album* (L.) and *Amaranthus retroflexus* (L.). Plant Physiology. 1987; 85(2):355–359. doi: 10.1104/pp.85.2.355.

Sage RF. The evolution of C4 photosynthesis. New Phytologist. 2004; 161(2):341–370. doi: 10.1111/j.1469-8137.2004.00974.x.

Sage RF, Christin PA, Edwards EJ. The C4 plant lineages of planet Earth. Journal of Experimental Botany. 2011; 62(9):3155–3169. doi: 10.1093/jxb/err048.

Sage RF, Sage TL, Kocacinar F. Photorespiration and the evolution of C4 photosynthesis. Annual Review of Plant Biology. 2012; 63(1):19–47. doi: 10.1146/annurev-arplant-042811-105511.

Saha R, Suthers PF, Maranas CD. *Zea mays i*RS1563: A comprehensive genome-scale metabolic reconstruction of maize metabolism. PLoS ONE. 2011; 6(7):e21784. doi: 10.1371/journal.pone.0021784.

Scheben A, Wolter F, Batley J, Puchta H, Edwards D. Towards CRISPR/Cas crops - Bringing together genomics and genome editing. New Phytologist. 2017; 216(3):682–698. doi: 10.1111/nph.14702.

Schlüter U, Bräutigam A, Gowik U, Melzer M, Christin PA, Kurz S, Mettler-Altmann T, Weber AP. Photosynthesis in C3–C4 intermediate *Moricandia* species. Journal of Experimental Botany. 2016; 68(2):191–206. doi: 10.1093/jxb/erw391.

Schlüter U, Denton AK, Bräutigam A. Understanding metabolite transport and metabolism in C4 plants through RNA-seq. Current Opinion in Plant Biology. 2016; 31:83–90. doi: 10.1016/j.pbi.2016.03.007.

Schomburg I, Chang A, Hofmann O, Ebeling C, Ehrentreich F, Schomburg D. BRENDA: A resource for enzyme data and metabolic information. Trends in Biochemical Sciences. 2002; 27(1):54–56. doi: 10.1016/s0968-0004(01)02027-8.

Schultz CJ, Coruzzi GM. The aspartate aminotransferase gene family of *Arabidopsis* encodes isoenzymes localized to three distinct subcellular compartments. The Plant Journal. 1995; 7(1):61–75. doi: 10.1046/j.1365-313x.1995.07010061.x.

Shikanai T. Regulatory network of proton motive force: Contribution of cyclic electron transport around photosystem I. Photosynthesis Research. 2016; 129(3):253–260. doi: 10.1007/s11120-016-0227-0.

Simons M, Misra A, Sriram G. Genome-scale models of plant metabolism. In: *Methods in Molecular Biology* Humana Press; 2013.p. 213–230. doi: 10.1007/978-1-62703-661-0_13.

Sonawane BV, Sharwood RE, Whitney S, Ghannoum O. Shade compromises the photosynthetic efficiency of NADP-ME less than that of PEP-CK and NAD-ME C4 grasses. Journal of Experimental Botany. 2018; 69(12):3053–3068. doi: 10.1093/jxb/ery129.

Spreitzer RJ, Salvucci ME. RUBISCO: Structure regulatory interactions, and possibilities for a better enzyme. Annual Review of Plant Biology. 2002; 53(1):449–475. doi: 10.1146/annurev.arplant.53.100301.135233.

Stitt M, Zeeman SC. Starch turnover: Pathways regulation and role in growth. Current Opinion in Plant Biology. 2012; 15(3):282–292. doi: 10.1016/j.pbi.2012.03.016.

Sun J, Okita TW, Edwards GE. modification of carbon partitioning, photosynthetic capacity, and O2 sensitivity in *Arabidopsis* plants with low adp-glucose pyrophosphorylase activity. Plant Physiology. 1999; 119(1):267–276. doi: 10.1104/pp.119.1.267.

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E. The *Arabidopsis* Information Resource (TAIR): Gene structure and function annotation. Nucleic Acids Research. 2007; 36(Database):D1009–D1014. doi: 10.1093/nar/gkm965.

Turgeon R. The sink-source transition in leaves. Annual Review of Plant Physiology and Plant Molecular Biology. 1989; 40(1):119–138. doi: 10.1146/annurev.pp.40.060189.001003.

Turina P, Petersen J, Gräber P. Thermodynamics of proton transport coupled ATP synthesis. Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics. 2016; 1857(6):653–664. doi: 10.1016/j.bbabio.2016.02.019.

Ueno O, Sentoku N. Comparison of leaf structure and photosynthetic characteristics of C3 and C4 *Alloteropsis semialata* subspecies. Plant, Cell and Environment. 2006; 29(2):257–268. doi: 10.1111/j.1365-3040.2005.01418.x.

Vishwakarma A, Tetali SD, Selinski J, Scheibe R, Padmasree K. Importance of the alternative oxidase (AOX) pathway in regulating cellular redox and ROS homeostasis to optimize photosynthesis during restriction of the cytochrome oxidase pathway in *Arabidopsis thaliana*. Annals of Botany. 2015; 116(4):555–569. doi: 10.1093/aob/mcv122.

Voznesenskaya EV, Koteyeva NK, Chuong SDX, Ivanova AN, Barroca J, Craven LA, Edwards GE. Physiological, anatomical and biochemical characterisation of photosynthetic types in genus *Cleome* (Cleomaceae). Functional Plant Biology. 2007; 34(4):247. doi: 10.1071/fp06287.

Wang Y, Bräutigam A, Weber APM, Zhu XG. Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. Journal of Experimental Botany. 2014; 65(13):3567–3578. doi: 10.1093/jxb/eru058.

Weissmann S, Ma F, Furuyama K, Gierse J, Berg H, Shao Y, Taniguchi M, Allen DK, Brutnell TP. Interactions of C4 subtype metabolic activities and transport in maize are revealed through the characterization of DCT2 mutants. The Plant Cell. 2016; 28(2):466–484. doi: 10.1105/tpc.15.00497.

Wilkinson TL, Douglas AE. Phloem amino acids and the host plant range of the polyphagous aphid, *Aphis fabae*. Entomologia Experimentalis et Applicata. 2003; 106(2):103–113. doi: 10.1046/j.1570-7458.2003.00014.x.

Williams BP, Johnston IG, Covshoff S, Hibberd JM. Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. eLife. 2013; 2. doi: 10.7554/elife.00961.

Wingler A, Walker RP, Chen ZH, Leegood RC. Phosphoenolpyruvate carboxykinase is involved in the decarboxylation of aspartate in the bundle sheath of maize. Plant Physiology. 1999; 120(2):539–546. doi: 10.1104/pp.120.2.539.

Yamamoto H, Peng L, Fukao Y, Shikanai T. An src homology 3 domain-like fold protein forms a ferredoxin binding site for the chloroplast NADH dehydrogenase-like complex in *Arabidopsis*. The Plant Cell. 2011; 23(4):1480–1493. doi: 10.1105/tpc.110.080291.

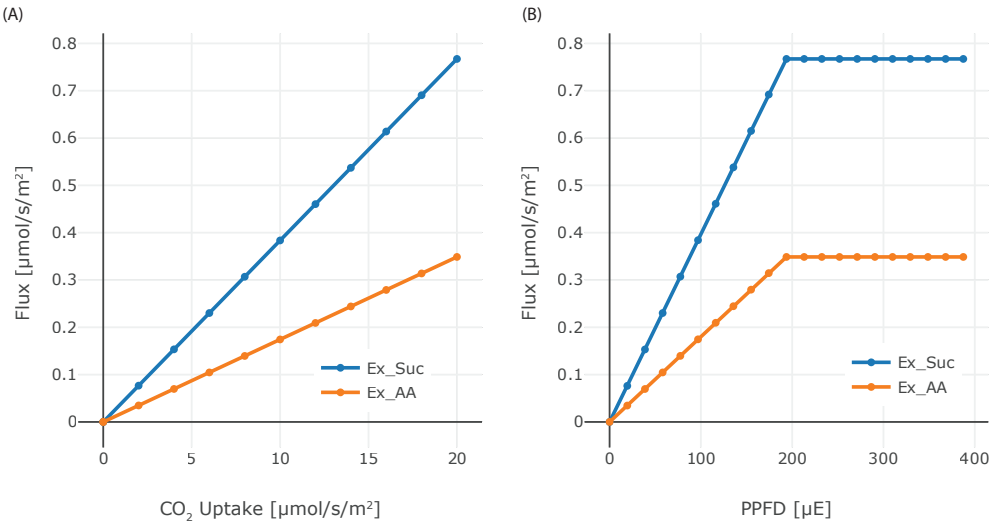


Figure 1–Figure supplement 1. Effect of CO_2 and PPFD variation. (A) Dependence of the phloem output on CO_2 input flux in the range 0 μmol/(m^2 s)–20 μmol/(m^2 s), (B) Dependence of phloem output on the PPFD in the range 0 μE–400 μE. Sucrose and starch are produced in the same amounts, each of them consists of 12 C-atoms.

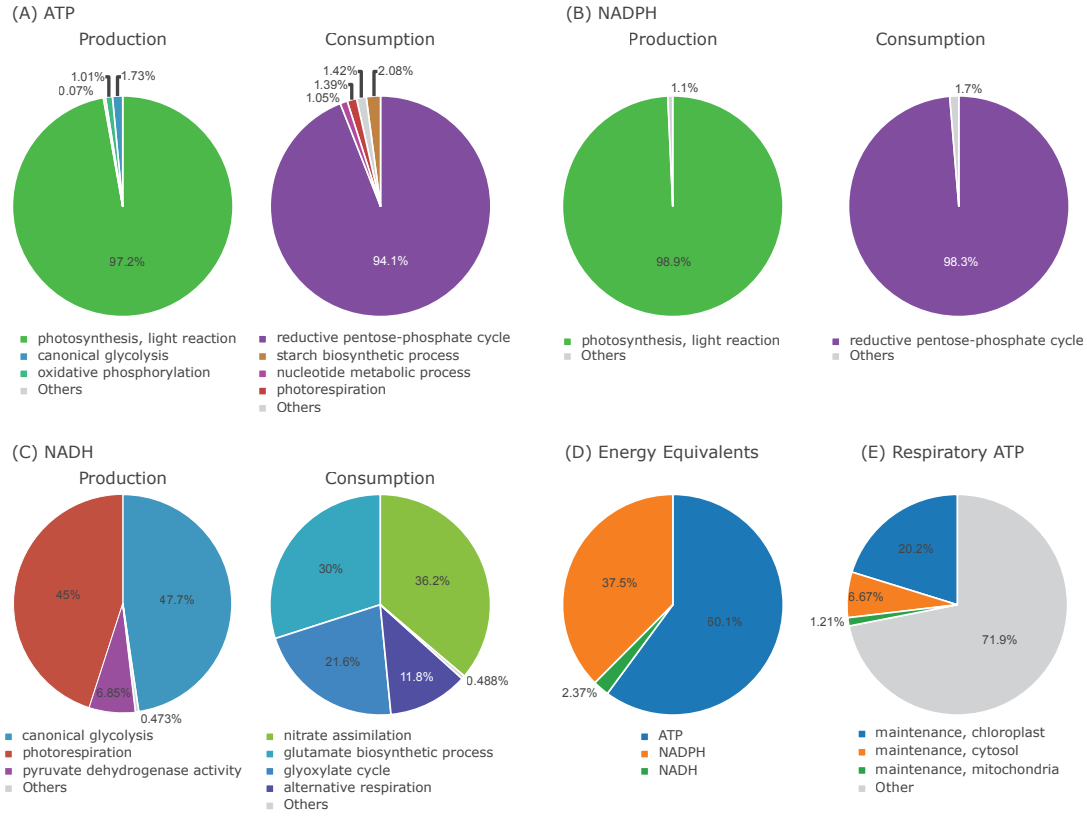


Figure 1–Figure supplement 2. Energy Flux Distribution in the *one-cell* Model. (A) ATP production and consumption, (B) NADPH production and consumption, (C) NADH production and consumption, (D) proportion of ATP, NADPH, NADH used as energy equivalent, (E) proportion of respiratory ATP used for maintenance

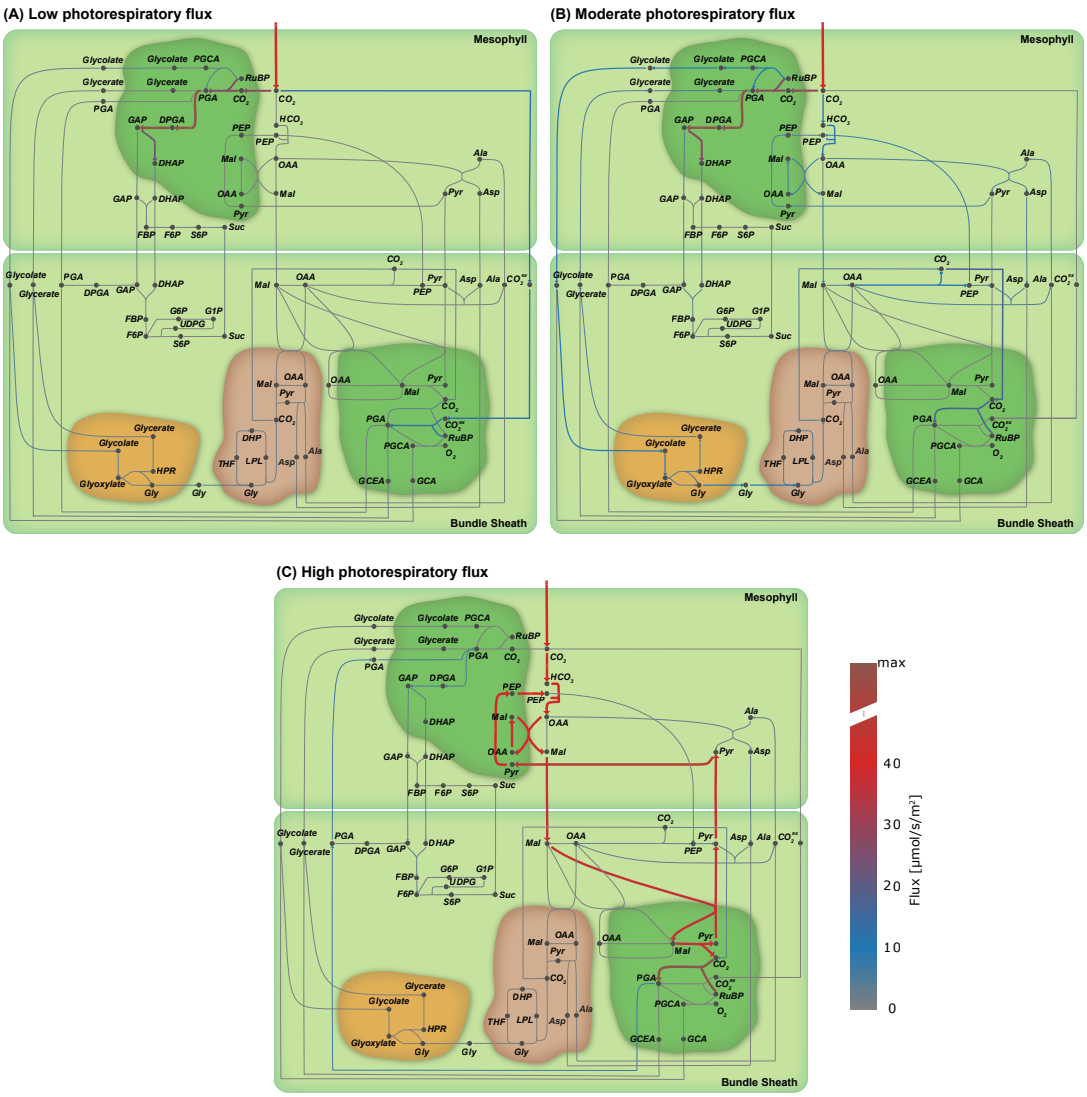


Figure 3–Figure supplement 1. Flux maps illustrating the effect of the proportion of photorespiratory flux through Rubisco. (A) Low photorespiratory flux; (B) Moderate photorespiratory flux; and (C) High photorespiratory flux. (Arc width and colour are set relative to flux values in flux, grey arcs – no flux)

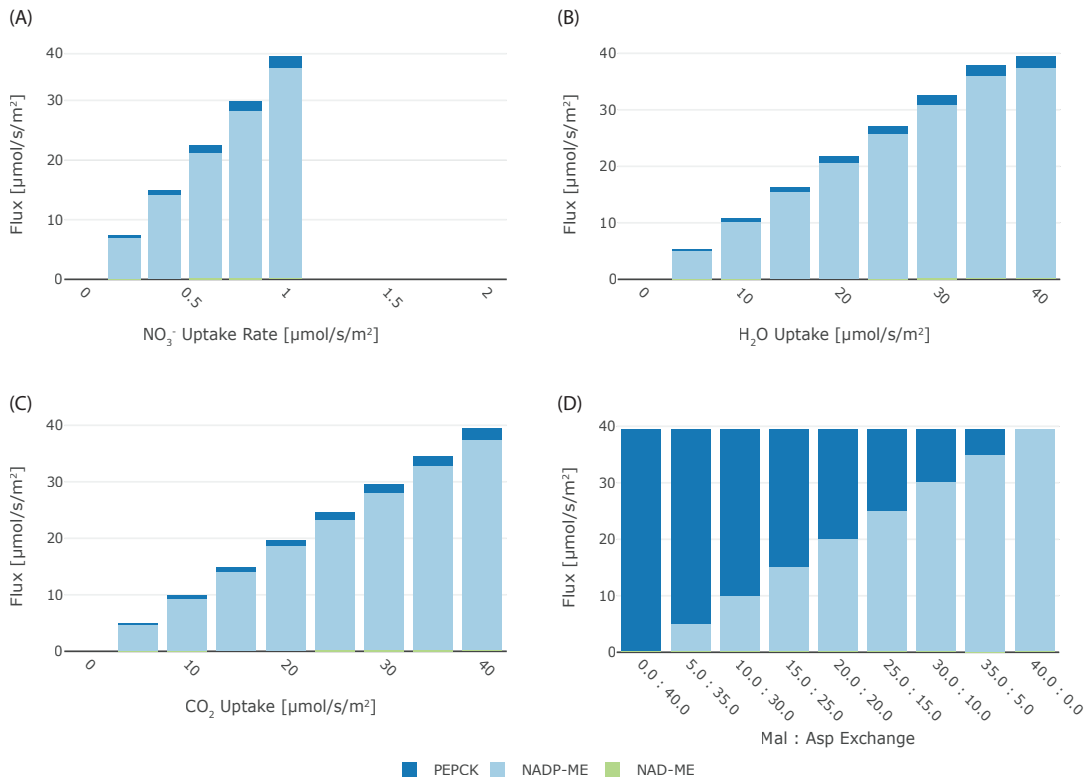


Figure 6–Figure supplement 1. Effect of (A) NO_3^-, (B) H_2O, and (C) CO_2 limitation, as well as (D) the malate : aspartate transport ratio on the C4 state.