Resource allocation modeling for autonomous prediction of plant cell phenotypes

Anne Goelzer, Loïc Rajjou, Fabien Chardon, Olivier Loudet, Vincent Fromion

Title: Resource allocation modeling for autonomous prediction of plant cell phenotypes

Authors: Anne Goelzer[1,*], Loïc Rajjou[2], Fabien Chardon[2], Olivier Loudet[2], Vincent Fromion[1,*]

[1]Université Paris-Saclay, INRAE, MaIAGE, 78350 Jouy-en-Josas, France
[2]Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), 78000, Versailles, France

*corresponding authors: anne.goelzer@inrae.fr. vincent.fromion@inrae.fr.

This is the accepted version of the manuscript https://doi.org/10.1016/j.ymben.2024.03.009. It is distributed under the CC-BY-ND-NC Creative Common license.

Please cite the published paper in your works:

Goelzer et al. Metabolic Engineering. 2024. https://doi.org/10.1016/j.ymben.2024.03.009

Abstract

Predicting the plant cell response in complex environmental conditions is a challenge in plant biology. Here we developed a resource allocation model of cellular and molecular scale for the leaf photosynthetic cell of *Arabidopsis thaliana*, based on the Resource Balance Analysis (RBA) constraint-based modeling framework. The RBA model contains the metabolic network and the major macromolecular processes involved in the plant cell growth and survival and localized in cellular compartments. We simulated the model for varying environmental conditions of temperature, irradiance, partial pressure of CO_2 and O_2, and compared RBA predictions to known resource distributions and quantitative phenotypic traits such as the relative growth rate, the C:N ratio, and finally to the empirical characteristics of CO_2 fixation given by the well-established Farquhar model. In comparison to other standard constraint-based modeling methods like Flux Balance Analysis, the RBA model makes accurate quantitative predictions without the need for empirical constraints. Altogether, we show that RBA significantly improves the autonomous prediction of plant cell phenotypes in complex environmental conditions, and provides mechanistic links between the genotype and the phenotype of the plant cell.

Keywords

Resource Balance Analysis; Constraint-based modeling; Genome-scale model; Farquhar model; Plant Systems Biology; *Arabidopsis thaliana*;

1. Introduction

Climate change, the scarcity of certain natural resources, the need to reduce crop inputs, increase biological diversity in agroecosystems, or change cropping practices have resulted in a greater variety and complexity of situations to be managed by agronomists. To assist them, agronomists need plant modeling with extended predictive ability to manage complex environmental conditions of farming systems. Usually, agronomists use individual plant models typically based on a systemic description of plant functioning such as ecophysiological models (Tuzet et al., 2003) or functional-structural plant models (FSPMs) that manage the production and distribution of mass (carbon or nitrogen) among plant organs (Chew et al., 2014; de Reffye et al., 2021). These models result in accurate predictions in controlled environments, especially for the exchanges of biological material (i.e. water, biomass, carbon or nitrogen depending on models) between organs. However, they usually fail to cope with complex -though realistic- environmental conditions (Yin et al. 2021), including other nutrient limiting conditions (such as phosphate or sulfur limitations) or combined environmental constraints (stresses). Indeed, the plants adopt complex strategies based on the integration of cell decisions to cope with the effect of multiple stresses (Zandalinas et al. 2021). In fact, individual plant models result from a trade-off between predictive capability and simplicity, which hides the intrinsic complexity of cellular functioning and makes it harder to predict a plant's responses to new challenging environmental conditions. Improving the modeling of cellular scales could thus significantly increase the predictive capabilities of individual plant models, which is made possible today by using the vast amount of existing knowledge about plant physiology (Bouchabke, et al. 2008, Tisné et al. 2013, Ikram, S., et al. 2012) as well as about the molecular mechanisms involved in the plant response to a wide range of environmental factors (Schulz, E., et al. 2021, Henriet, C., et al. 2021 , Zandalinas et al. 2021).

Plant cell modeling has been partly achieved through the use of constraint-based models. Constraint-based modeling -CBM- describes the cell by a set of biophysical, biochemical and structural constraints. From an engineering perspective, these functional constraints are

mathematically formalized as a set of linear equalities and inequalities that define a feasibility domain for cellular states, which describe all possible cell phenotypes. Satisfying these constraints is necessary but not enough to accurately predict a cellular state. To select a single state within this set of feasible states, modelers typically apply general optimality principles, such as the maximization of biomass production (Feist et al. 2010). This optimality principle is formalized as a linear objective function, which, along with the linear equalities and inequalities, leads to linear optimization methods (Nesterov et al. 1994; Nesterov 2012). The most famous and widely used CBM method is called Flux Balance Analysis (FBA) (Varma et al. 1994). FBA describes the cell functioning through (i) the metabolic network, and (ii) a chemical reaction describing the biomass formation with a fixed average biomass composition. The metabolic network is described by equalities on metabolic fluxes, each one representing the mass conservation in steady-state for each metabolite, i.e. the sum of the fluxes that produces a metabolite is equal to the sum of the fluxes that consumes this metabolite. FBA is thus based on the stoichiometry of chemical reactions only. FBA models have already been developed to predict plant cell phenotypes for different plant species, that is, in the context of FBA, the metabolic flux distribution (see Poolman, M.G., et al. 2009; and de Oliveira Dal'Molin et al. 2010; for the first FBA models of *Arabidopsis thaliana*). For Arabidopsis, existing FBA plant cell models differ from each other by the number of metabolites, metabolic reactions, or compartments and by their application: autotrophic or heterotrophic regimen for leaf (Poolman et al. 2009; de Oliveira Dal'Molin et al. 2010; Arnold et al. 2014), day/night alternation for guard cells (Tan et al. 2020), or for multi-organs by considering one FBA model per organ (Grafahrend-Belau, E., et al. 2013; de Oliveira Dal'Molin et al. 2015; Shaw, R., and Cheung, C. M. 2018). These models undoubtedly helped to understand the metabolic flux distribution of plant cells in various conditions (see for instance Cheung, et al. 2013, Cheung, et al. 2014; Töpfer et al. 2020).

To predict flux distributions within each organ and between plant organs by using FBA, one needs to define (i) an appropriate objective function (e.g. minimization of photon influx for a fixed relative growth rate or maximization of the relative growth rate) and (ii) a set of constraints on the fluxes exchanged between the organs, their local environment, and other organs (de Oliveira et al. 2015, Gerlin et al. 2022). This means that the whole-plant configuration (or the plant cell configuration) is imposed, instead of being autonomously predicted by the model. Beyond the choice of an adequate objective function, the quality of FBA predictions is very sensitive to the constraints applied on exchange fluxes. This dependency – and hence the impossibility to obtain precise predictions without knowing the exchange fluxes – holds across organisms, and is well known in the CBM community. Even for bacterial cells, the organisms on which CBM was originally developed, a correct quantitative prediction of the growth rate depends crucially on imposed exchange fluxes. This has led to numerous CBM approaches (reviewed in Goelzer et al. 2017) for improving FBA, and, among them, CBM methods integrating structural constraints other than mass conservation for describing the cell functioning are particularly relevant.

To overcome FBA limitations, we developed a new CBM method called Resource Balance Analysis (RBA; Goelzer et al. 2009, Goelzer et al. 2011). RBA considers the cell as an autonomous system that has to determine whether the resources available in the environment are sufficient to ensure growth, or survival. At the cellular level, an enzyme needs to be sufficiently abundant to produce its metabolic flux; the metabolic network needs to provide sufficient fluxes of building blocks and energy to ensure the production of all enzymes, ribosomes and other cellular constituents; there need to be enough ribosomes to build all cellular proteins; and all these components need to fit into a limited cellular space. All this results in a global trade-off with respect to the use of resources by cellular processes in the entire cell. For instance, let us consider a cell environment in which no amino acids but only nitrate is available. In this case, the cell needs to synthetize the

amino acids *de novo*. This synthesis consumes or binds resources in terms of building blocks and energy, of ribosomes and chaperones for enzyme synthesis involved in de novo amino acid synthesis pathways, and of cellular space occupied by enzymes. In contrast, if all amino acids are already present in the medium, these resources are saved and can be invested in other cellular functions, especially in the translation apparatus and in metabolic pathways to increase growth rate. At a given growth rate, three main sets of cellular constraints governing resource allocation between cellular processes can be distinguished (Goelzer et al. 2009, Goelzer et al. 2011, Scott et al. 2010) and are included in RBA:

(C1) the mass balance is conserved;
(C2) the capacity of molecular machines (such as ribosomes or enzymes) is sufficient to fulfill their functions;
(C3) the molecular machines occupy the limited space available in all cellular compartments and their membranes.

Satisfying these constraints at a given growth rate leads to a linear convex optimization problem. This problem is tractable and solvable rapidly even at genome-scale in a matter of seconds (Bulovic et al. 2019). RBA was originally developed and experimentally validated on bacterial cells (Goelzer et al. 2015) and predicts the set of possible cellular configurations (growth rate, metabolic fluxes, abundances of molecular machines, including ribosomes, enzymes, transporters…) compatible with the available external resources (i.e. the extracellular concentration of nutrients), for a given environmental condition. Importantly, RBA takes the extracellular nutrient concentrations as model inputs, and does not need constraints on exchange fluxes to obtain quantitative predictions. The cellular configuration is not imposed by constraints on exchange fluxes, but instead results from a parsimonious allocation of resources to different cellular processes. This makes RBA more autonomous than FBA in terms of predictive capability especially in complex environmental conditions (Tournier et al. 2017).

In this paper, we present a RBA model for the rosette leaves of *Arabidopsis thaliana* (accession Col-0, the reference genotype of the Arabidopsis community), based on the RBA framework for eukaryotic cells (Goelzer et al. 2019). The mathematical framework of RBA for eukaryotic cells resembles the one for prokaryotic cells, except for how it treats the maximal available space of each cellular compartment (which is predefined in the case of prokaryotic cells, and treated as a free model variable, and therefore predicted, in eukaryotic cells; see Methods). The RBA model proposed in this paper describes the functioning of a leaf cell fixing CO_2 during vegetative growth, which contains a complete photosynthetic apparatus (including the RuBisCO, the most abundant protein on earth) in the chloroplast and thus concentrates a large part of the rosette resources of *Arabidopsis thaliana*. We specifically investigated if the cellular configuration might be parsimonious in terms of resource use for complex environmental conditions. For this purpose, we predicted the rosette cellular configuration that maximizes the relative growth rate for varying temperature, irradiance, partial pressure of CO_2 and O_2. We compared RBA predictions to a known resource distribution (Pyl et al. 2012), to a known flux distribution (Ma et al. 2014), to known quantitative phenotypic traits such as the C:N ratio (Chietera et al. 2018), and to the empirical characteristics of CO_2 fixation given by the well-established Farquhar model (Farquhar et al. 1980). We find that the RBA model provides accurate quantitative predictions when compared to existing experimental data. Finally, the sensitivity of this model to parameter variations is explored. Altogether, our results show that RBA is a significant step forward in terms of autonomous predictions of plant cell phenotypes in complex environmental conditions.

2. Materials and Methods

2.1 RBA framework for eukaryotic cells

In our model, an eukaryotic cell is composed of N_c subcellular compartments, each of them occupying a relative subcellular volume f_i (and noting f the vector of $(f_1,...,f_{N_c})^T$) and of the following molecular species:

- N_y molecular machines, which can be subdivided in N_m enzymes and transporters involved in the metabolic network (i.e. enzymes, transporters) at the concentrations $E = (E_1,...,E_{N_m})^T$ and with the fluxes $\nu = (\nu_1,...,\nu_{N_m})^T$; and N_p macromolecular machines involved in non-metabolic cellular processes such as the translation apparatus, at the concentrations $M = (M_1,...,M_{N_p})^T$;
- N_g proteins for which the cellular process to which the proteins belong is not specified and $P_G = (P_{G_1},...,P_{G_{Ng}})^T$ denotes the corresponding set of protein concentrations;
- N_s metabolites at the concentrations $S = (S_1,...,S_{N_s})^T$. Moreover, we distinguished a subset of $N_{\bar{b}}$target metabolites with fixed concentrations $\bar{B} = (\bar{B}_1,...,\bar{B}_{N_{\bar{b}}})^T$ and a subset of $N_{\hat{b}}$ metabolites involved in the building of organelles and with fixed concentrations $\hat{B} = (\hat{B}_1,...,\hat{B}_{N_{\hat{b}}})^T$ per unit of surface.

For this set of molecule species, the problem of resource allocation between cell functions -i.e. the RBA optimization problem - is formally stated as a set of linear inequalities and equalities to be satisfied, which together form a convex feasibility problem at fixed relative growth rate $\mu \geq 0$- i.e. the amount of produced dry biomass per cell per day. More precisely, the constraints that molecular species must fulfill are:

(C1). the metabolic network has to produce all metabolic precursors necessary for biomass production and mass balance must be satisfied for all molecule species;

(C2). the capacity of all molecular machines must be sufficient to ensure their function, i.e. to catalyze chemical conversions at a sufficient rate;

(C3). the intracellular space of compartments and the occupancy of membranes are limited.

The problem of parsimonious resource allocation between cellular functions in eukaryotic cells is formalized mathematically as follows (Goelzer and Fromion, 2019). For a fixed vector of concentrations $P_G \geq 0$

$$
\begin{aligned}
&\underset{Y\in\mathbb{R}^{N_y}_{\geq 0},\nu\in\mathbb{R}^{N_m},f\in\mathbb{R}^{N_c}_{\geq 0}}{\text{maximize}} && \mu \in \mathbb{R}_{\geq 0},\\
&(C_1) && -\Omega\nu + \mu(C^S_Y Y + C^S_G P_G + C^S_B \bar{B} + C^S_F f\hat{B}) = 0\\
&(C_2) && \mu(C^M_Y Y + C^M_G P_G) - K_T Y \leq 0\\
& && -K^{'}_E Y \leq \nu \leq K_E Y\\
&(C_3) && C^D_Y Y + C^D_G P_G - C^D_F f \leq 0\\
& && C^F_F f - \bar{C} = 0
\end{aligned}
$$

where the decision variables are the vector f, the flux vector ν, the vector of concentration of molecular machines $Y = (E^T, M^T)^T$.

Moreover:

- Ω is the stoichiometry matrix of the metabolic network of size $N_s \times N_m$, where Ω_{ij} corresponds to the stoichiometry of metabolite S_i in the j-th enzymatic reaction;
- C^S_Y (resp. C^S_G) is a $N_s \times N_y$ (resp. $N_s \times N_g$) matrix where each coefficient $C^S_{Y_{ij}}$ (resp. $C^S_{G_{ij}}$) corresponds to the number of metabolite S_i consumed (or produced) for the synthesis of one machine Y_j (resp. P_{G_j}); $C^S_{Y_{ij}}$ (resp. $C^S_{G_{ij}}$) is then positive, negative or null if S_i is produced, consumed or not involved in the synthesis of one machine Y_j (resp. P_{G_j}). ;
- C^S_B (resp. C^S_F) is a $N_s \times N_{\bar{b}}$ matrix (resp. $N_s \times N_{\hat{b}}$) where each coefficient $C^S_{B_{ij}}$ (resp. $C^S_{F_{ij}}$) corresponds to the number of metabolite S_i consumed (or produced) for the synthesis of one $\bar{B}_j$ (resp. one $\hat{B}_j$);
- K_T (K_E and $K^{'}_E$, respectively) of size $N_p \times N_y$ ($N_m \times N_y$, respectively) is a diagonal matrix where each coefficient K_{T_i} (K_{E_i} and $K^{'}_{E_i}$, respectively) is positive and corresponds to the efficiency of the molecular machine Y_i, i.e. the rate of the process per amount of the catalyzing molecular machine, (resp. the efficiency of the enzyme E_i in forward and backward sense, respectively);
- C^M_Y (resp. C^M_G) is a $N_p \times N_y$ (resp. $N_p \times N_g$) matrix where each coefficient $C^M_{Y_{ij}}$ typically corresponds to the length in amino acids of the machine Y_j (resp. P_{G_j}). In some cases (for instance for the constraints on protein chaperoning), the length in amino acids can be multiplied by a coefficient, such as the fraction of the whole proteome that necessitates chaperoning;
- C^D_F is a $N_c \times N_c$ diagonal matrix such as $C^D_F = d_T I_{N_c}$ where I_{N_c} is the identity matrix of size N_c, and d_T is the total density of the cell, expressed in equivalent amino acid residues per gram of cell dry weight as in the RBA framework for prokaryotic cells;

- C_Y^D (resp. C_G^D) is a $N_c \times N_y$ (resp. $N_c \times N_g$) matrix where each coefficient $C_{Y_{ij}}^D$ (resp. $C_{G_{ij}}^D$) corresponds to the density of one machine Y_j (resp. P_{G_j}) in the i-th compartment. By construction, we have one unique localization per machine. Since densities are expressed as a number of amino-acid residues per gram of cell dry weight, $C_{Y_{ij}}^D$ and $C_{G_{ij}}^D$ correspond in most of the cases to the number of amino acids residues per protein, except for ribosomes, where the density of nucleotides residues of rRNAs are converted to an equivalent density of amino acids as proposed in Marr 1991.
- The last equality constraint can contain additional constraints between volumes or surfaces for complex structures of compartments. In this paper, we only consider the equality constraint where all relative cellular compartment volumes or surfaces must sum to one.

Compared to the RBA framework for prokaryotic cells (Goelzer et al. 2009; Goelzer et al. 2011), the eukaryotic RBA framework introduces the volume fractions *f* as new decision variables. This means that the volume fraction of the cell occupied by organelles could theoretically be predicted. However, this would require coefficients C_Y^D and C_G^D that convert each molecular machine concentration into its volume. Since these coefficients are currently unknown, we chose instead to use the limitation of the cellular density (expressed in equivalent millimoles of amino acid residues per gram of cell dry weight) as a proxy of the limitation of the cellular volume, like for prokaryotic cells. Coefficients C_Y^D and C_G^D thus correspond to the densities of molecular machines (see above and Goelzer et al. 2011; Goelzer et al. 2015). Therefore, in a first approximation, the variables f_i can be interpreted as the fractions of total protein allocated to each cellular compartment (see Fig.2A).

Including environmental conditions into the model. Environmental conditions are included through the concentration of nutrients at the cell boundary - i.e. in xylem vessels-, partial pressure of CO_2, O_2 and temperature. Concentrations of nutrients are accounted for in RBA through Michaelis-Menten type efficiencies for transporters located in the plasma membrane. For extracellular nutrient S_{ext}, the efficiency of the i-th transporter is equal to $K_{E_i} = \frac{k_{max_i} S_{ext}}{K_{m_i} + S_{ext}}$. Temperature and the partial pressures of CO_2 and O_2 are included in RBA through the carboxylase and oxygenase activities of RuBisCO as in the Farquhar model (Caemmerer, S., et al. (2009), see below). Enzyme and transporter efficiencies that depend on extracellular conditions are evaluated once during the initialization of the numerical resolution, and used in constraints (C2).

Modeling RuBisCO and PS2 efficiencies. RuBisCO efficiency is decomposed into oxygenase ($K_{RuBisCO,o} = \frac{k_{o\,max} O}{O + K_o(1+C/K_c)}$) and carboxylase ($K_{RuBisCO,c} = \frac{k_{c\,max} C}{C + K_c(1+O/K_o)}$) nonlinear activities for given extracellular partial pressure of CO_2 (-C-) and O_2 (-O-), where $k_{c\,max}$, $k_{o\,max}$, K_c and K_o are the kinetics parameters of RuBisCO and are temperature-dependent as in the Farquhar model (see below). Moreover, the thermostability of the RuBisCO activase, - the chaperone responsible for RuBisCO folding-, is known to be a limiting factor in plant productivity under heat stress (Kurek et al. 2007). We modeled this effect like in the Farquhar model (Caemmerer, S., et al. (2009)) by a multiplicative term $a_R = 1/(1 + e^{0.3(T-T_{sat})})$ to $K_{RuBisCO,o}$ and $K_{RuBisCO,c}$. The term a_R describes

the fraction of RuBisCO that is active, T is the temperature, and T_{sat} is the temperature at which half of the RuBisCO is inactive. The RuBisCO oxygenase and carboxylase efficiencies are computed for given temperature and partial pressures of CO_2 and O_2. Maximal PS2 efficiency is set to be 2.44 times greater than the maximal efficiency of RuBisCO carboxylase $k_{c\,max}$ (as in Walker, B., et al. (2013)) and modulated by temperature (see Jmax(T) formula below). At the beginning of each simulation, the RuBisCO and PS2 efficiencies are computed using these nonlinear relationships and using the C, O and T values of the simulated environment, and used in constraints (C2). Then the RBA optimization problem is built and solved.

Numerical resolution. At a fixed growth rate, the RBA problem for eukaryotic cells is a Linear Programming (LP) optimization problem that can be solved efficiently at the whole-cell scale (Nesterov, Y., et al. (1994)). It has the same numerical complexity as the RBA problem of prokaryotic cells (Goelzer et al. 2009, Goelzer et al. 2011). Thus, the specificities of the RBA problem for eukaryotic cells (i.e. the f variables) did not increase the numerical complexity of the optimization problem. The maximal growth rate compatible with the available external resource can be computed by solving a series of LP optimization problems for different growth rate values. During model construction in the RBApy tool, when several enzymatic complexes catalyze the same chemical reaction, reactions are duplicated in order to obtain one reaction per enzymatic complex and having its own efficiency. However, the presence of (numerous) alternative isoenzymes can cause numerical problems when the LP problem is solved (i.e. the gradient is degenerating). Therefore, we chose to consider (during the resolution) a single average enzyme for which its amino acid and cofactor composition is equal to the average composition of isoenzymes. Algorithms of simulations have been implemented in Matlab R2018b using Cplex 12.8 (http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/) as the LP solver. The source code is distributed in an open source research software, named PlantCellRBA, under the EUPL v1.2 license, and available at https://forgemia.inra.fr/anne.goelzer/rba-plant-cell-model.

Uniqueness of the optimal solution. We implemented the numerical test of Appa 2002 to test the optimal solution at maximal growth rate for uniqueness. Briefly, the numerical test is performed in two steps and necessitates the resolution of two LP optimization problems (see Appa 2002 for proofs). First, the RBA LP problem at maximal growth rate has to be rewritten and solved in a standard form $max_z\ 0\ s.t.\ Az = b,\ z \geq 0$, which leads to a feasible solution z^*. Let T represent the set of all variables in z^* which have value zero. Then a second LP optimization problem is solved: $max_z\ dz\ s.t.\ Az = b,\ z \geq 0$, where $d_j = 1$ if $j \in T$ and 0 otherwise, which leads to a second feasible solution z^{**}. The original RBA problem at maximal growth rate has an unique solution if and only if $z^* = z^{**}$.

2.2 FBA model

We used the metabolic model provided in Supplementary file 2 to perform FBA simulations. It corresponds to the AraCore model of Arnold et al. 2014 that was updated with a few additional metabolic pathways - the synthesis of pigments (chlorophyll a and b), some cofactors (pyridoxal-5P, riboflavin, NAD(P)), and the transport of ions within the cell and between cellular compartments. A FBA optimization problem is formalized as $max_v\ c^T v\ s.t. \Omega v = 0,\ \alpha \leq v \leq \beta$, where Ω is the stoichiometry matrix, the vectors α and β are the lower and upper bounds on individual fluxes

respectively, and $c^T v$ is the criterion to be maximized (Varma et al. 1994). In our FBA simulations, either we maximized the biomass production (i.e. the flux called R_Bio_opt), or we minimized the use of photons (i.e. the flux called R_Im_hnu). We also set the flux of maintenance ATP to 7.3 millimoles/gDW/day as in the RBA model (i.e. the flux called R_NGAM_c) in all simulations. The other constraints described in this paper are added as equalities or inequalities in the FBA problem.

2.3 Farquhar model

We implemented the equations of the Farquhar model of (von Caemmer et al. 2009). Briefly, for given partial pressure of CO_2 (denoted C), O_2 (denoted O), irradiance (denoted I), temperature (denoted T), day respiration - mitochondrial CO_2 release (denoted R_d)-, the rate of triose phosphate utilization in the chloroplast (denoted T_p), and amount of RuBisCO, the Farquhar model computes the net assimilation rate of CO_2 (denoted A) as:

$A = (1 - \frac{\Gamma^*}{C})\, min(W_c, W_j, W_p) - R_d$ where

- $W_c = \frac{C\, V_{c\,max}}{C + K_c(1+O/K_o)}$ is the RuBisCO limited CO_2 assimilation rate. K_c, K_o and $V_{c\,max}$ are kinetic constants.
- $W_j = \frac{J}{4 + 8\,\Gamma^*/C}$ is the electron transport limited CO_2 assimilation rate. J is the potential electron transport rate which depends directly on irradiance I as $\theta J^2 - J(I_2 + J_{max}) + I_2 J_{max} = 0$, and $I_2 = I\, abs(1 - f)/2$. Parameters are θ, f and the kinetic constant J_{max}.
- $W_p = \frac{3\,T_p}{1 - \Gamma^*/C}$ is the export limited CO_2 assimilation rate
- $\Gamma^* = \frac{0.5\,O}{S_{c/o}}$ is the carbon dioxide compensation point and $S_{c/o} = (\frac{V_{o\,max}}{K_o}\,\frac{K_c}{V_{c\,max}})\frac{O}{C}$ is the relative specificity of RuBisCO.

The amount of active RuBisCO directly scales the maximal velocities $V_{c\,max}$ and $V_{o\,max}$, as $V_{c\,max} = k_{c\,max} Y_{RuBisCO} a_R$. $Y_{RuBisCO}$ is the amount of RuBisCO and $a_R = 1/(1 + e^{0.3(T - T_{sat})})$ is the fraction of RuBisCO that is active and T_{sat} is the temperature at which half of the RuBisCO is inactive.

The temperature dependence of the kinetic constants K_c, K_o and $V_{c\,max}$ is described by an Arrhenius function of the form Parameter(T) = Parameter(25°C) $\exp(\frac{(T-25)E_p}{298\,R\,(273+T)})$ where R is the universal gas constant and E_p is the activation energy of the parameter.

Finally, the temperature dependence of parameter J_{max} is given by

$J_{max}(T) = J_{max}(25) exp(\frac{(T-298)E_J}{298\,R\,T})\frac{1+exp(\frac{298\,S-H}{298\,R})}{1+exp(\frac{ST-H}{RT})}$, where the parameters S and H are the entropy and

enthalpy of a hypothetical equilibrium between an active and inactive form of the limiting component of electron transport, E_J is the apparent activation energy for low temperature limited electron transport (see (von Caemmer et al. 2009) for details). All parameter values were extracted from (von Caemmer et al. 2009) and (Walker, B., et al. 2013). To perform simulations, the

amount of RuBisCO, R_d, and T_p are fixed respectively to the predicted abundance of RuBisCO and to the predicted flux of CO_2 released by mitochondria (reaction R_Tr_CO2m) and to the rate of inorganic phosphate supply to the chloroplast (sum of the flux of reactions R_Tr_X5P, R_Tr_GPT1, R_Tr_PPT, R_Tr_TPT3, R_Tr_TPT2, R_Tr_TPT1,R_Tr_NTT) by RBA. Equations have been implemented and simulated using Matlab 2018b.

3. Results

3.1 A Resource Balance Analysis model for the leaf cell of *Arabidopsis thaliana*

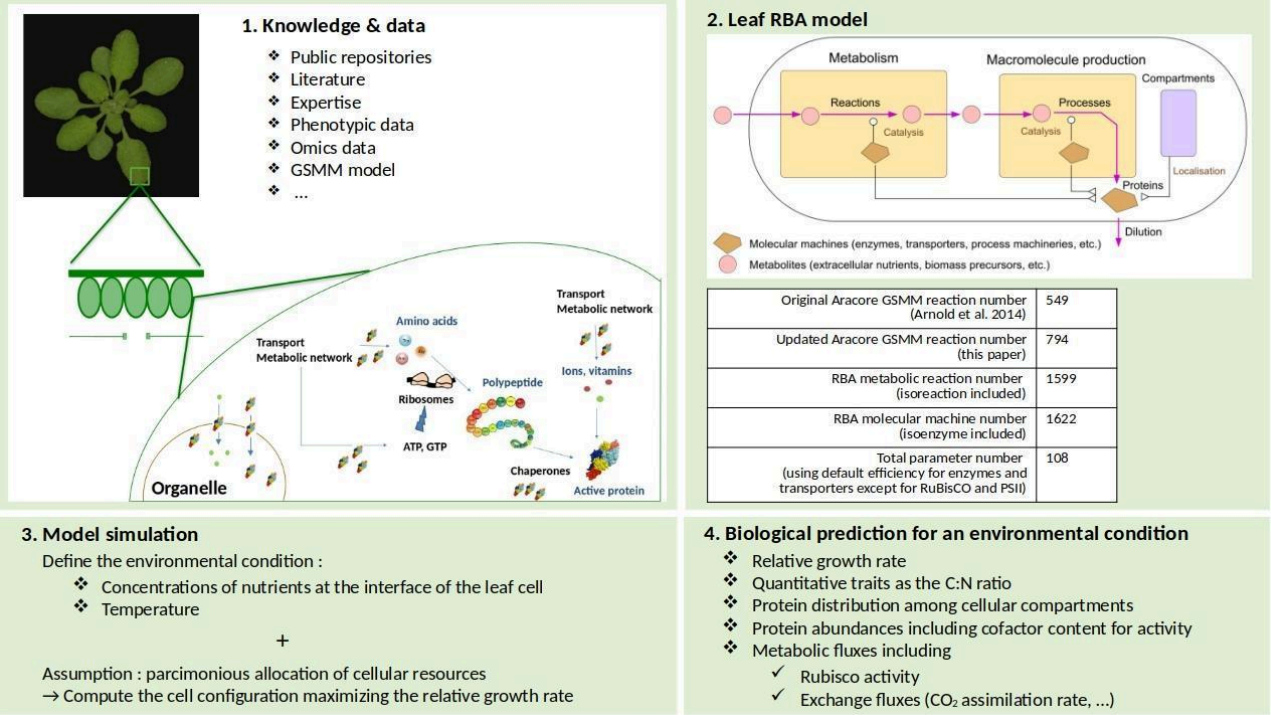

Fig1. Principle of RBA model generation and simulation. (**1**) Knowledge and information necessary for model building are extracted from various sources (e.g. public repositories, literature, etc.; see Supplementary Table 1). (**2**) The RBA model is generated using the software RBApy (Bulovic et al. 2019) and (**3**-**4**) simulated using the RBA modeling framework for eukaryotic cells (see Goelzer et al. 2019 and Methods). GSMM: Genome-scale metabolic model.

Our aim was to conduct a first *in silico* investigation of the resource allocation at the cellular and molecular scales in *Arabidopsis thaliana.* Focusing on a single cell type, we had to choose (i) a developmental stage and (ii) a cell type that is representative of plant functioning during this developmental stage and that can be handled by the RBA framework. We decided to focus on the vegetative growth phase, which corresponds in *A. thaliana* to half of its life cycle, beginning after the transition from the heterotrophic to autotrophic stage and ending before the flowering initiation; on short period of times and in optimal conditions the relative growth rate can be considered as constant and, hence, the growth as exponential. During this phase, the shoot part represents between 77% and 90% of the dry weight of the plant (Schulze, W., et al. 1991, Menz et al. 2018). Within the shoot part, the leaf cells are the main location of photosynthesis and concentrate most of the plant resources.

The RBA model for the leaf cell of *A. thaliana* integrates the cellular organelles (e.g. mitochondrion, chloroplast, etc.) and a set of cellular processes associated with each organelle. It includes the main cellular processes of gene expression, the protein chaperoning and transport (between organelles), the rRNA and tRNA production and transport, and the primary metabolism, to mention the ones requiring the larger part of resources. For gene expression, we included in

particular three translational processes, located in the cytosol, in the mitochondrial matrix and in the chloroplast respectively, in order to be biologically realistic and to illustrate the level of cellular details that can be managed by RBA. This implies that the RBA model also includes the TIC/TOC and TIM/TOM molecular machines (Thomson, S. M., el al. (2020); Ghifari, A. S., et al. (2018)) that import proteins from the cytosol into the chloroplast or into the mitochondrion respectively. Organelles are further detailed into sub-compartments such as chloroplast envelope, stroma, thylakoid membrane, and thylakoid lumen. This allows us to describe localized proteins and metabolites, and thus to have a detailed description of photosynthesis (including the chlorophyll needs for the functioning of photosynthetic complexes) and of carbon fixation reactions. Consequently, the model also includes the metabolite transporters necessary for metabolic exchanges between the different sub-compartments.

Generating an RBA model requires information for formalizing cellular constraints, in particular: (i) the localization and the composition of the molecular machines, (ii) the molecules that are consumed and released by the functioning of molecular machines; (iii) the efficiencies of molecular machines, i.e. the rates of the process per amount of the catalyzing molecular machine; (iv) other parameters such as the fraction of housekeeping proteins for each compartment. To build the model we used the software RBApy, conceived to assist modelers in the creation of RBA models by helping in information extraction and semi-automatic model generation (Bulovic et al 2019). RBApy necessitates as input a genome-scale metabolic model that accounts for a description of enzymatic complexes as Boolean (and/or) rules in SBML (Systems Biology Markup Language) format (Hucka et al. 2003), and a description of macromolecular machines - i.e ribosomes, chaperones, etc- in FASTA files. Although several genome-scale metabolic models exist for *A. thaliana* (Poolman et al. 2009; de Oliveira Dal'Molin et al. 2010; Seaver et al. 2014), only one - the AraCore model (Arnold et al. 2014) - contains a precise description of enzymatic complexes. The AraCore model specifically describes a leaf cell, that is, (i) it contains the protein subunits of enzymatic complexes that are localized in the leaf; (ii) the biomass composition corresponds to the leaf cell. Moreover, it has been manually curated, which makes it more trustworthy than non-curated models. We thus selected the AraCore model as a starting point and completed it manually with a few additional metabolic pathways - the synthesis of pigments (chlorophyll a and b) and some cofactors (pyridoxal-5P, riboflavin, NAD(P)), and the transport of ions within the cell and between cellular compartments - to describe in more detail the enzymatic complexes containing ions, cofactors or photosynthetic pigments. Finally, we added transporters of molecules in the membrane of organelles that are necessary to build the different molecular machines and to ensure their functioning. For modelling the macromolecular processes, we reconstructed manually the composition of molecular machines using information on UniProt (UniProt consortium 2017), KEGG (Kanehisa et al. 2000), ChloroKB (Gloaguen et al. 2017), literature and biological expertise.

Using all this information, we generated the RBA model of the leaf cell using the *prerba* module of RBApy (Bulovic et al 2019). The model contains 21 compartments (including different membranes of cellular organelles), 1599 metabolic reactions, 728 metabolites, 27 macromolecular processes (e.g. cytoplasmic translation), 1622 distinct molecular machines (including enzymes and transporters) composed of 1332 proteins, and 108 parameters. The RBA model is encoded in XML files using the RBAxml format (Bulovic et al 2019) (Supplementary file 1 and supplementary Table 1). The updated AraCore metabolic model is available in Supplementary file 2.

3.2 Model parameters

Model parameters include (i) the molecular machine efficiencies including, for instance, the ribosome efficiency (translation elongation rate) or the RuBisCO efficiency; (ii) the total amount of amino acids contained in proteins and the amount of housekeeping proteins allocated to each cellular compartment (called Pg parameters); (iii) the amount of macromolecular components such as DNA, mRNA, lipids or cellulose, as well as metabolic reserves such as starch, sucrose or free amino acids (called $\underline{B}$ parameters). For parameter estimation, we had to select one (ideally) or several datasets -especially quantitative proteomics data for the estimation of housekeeping protein abundance- acquired in comparable plant growth conditions (but being aware that variance in plant samples is high). We chose the datasets from Pyl et al. 2012 and Sulpice et al. 2014 acquired in short day experiments (8 hours of photoperiod) since they have been produced by the same group in comparable environmental conditions.

Following the parametrization procedure for prokaryotic RBA models, amounts of macromolecular components and metabolic reserves were extracted from the biomass composition used in the AraCore FBA model and were set as metabolic targets to be produced in the RBA model. The total amino acid content incorporated in proteins was computed using the total protein content of 18.97 mg per gram fresh weight (mg/gFW) and the dry over fresh matter ratio of 8.35% reported in Sulpice et al. 2014 for *A. thaliana* grown in short days and the amino acid composition of the Aracore model (Arnold et al. 2014).

In RBA models, housekeeping proteins correspond to proteins for which the function is not made explicit. The parameter Pg - the fraction of such "non-modeled" proteins in the whole proteome- can be roughly estimated from data by summing the measured protein mass of non-modeled proteins and dividing by the total measured protein mass. Using the quantitative proteomics datasets of Pyl et al. 2012, we obtained a value of 23% for the Pg parameter. However, this number probably underestimates the real value due to a limited coverage in the proteomics dataset (only a total of 273 proteins was quantified with a good confidence level). For bacterial cells, Pg represents approximately 50% of the total mass proteome (Scott et al. 2012, Goelzer et al. 2015). We thus decided to choose the intermediate value of 35% as a more realistic value, knowing that RuBisCO may represent between 23% (Li et al. 2017), and up to 50% of the leaf mass proteome (Arrivault et al. 2019).

The efficiency of RuBisCO was decomposed into nonlinear oxygenase and carboxylase activities as in the empirical Farquhar model of carbon fixation (Farquhar G, et al. 1980), using *in vivo* kinetic values reported in (Walker et al. 2013) and extracellular CO_2 and O_2 as variables (see Methods). PS2 maximal catalytic rate was assumed to be 2.44 higher than the maximal carboxylase activity of RuBisCO (Walker et al. 2013). The *in vivo* efficiencies of other metabolic processes were assigned to a default value of $7s^{-1}$, which was reported as the mean value of *in vivo* estimates in glucose minimal medium (Goelzer et al. 2015) for bacteria, and lower than the median maximal catalytic rates of 13.7 s^{-1} reported in (Bar-even et al 2011) for all organisms (including plants). Maintenance ATP flux was set to 7.3 millimoles/gDW/day in agreement with values reported in Cheung et al. 2013 in heterotrophic growth of cell cultures. Ribosome efficiency was set to an elongation rate of 3 amino acids per second as in (Piques et al. 2009).

3.3 Predicted cell configuration in standard non-limiting growth condition

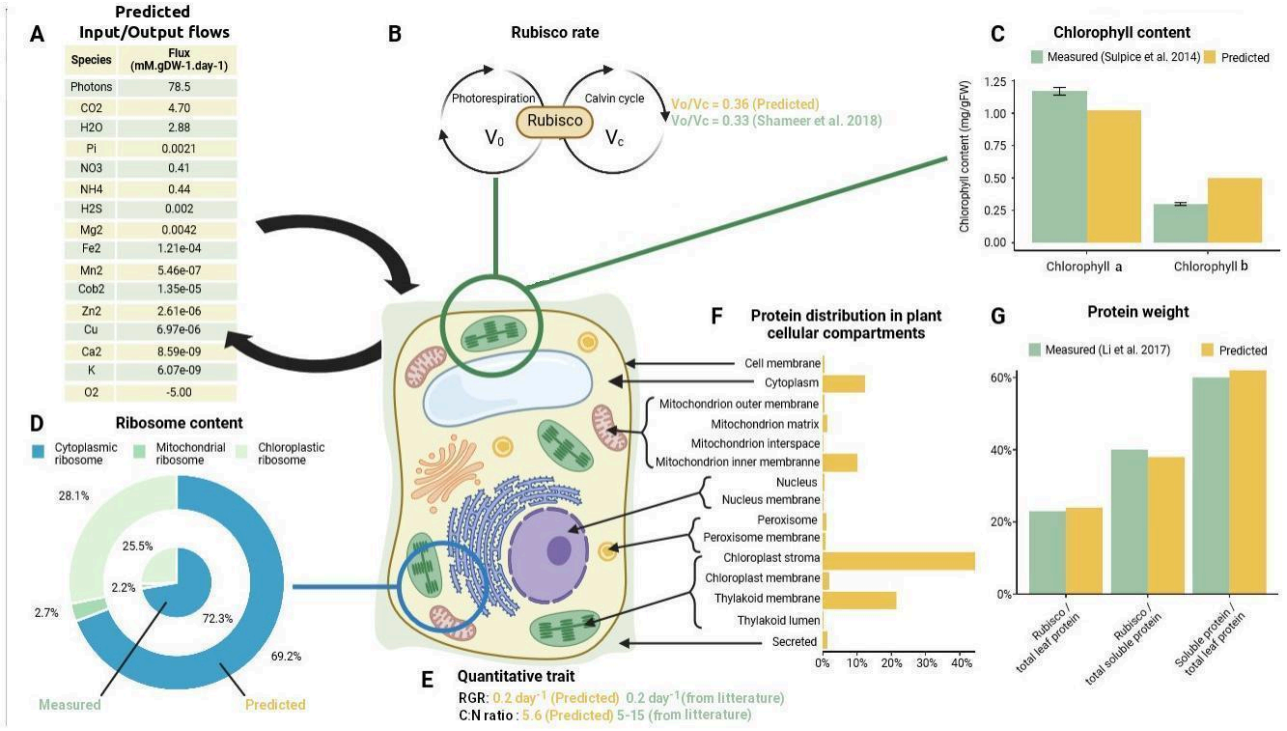

Fig2. RBA predictions for a standard non-limiting environmental growth condition at 21°C (irradiance: 1000µE/m2/s; partial pressures of CO_2: 200ubar and O_2: 210 mbar). Numbers shown in the different graphs and tables are a subset of RBA predictions obtained at the maximal relative growth rate. The measured chlorophyll content and protein weight were extracted from Sulpice et al. 2014 and Li et al. 2017 respectively. **(A)** Assimilation and excretion rates of nutrients in millimoles/gDW/day. (**B**) Ratio between oxygenase (Vo) and carboxylase (Vc) activity of RuBisCO. (**C**) Chlorophyll content (in mg/gFW). (**D**) Ribosome content. **(E)** Quantitative traits, i.e. the relative growth rate (RGR), the C over N ratio. (**F**) Comparison of predicted and measured weight of RuBisCO, soluble proteins in total proteins (in %). **(G)** Protein distribution among cellular compartments (in %). gDW, gFW stand for gram dry weight and gram fresh weight respectively and we considered that gDW is 8.35% of gFW as in Sulpice and al. 2014.

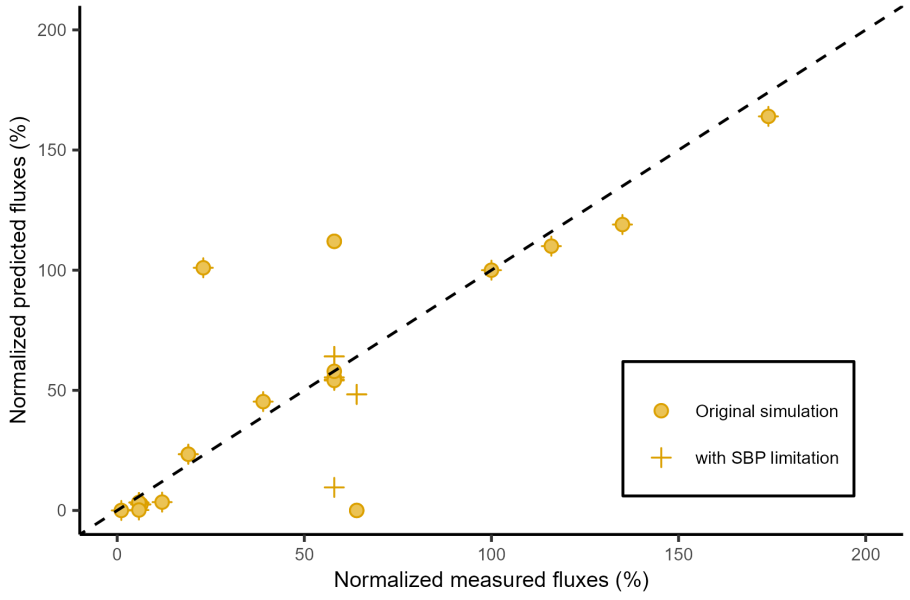

Fig3. Comparison of the RBA predicted flux distribution with the fluxomics dataset of (Ma et al. 2014). Predictions were performed for a standard non-limiting environmental growth condition at 21°C (irradiance: 1000µE/m2/s; partial pressures of CO_2: 200µbar and O_2: 210 mbar) with (red diamonds) and without (blue cross) limiting the

sedoheptulose 1,7 biphosphate (SBP) aldolase. For comparison, the predicted fluxes were normalized by the predicted net assimilation rate of CO_2 as it was done in Ma et al. 2014.

Fig4. Protein cost of cellular processes divided by the total amino acid content. The protein cost of a cellular process (shown on x-axis on log10 scale) is computed as the sum of the predicted abundance (in millimoles.gDW^{-1}) of the proteins belonging to the cellular process multiplied by the number of amino acid residues in the protein. The correspondence between the proteins and the cellular processes is given in Supplementary table 1.

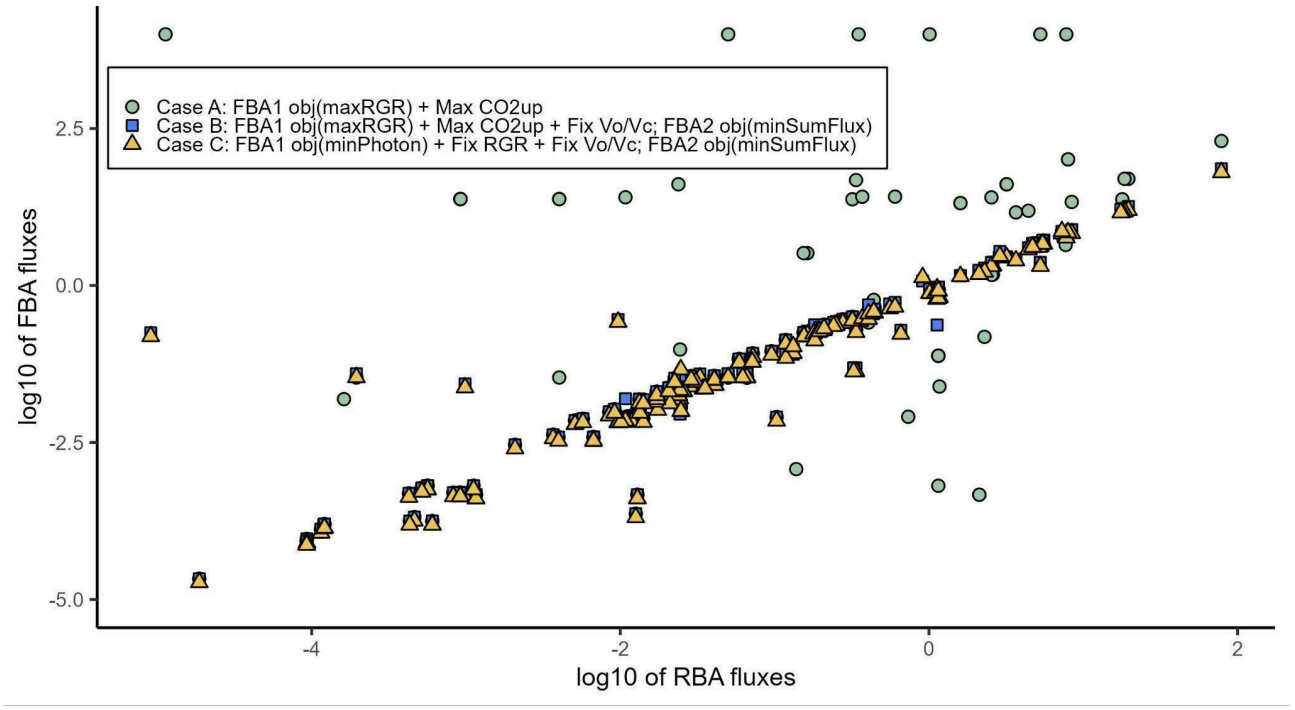

Fig5. Metabolic flux predictions, compared between RBA and three different variants of Flux Balance Analysis (FBA) with the following respective assumptions: case (A): Biomass production (i.e. relative growth rate) is maximized and the maximal uptake rate of CO_2 is set to the RBA predicted value (green circles); case (B): Same as case (A), the flux ratio between oxygenase and carboxylase activities of RuBisCO is set to 0.33; then a second FBA problem is solved in which the same constraints are applied, the biomass production is set to the optimal value obtained in case (A), and the sum of fluxes is minimized (blue squares); case (C): First, photon influx is minimized at a predefined biomass production and a predefined flux ratio between oxygenase and carboxylase activities of RuBisCO; then a second FBA problem is solved in which the same constraints are applied, the photon influx is set to its previous optimal value, and the sum of fluxes is minimized and the photon influx is set to the optimal value (yellow triangles).

Our primary objective is to investigate if an assumption of parsimonious resource allocation between the plant cellular processes leads to realistic predictions of cell configurations. To this purpose, we first focused on standard non-limiting growth conditions (see Methods), setting the concentrations of nutrients at the interface of the leaf cell (i.e. metabolites in xylem, partial pressure of CO_2 and oxygen, photons) and the temperature (21°C). We then computed the cell configuration that maximized the relative growth rate of the leaf cell and compared the predicted configuration - protein abundances and metabolic fluxes including cell exchange fluxes - to measured quantitative traits in similar environmental conditions (see Fig.2), that is, the relative growth rate, the C:N ratio, the ratio of oxygenase/carboxylase activities of RuBisCO, the chlorophyll content, the weight of RuBisCO or soluble proteins with respect to the total protein content, and the ribosome concentration.

Metabolic fluxes and Quantitative traits (Fig.2A,B,E, Fig.3). Predicted cell exchange fluxes (i.e. import and export cell fluxes) are shown in Fig.2A, with (i) a predicted influx of photosynthetically active photons that is actually absorbed and used in photosynthesis of 80 millimoles.gDW^{-1}.day^{-1}, and (ii) a CO_2 assimilation rate of 4.7 millimoles.gDW^{-1}.day^{-1}. By using a specific leaf area of 31.6 m^2.$kgDW^{-1}$ for a plant grown in short days (8 hours of day light) reported in Pyl et al. 2012, the predicted influx of photosynthetically active photons corresponds to 87.9 µmol.m^{-2}.s^{-1} and the predicted CO_2 assimilation rate during the day corresponds to 5.2 µmol.m^{-2}.s^{-1}. These values are in reasonable agreement with the values reported for a whole-plant chamber, i.e. a CO_2 assimilation

rate between 2.9 and 7.1 µmol.$m^{-2}.s^{-1}$ for an illumination between 60 and 160 µmol.$m^{-2}.s^{-1}$ (Kölling et al. 2015). Recall that compared to FBA, we did not impose any bounds on cell exchange fluxes. The flux distribution results from a compromise with respect to the use of resources. The predicted flux distribution is realistic and consistent with the literature. The predicted relative growth rate (0.2 day^{-1}) falls in the range of the measured one (0.17 to 0.20 ± 0.02 day^{-1}, Fig.2E see Sulpice et al. 2014); this also holds for the predicted C:N ratio of 5.6 (expected to be between 5 and 15) (Chietera et al. 2018). The ratio between the oxygenase and carboxylase flux catalyzed RuBisCO is predicted to be 0.38 (Fig.2B), in agreement with the Farquhar and FBA models (Farquhar et al. 1981; Shameer et al. 2018).

We compared the predicted metabolic flux distribution with the measured fluxomics data of Ma et al. 2014 (using their light intensity of 200 µmol.$m^{-2}.s^{-1}$). For that, we normalized, as it was done by the authors, the flux values by the net assimilation rate of CO_2. Predictions and measurements are in reasonable agreement (see Supplementary Table 1, Fig.3), with for instance a predicted flux through the carboxylase and oxygenase of Rubisco of 119 and 45 compared to the measured values of 135 and 39 respectively. Discrepancies first concern the production of starch and sucrose. This is expected since the sucrose and starch content are different in the data of Ma et al. 2014 and in our primary source of information (Arnold et al. 2014). A second discrepancy concerns the second half of the Calvin cycle. The predicted synthesis of chloroplastic F6P is produced by the pentose phosphate pathway (transketolase and transaldolase) and not by the FBP aldolase and FBP biphosphatase. However, the network used to compute the flux distribution in Ma et al. 2014 does not contain all the metabolic intermediates in the pentose phosphate pathway, which could have an impact on the final computation of the flux. To test further, we limited the flux through the sedoheptulose 1,7 biphosphate aldolase (see Supplementary table 1, Fig.3). The impact on the predicted growth rate is negligible with a decrease of only 0.1% and no reshaping in the flux distribution. It means that the two pathways have similar cellular costs and can be used equally.

Chlorophyll content (Fig.2C). To determine the predicted chlorophyll content, we summed the amounts of chlorophyll a and b associated with the predicted abundances of photosystem 1 (PS1), photosystem 2 (PS2) complexes and compared them to the measurements of chlorophyll a and b acquired at the end of the day in Sulpice et al. 2014. The predicted chlorophyll a content is already in a realistic range, while chlorophyll b is moderately overestimated. Actually, we assumed that 165 and 170 copies of chlorophyll a (and respectively 95 and 20 copies of chlorophyll b) were contained in PS1 and PS2 respectively. This was based on the values reported for plants in Nobel 1999. For *A. thaliana*, the number of chlorophyll b per photosynthetic complex in the model may be too high.

Ribosome content (Fig.2D). Piques et al. 2009 reported a measured concentration of ribosomes at 0.1 nmol/gFW (summed cytosolic, plastid and mitochondrial ribosomes) while we obtained a predicted total concentration of ribosomes of 0.14 nmol/gFW. Moreover, predicted cytosolic ribosomes were 2.4-fold more abundant than predicted plastid ribosomes, and 26-fold more abundant than predicted mitochondrial ribosomes. These numbers agree well with known experimental data (Piques et al. reported values of three-fold and 30-fold respectively).

Protein weight (Fig.2F,G). In Li et al. 2017, the authors reported the weight of RuBisCO with respect to leaf or soluble proteins, and the weight of soluble proteins with respect to the total proteome. Fig.2G shows that our predictions are well within the range of expected values. Fig.2F shows the predicted distribution of proteins (i.e. expressed in amino acid residues per gram of cell dry weight) among the cellular compartments. The prediction states that 68% of the protein mass are located in the chloroplast, 12% in the cytosol and 12% in the mitochondrion, which is globally in the range of the values reported in Heinemann et al. 2021 (i.e. 66%, 8.6% 2.1% for the same compartments respectively). 45% of proteins are located in the stroma of the chloroplast, in

agreement with the fact that at least 23% of the leaf protein is RuBisCO (Li et al. 2017; Heinemann et al. 2021), and thus located in this compartment.

Protein cost of cellular processes (Fig.4). We defined a set of cellular processes, composed of metabolic pathways and the macromolecular processes of the model, and assigned each metabolic reaction (and thus its associated enzymatic complex) to a single metabolic pathway (see supplementary table 1). We then computed the predicted protein cost of each cellular process by summing the predicted abundances (in millimoles.gDW^{-1}) of all molecular machines associated with the process and multiplied by the number of amino acid residues (Fig.4). Expectedly, the most costly cellular processes are the Calvin-Benson-Bassham cycle and the light reactions that together totalize around 52% of the cell mass (recall that RuBisCO by itself represents 23% of the leaf proteome (Li et al. 2017)). Mitochondrial metabolism (i.e. oxidative phosphorylation, TCA cycle, metabolite transport) represents 6.3% of cell mass. One will note the high weight of metabolite transport into organelles (chloroplast and mitochondrion) compared to other processes. Obviously, the predicted protein cost depends on the efficiency coefficients of molecular machines. For a given flux, the higher is the efficiency, the lower is the abundance of the molecular machine necessary to maintain the flux. By using a default unique value for all enzymes and transporters (except for RuBisCO and PS2) efficiencies, we averaged the variability of *in vivo* apparent catalytic rates within a cellular process, which led to compute an averaged (but representative) protein cost of cellular process.

Cellular ATP balance. The model predicts that cellular ATP is produced mainly in the chloroplast by the chloroplastic ATP synthase (67%), and to a lesser extent by the mitochondrial ATP synthase (21%) and by the cytosolic phosphoglycerate kinase (12%). ATP is also exchanged between organelles. Around 86% of chloroplastic ATP originates from the chloroplastic ATP synthase, while 13% is translocated from the cytosol by the chloroplastic nucleoside triphosphate transporter (NTT). Predicted mitochondrial ATP is produced exclusively by the mitochondrial ATP synthase and furnishes 63% of cytosolic ATP by mitochondrial adenylate nucleotide translocase. Glycolysis produces the remaining 37% of cytosolic ATP. Chloroplastic ATP is mainly consumed by the Calvin-Benson-Bassham cycle while cytosolic ATP is mainly consumed by maintenance processes and by translocation into the chloroplast (see Supplementary table 1).

In mature illuminated chloroplasts, the main source of chloroplastic ATP is the chloroplastic ATP synthase (Voon and Lim 2019), which is in agreement with RBA predictions. However, the fact that the NTT transporter is active in illuminated chloroplasts still remains a matter of debate (see Lim, S. L., et al (2022); Flügge, U. I., et al. (2011) and references therein). We thus prevented the use of the NTT transporter and runned a new simulation for the same standard non-limiting growth conditions. The cell is viable, but the predicted growth rate decreased by 20% and the flux distribution changed (see Supplementary table 1): nitrate and sulfate are now the only sources of nitrogen and sulfur and imported by proton-symporters in contrast to the original simulation, where NH4 and H2S were also imported (Fig.2A). Glycine is now excreted which causes loss of carbon and nitrogen. The other major changes involve exchanges of glycolytic intermediates (3-phosphoglycerate, phosphoenolpyruvate, Glyceraldehyde 3-phosphate), carboxylic acids (Malate, 2-Oxoglutarate) and glutamate between the chloroplast and the cytosol, and between the cytosol and the mitochondrion (isocitrate, succinate, proline, glutamate) to a lesser extent. Thus, according to the model, the NTT transporter is not necessary, but significantly improves the growth when turned on (i.e. when chloroplast is supplemented with cytosolic ATP). This simulation also

indicates that an alternative pathway for chloroplastic ATP supplementation does not seem to exist.

Uniqueness of the solution. In standard CBM methods, such as FBA, multiple flux distributions are compatible with the optimal growth rate. To analyze the existence of alternative flux distributions at the optimum, we implemented the numerical test of Appa, 2002 (see Material and methods), which consists in building (if possible) an alternative solution at the optimal growth rate. If an alternative solution does not exist, the solution is unique. Our RBA solution is very close to being unique (see supplementary table 1). Alternative solutions would involve a combination of different mitochondrial carriers of carboxylic acids and a combination of different chloroplastic translocators of phosphate/triose phosphates. These combinations actually have similar protein costs in the model. This has a simple reason. Since the genes coding for these carriers were either the same, or unknown in *A. thaliana*, we had assigned to them a default protein of mean amino acid composition when constructing the model, resulting in different combinations of metabolic reactions that have the same function and the same cost in terms of resources (cofactors, proteins, etc.). However, the metabolic fluxes through these alternative reactions are weak compared to the flux of carbon fixation or energy generation.

Comparison with FBA simulations (Fig.5). We finally compared the flux distribution predicted by RBA to fluxes obtained by FBA. We used our updated metabolic model Aracore (given in Supplementary file 2) that contains the same metabolic reactions as the RBA model. We considered different FBA optimization problems including different constraints and objectives used in plant modeling. Possible FBA constraints are setting (i) the relative growth rate to the one predicted by RBA, (ii) the CO_2 assimilation rate to the one predicted by RBA, and (iii) the ratio between oxygenase and carboxylase activities of RuBisCO to 0.33 as in Shameer et al. 2018. Possible objectives are the maximization of the relative growth rate or the minimization of the photon influx for fixed relative growth rate, and followed by the minimization of the sum of the fluxes (see supplementary table 1 and case A in Fig.5). The flux distribution predicted by FBA is close to the RBA prediction when (i) the ratio between oxygenase and carboxylase activities of RuBisCO is set, (ii) the sum of the flux is minimized, and either the CO_2 assimilation rate or the relative growth rate is set to the value predicted by RBA (see cases B and C in Fig.5). Measures of CO_2 assimilation rates are rarely available, in contrast to measured relative growth rates. Consequently, in practice, FBA modelers may set (i) the flux of produced biomass to the measured relative growth rate, (ii) the ratio of RuBisCO activities to a known value (as 0.33) while minimizing the flux of photons (de Oliveira et al. 2010). Then, they may solve a second optimization problem where the sum of the flux is minimized while imposing the photon influx to the optimal value of the first FBA problem (Gerlin et al. 2022). This scenario corresponds to the yellow triangles of Fig.5 (case C), which is a good approximation of the RBA solution in standard non-limiting conditions.

Altogether, these results show that the predicted cell configuration fits quantitatively with a realistic configuration of a photosynthetic cell in the rosette. In particular, the RBA model accurately predicts the relative growth rate and the exchange cell fluxes in an autonomous way (i.e. without a need for predefined values or bounds on exchange fluxes) for a given environmental condition. This constitutes a major breakthrough in the prediction of plant cellular phenotypes compared to the state of the art in plant constraint-based modeling.

3.4 Impact of parameter variations on RBA predictions

Model parameters are compiled from different sources and are consequently subject to uncertainties. We thus investigated the sensitivity of our predictions to parameter variation by performing a so-called One-At-a-Time (OAT) local sensitivity method. We multiplied each parameter by 0.5 and 2 (two-fold increase or decrease). We then computed again the resource allocation at maximal growth and quantified the impact of parameter variations on several phenotypic traits - relative growth rate, CO_2 assimilation rate and C over N ratio- by the relative variation of the predicted phenotypic trait with respect to the nominal prediction. Variations of thirty-five of the parameters (given in Supplementary Table 1) led to changes in growth rate greater than 1% (Fig.6). The greatest impacts were observed with parameters related to total protein content (total amino acid concentration to be allocated), kinetic parameters of RuBisCO (affinity parameters for CO_2 and O_2, maximal turnover rates for oxygenase and carboxylase activities), and RuBisCO activation (temperature for which half of the RuBisCO is active), the maximal turnover rate of PS2 photosynthetic systems, mean efficiencies of enzymes and transporters, parameters involved in the definition of housekeeping protein abundances (fraction of total proteins and of Pg proteins allocated to the different cellular compartments), efficiencies of ribosomes and chaperones, value of maintenance ATP, and metabolic requirements for membrane (Malonyl-ACP, ACP) and cell wall (cellulose) synthesis and for metabolic reserves (sucrose, starch). Moreover, thirty-three of these thirty-four parameters also correspond to the top parameters impacting the predicted CO_2 assimilation rate (Supplementary figure 1). In contrast, only nine parameters led to variations on the C:N ratio greater than 1%: the total protein content, the default transporter efficiency, metabolic reserves (sucrose, fumarate, starch, nitrate), cell wall (cellulose) and lipid requirements (Malonyl-ACP, ACP), see Supplementary figure 2. Moreover, the direction of the effect of parameter variations is in agreement with the theoretical results of Goelzer et al. 2011. Any parameter leading to an increase in the abundance of housekeeping proteins (the fraction of Pg and of proteins in cellular compartments) and in the abundance of cellular macro-components, i.e. lipid requirements (concentration of Malonyl-ACP, the major building block of fatty acid synthesis), cell wall requirements (cellulose) or metabolic reserves (concentrations of starch, sucrose and fumarate), decreases the growth rate. Increasing the efficiency of molecular machines (default efficiency, default transporter efficiency, ribosome and chaperone efficiencies) increases the growth rate.

In addition to variations related to RuBisCO and PS2 complex (above), we extended the OAT sensitivity test to individual enzyme and metabolic transporter efficiencies to identify those that have the greatest impact on predictions. We increased or decreased each individual enzyme efficiency by a factor of two, while the other efficiencies remained fixed to their default value ($7 s^{-1}$). Then we computed the resource allocation that maximizes growth rate and quantified the impact of parameter variation on the phenotypic traits as described above. The impact of variations of individual enzyme or transporter efficiencies is generally weaker than with the other parameters. Variation in efficiency of only 18 of the enzymes led to growth rate variations greater than 0.4% and correspond to efficiencies of enzyme or transporter catalyzing high fluxes (supplementary table 1). Enzymes are involved in energy generation (oxidative phosphorylation chain, chloroplastic and mitochondrial ATP synthase), and carbon fixation (Photosynthetic complexes, Calvin-Benson-Bassham cycle and glycolytic enzymes) (Fig.7; supplementary figures 3-4). In supplementary figures 5-6, we show the variability of individual flux values caused by variations of each individual enzyme or transporter efficiency. Variations of some of the efficiencies induce changes in the cell configuration. They are mainly associated with enzymes that catalyze high fluxes or metabolic carriers into organelles and belonging to cellular processes having

a high protein cost (i.e. the categories named photosynthesis, oxidative phosphorylation, central carbon metabolism and metabolite transport on Fig.4).

We finally explored the sensitivity of predictions to joint variations of molecular machine efficiencies having high uncertainties, i.e. those of enzymes (except for RuBisCO and PS2 that have been assigned experimentally-measured values for Arabidopsis), transporters, ribosomes, chaperones, protein translocation machineries into organelles, RNA and DNA polymerases. To ensure biological realism, we did not assume a predetermined interval of variation for each efficiency, but rather assumed that the empirical distribution of efficiencies follows a log-normal distribution, with the original value identified in the literature as the mean (given in Supplementary Table 1) and a conservative standard deviation of 0.7, as it was shown in Goelzer et al. 2015. We then randomly sampled 1000 values in the distribution of each molecular machine, leading to 1000 distinct combinations of molecular machine efficiencies. For each combination, we computed the resource allocation that maximizes growth rate and quantified the impact of parameter variation on the phenotypic traits as described above. Supplementary figures 7-9 show the empirical distribution of predicted growth rate and CO_2 assimilation rate and the variability in individual flux predictions. The range of predicted growth rate and CO2 assimilation rate increased by ~30% expectedly with the high standard deviation that we chose, but still display a nice unimodal distribution. It indicates that two (or more) alternative cellular configurations with very different protein costs that could be turned on or off (and thus lead to a multimodal distribution) should not exist. The model behavior is thus robust to variations of individual efficiency parameters (as in bacteria; see Goelzer et al. 2015).

Altogether, predictions are unsurprisingly sensitive to the efficiencies of costly molecular machines (e.g. RuBisCO) and to parameters related to plant cell physiology (e.g. the total protein content, abundance of housekeeping proteins, etc), which are associated with significant portions of the cellular proteome (abundance of housekeeping protein at 35% and RubisCO at 25%; Figure 2G). Any changes affecting the final concentration of these proteins can lead to significant alterations in resource consumption, thereby strongly influencing the growth rate. Small variations in individual efficiency parameters are expected to have a lesser impact on the growth rate due to their lower weight contribution to the total proteome, the local sensitivity test being performed at the same efficiency value ($7s^{-1}$). The known dynamics of individual molecular machine efficiencies collected on many organisms and across metabolic pathways is larger (from 10^{-1} to 10^{3} s^{-1} for enzymes; see Bar-even et al. 2011 and Goelzer et al. 2015). Accordingly, we obtained a higher impact on predictions when joint variations on efficiencies are tested (Supplementary figures 7-9). To go further, we need a more precise dynamic of each parameter for *Arabidopsis thaliana*, *i.e.* estimated on dedicated datasets. This could enable us to assess more precisely the contribution of each individual molecular machine or of alternative metabolic pathways to the growth rate, and to highlight the plant strategies with respect to cellular economics.

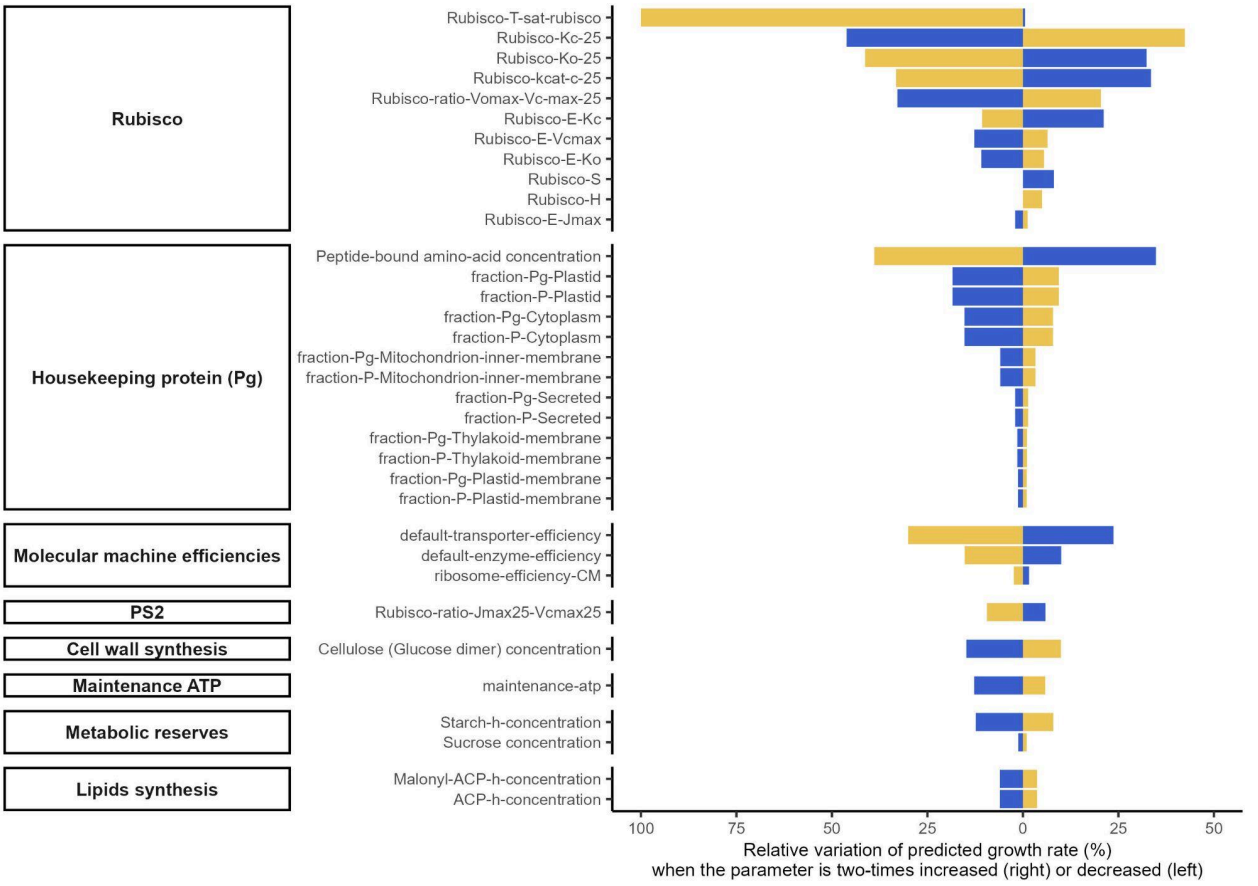

Fig6. Sensitivity of relative growth rate prediction to individual parameter variation. RBA predictions were obtained by increasing (right-side of the 0 axis) or decreasing (left-side of the 0 axis) the parameter value by a factor of two. Colors indicate if the parameter variation leads to an increase (blue) or to a decrease (yellow) on the relative growth rate. Only parameters that lead to a relative variation higher than 1% are displayed.

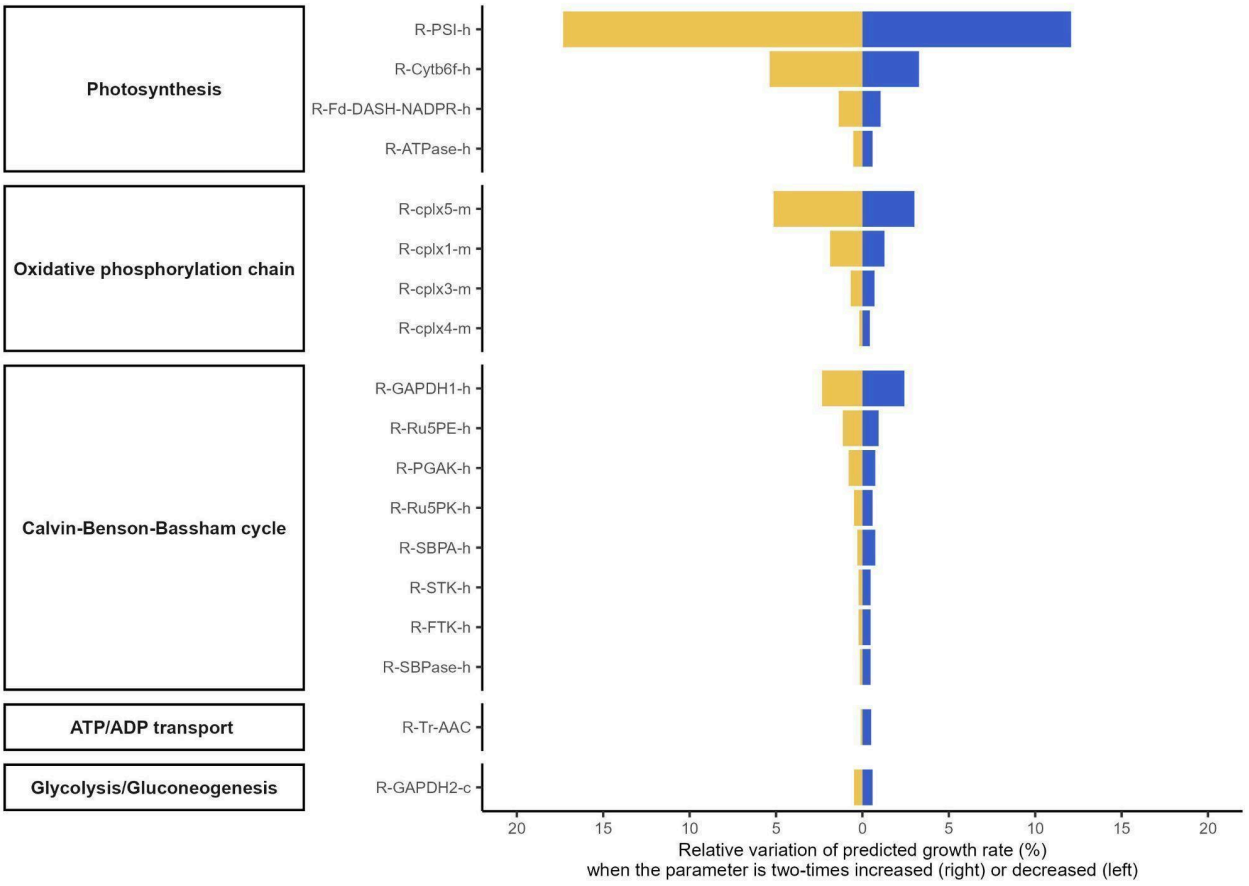

Fig7. Sensitivity of predicted relative growth rate to individual enzyme efficiency. RBA predictions were obtained by increasing (right-side of the 0 axis) or decreasing (left-side of the 0 axis) the parameter value by a factor of two. Colors indicate if the parameter variation leads to an increase (blue) or to a decrease (yellow) of the relative growth rate. Only parameters that lead to a relative variation higher than 0.4% are displayed. Parameters are grouped by the biological processes to which they belong.

3.5 Prediction of plant phenotypic response under complex environmental conditions

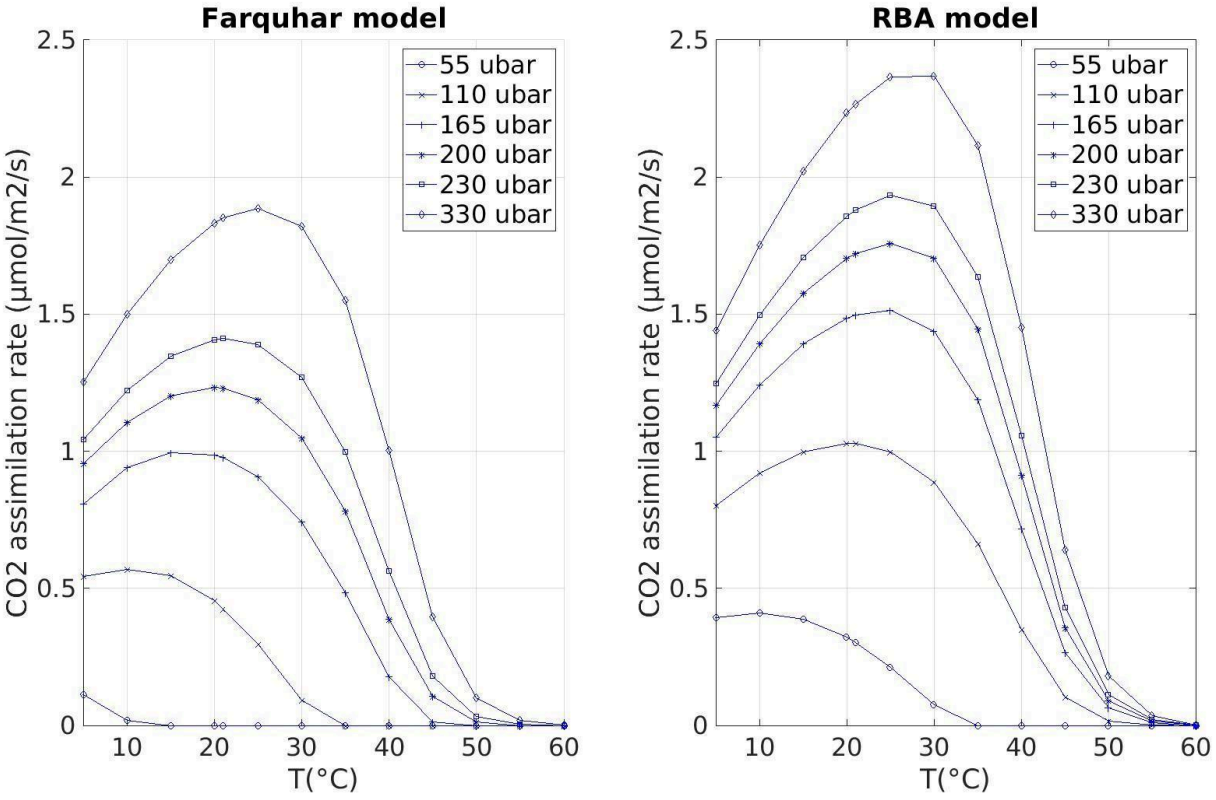

Fig8. Comparison of predicted CO_2 assimilation rates response for different partial pressure of CO_2 (in ubar) and temperature (°C). The plots show results from the Farquhar model (Farquhar et al 1981; Caemmerer, S., et al. (2009) calibrated for *A. thaliana* (**left**) and the RBA model (**right**). For the RBA model, each point corresponds to a simulation in steady-state where the growth rate is maximized.

Even if the literature on plant constraint-based modeling is vast, papers usually present *in silico* analyses of the distribution of metabolic fluxes, and rarely confront predictions to experimental data. A main reason for this is the experimental cost (and difficulties) of acquiring fluxome in the whole-plant. An important difficulty in plant constraint-based modeling is thus to validate quantitatively the model behavior. While a dedicated dataset for full experimental validation is not available, we can still evaluate the model behavior by comparing the RBA model predictions with simple and well-established empirical models that capture the macroscopic behavior of the organism. For plants, the most well-established empirical model mimicking plant cell behavior is the model of photosynthesis and carbon fixation developed by Farquhar and colleagues (Farquhar et al. 1981, von Caemmerer et al. 2009).

For the comparison, we first implemented the temperature dependency in RBA as in the Farquhar model (i) by the Arrhenius law for the kinetic parameters of RuBisCO using the parameter values given by Walker, B., et al**.** 2013 and (ii) by modeling the inactivation of RuBisCO by the inhibition of the RuBisCO activase at moderate to high heat stress (see Methods and Kurek et al. 2007). Then, we implemented the equations of the Farquhar model provided by von Caemmerer et al. 2009 and using the kinetic *in vivo* parameters of RuBisCO described by Walker et al. 2013. In addition to the kinetic parameters of RuBisCO and of the photosynthesis system II, the Farquhar model is set up by the RuBisCO content in g/m2, by the rate of CO_2 released by respiration (parameter R_d, usually given in $mol.m^{-2}.s^{-1}$), and by the rate of use of triose phosphate in chloroplast (parameter T_p, usually given in $mol.m^{-2}.s^{-1}$), i.e. three quantities that are predicted by the RBA model. Parameter

T_p is the rate of inorganic phosphate supply to the chloroplast (von Caemmerer et al. 2009), which corresponds to the sum of the flux of phosphate-dependent carriers in the RBA model. We first ran the RBA model for varying partial pressure of CO_2 (55 to 330 ubar), O_2 (10 to 330 mbar), temperature (5 to 60°C) and irradiance (0 to 2000 µmol.m^{-2}.s^{-1}) as defined by Farquhar et al 1981 and von Caemmer et al. 2009. For each simulation, we extracted the predicted amount of RuBisCO (in millimole.gDW^{-1}) and the predicted CO_2 excretion rate by respiration and chloroplastic phosphate import (both in millimole.gDW^{-1}.day^{-1}), converted them to the relevant units (g.m^{-2}, µmol.m^{-2}.s^{-1}) using a specific leaf area of 31.6 m^2.$kgDW^{-1}$ reported in Pyl et al. 2012, and finally used them as input variables for the Farquhar model. The comparison of the CO_2 assimilation rates for varying CO_2 partial pressure and temperature between both models is given in Fig.8 (R^2=0.93, p-value=1.5e^{-48} of the F-statistics). Supplementary figures 10-11 show results for other scenarios such as varying O_2-temperature, varying CO_2-irradiance.

Both the curve shapes and the order of magnitude of the CO_2 assimilation rate are similar between the two models (Fig.8). This indicates that the RBA model captures the macroscopic behavior of the plant cell as given by the Farquhar model and that it can already quantitatively predict the assimilation rate of CO_2 for varying CO_2 and O_2 partial pressures, irradiance and temperature. However, and in contrast to the Farquhar model, the RBA model is more autonomous: the RuBisCO content, the CO_2 excretion rate by respiration, the electron transport rate are not fixed a priori or determined through empirical curves but result from a parsimonious trade-off between cellular processes with respect to resources.

4. Discussion

Using the RBA framework (Goelzer et al. 2009, Goelzer et al. 2011, Goelzer et al. 2019), we built the first model of a plant cell on a cellular scale that accounts for resource allocation between the major cellular processes. The RBA model defines a set of possible phenotypes compatible with a given specific growth rate. Among the possible phenotypes, we investigated if an assumption of parsimonious resource allocation - i.e. growth rate maximization - leads to realistic predictions of plant cell configurations as in bacterial cells. Using realistic biological parameters, simulating the RBA model leads to accurate and realistic predictions not only in a standard non-limiting growth condition, but also when different abiotic limitations vary in combination. Hence, the RBA model captures the behavior of the well-established model for carbon fixation and photosynthesis from Farquhar and colleagues (Farquhar et al. 1982) under combinations of carbon, irradiance and temperature condition variations. Further validation of the extent of the model genericity remains to be performed under other abiotic conditions such as for instance combinations of ions, sulfur, phosphorus, nitrogen or carbon limitations and/or across diverse genotypes to test specific pathways' limiting effect.

A more autonomous CBM model. Compared to other constraint-based models based on the Flux Balance Analysis framework (Varma and Palsson 1994), the cell configuration in RBA is not fixed *a priori*, but results from a global compromise with respect to the parsimonious use of resources. RBA overcomes the need to predefine several biological objectives or constraints in order to obtain accurate quantitative predictions, such as the bounds on nutrient uptake rates, minimization of photon influx, minimization of the sum of the fluxes (see supplementary Table 1) which are commonly used in other plant constraint-based modeling. In parallel, our results also demonstrate that some combinations of criteria or constraints imposed in FBA allow the recovery of the RBA flux distribution on standard non-limiting growth conditions. This highlights the fact that FBA is a useful tool for guided quantitative cellular exploration, when adequate targets such as the relative

growth rate, or the uptake rate of the limiting nutrient can be set to measured (or relevant) quantitative values. Moreover, the detailed biomass composition of RBA models is not fixed and given as input information as in FBA, but is predicted by the model itself. The model predicts the optimal macromolecule composition of the cell, and this prediction is very detailed: it follows from the predicted proteome and may comprise details like trace elements (e.g. iron, which appears in some of the proteins) or cofactors (e.g. chlorophyll a and b requirements in photosynthetic complexes) and whose contribution to the overall biomass will vary with the expression level of those proteins. RBA could thus assist FBA model parametrization, by providing (i) a biomass composition adjusted to the environmental condition (Bodeit et al. 2023), and (ii) realistic bounds for nutrient uptake rates. Our RBA model is thus more autonomous and represents a significant step forward in predictive plant biology in view of the coverage of cellular functions, the accuracy of predictions, and numerical tractability (the capability of simulating the model in reasonable time).

Range of model validity, calibration and further model validation. The increased prediction capability of RBA, or more generally of other resource allocation CBMs such as FBAwMC (Beg et al. 2007), ME-model (O'Brien et al. 2015) or Gecko-model (Sánchez, B. J., et al. 2017), compared to other CBM methods, necessitates a model of greater complexity, i.e. with a higher number of parameters, and in particular the in vivo efficiencies of metabolic enzymes and transporters. This raises the question of the range of validity of these parameters for different environmental conditions. Actually, metabolic enzyme or transporter efficiencies are usually either approximated by the maximal turnover rates of enzymes as done in Beg et al. 2007 or Sánchez, B. J., et al. 2017, and can be extracted (to some extent) from public repositories like the Brenda database (Schomburg et al. 2002), or estimated from omics data (especially quantitative proteomics and fluxomics) for different environmental conditions as was done for bacterial cells (Goelzer et al. 2015, Bulovic et al. 2019). Other parameters, such as the total protein content, the fraction of proteins dedicated to maintenance (parameter Pg) and the main metabolic stocks (parameter B), are usually estimated from quantitative proteomics, quantitative metabolomics and from the biomass composition (Goelzer et al. 2015). For bacteria, most of the estimated parameters were found to vary linearly with growth rate, which allowed the prediction of the parameter value as a function of the growth rate (Goelzer et al. 2015). Currently, the RBA model of the plant cell contains constant parameters compiled from the literature which, in spite of that, allows consistent and realistic predictions in standard non-limiting growth conditions. The calibration for the vegetative growth stage would necessitate dedicated datasets (composed of quantitative proteomics, quantitative metabolomics to determine the metabolic reserves, and the measurements of some fluxes such as the CO_2 or nitrate assimilation rates) that will be designed to explore the space of parameter values. It would help to determine the dynamics of parameters, especially the ones of molecular machine efficiencies, and if some parameters display a growth-rate dependent behavior or vary with the environment. Altogether, the model calibration will help to explore the range of model validity and robustness quantitatively.

A resource for detecting gaps in biological knowledge, checking the integrity of information and testing biological hypotheses. Adaptation in plant cells is a highly dynamic process that unfolds on multiple scales of time, space, and complexity. Predictive cell models remain imperfect, which is a sign that eukaryotic cell functioning is still not fully understood. The modeling of complex biological systems such as plant cells could help us see more clearly certain limits of biological knowledge and could point us to the molecular targets to be studied in priority. To do so, it is essential to have a fine molecular-scale description of how the plant cell system operates in different environments, as we propose in the RBA modeling framework. Our model is flexible and

is intended to be extended and improved as plant cellular biology progresses. Any new cellular function or cellular compartments can be added as long as their behavior can be described as linear constraints. In this way, metabolic costs or capacity of molecular machines associated with a new cellular function can be assessed and evaluated in the context of an entire cell, in order to check whether the newly added cellular function is compatible with cell growth and survival. Moreover, since the RBA model explicitly integrates costs associated with the import, use, and recycling of resources within the plant cell, several biological questions about cellular economics (e.g. the costs linked to protein or metabolite production and turnover) can be addressed with the model in non-limiting or limiting growth conditions and for different cell types. Finally, since RBA is a CBM method, biological assumptions on the plant cell behavior can be tested *in silico* as soon as they can be formalized as linear constraints. Their impact can be quantified, as we've done in this work for the inactivation of the NTT transporter, or for the use of the calvin benson cycle. We thus expect that the RBA model can be a useful resource for the plant scientific community.

Towards the prediction of the organelle volumes. Our RBA modeling framework for eukaryotic cells has been conceived to (i) predict cellular organelle volumes, based on coefficients that convert molecular machine abundances to volumes, and (ii) to ensure that the volume of organelles does not exceed the total cell volume (see Methods and Goelzer et al. 2019). The volume/density conversion coefficients are currently unknown parameters and call for further biological investigations on the biophysical properties of organelles. Thus, to account for a limited total cell volume, we assumed that the organelle volume is driven by the protein density of organelle, and that the sum of protein density of organelles does not exceed the total density of the cell. The current plant cell RBA model thus predicts the density of proteins allocated to the different organelles (see Fig.2FG), and predictions are consistent with experimental data (Li et al. 2017, Heinemann et al. 2021). Knowing coefficients converting the molecular machine abundance to volume, we expect that RBA could predict in the future the adaptation of organelles in complex environmental conditions in an autonomous way. However, this remains a challenging and open question, since the volume variation of some organelles like the vacuole can be large and may be driven by other biophysical constraints (such as ecophysiological constraints) and not by density variations only.

Towards the prediction of dynamical behavior such as day/night cycles. The model contains the known metabolic functions involved in the management of day/night cycle, such as starch or sucrose remobilization. In theory, the model could already simulate the leaf cell configurations during day/night cycles as it is done in the dynamical FBA (dFBA) literature. dFBA is an iterative process based on the division of time into small intervals into which the metabolism is assumed to be in steady-state, and consists of solving successive FBA problems, representing the metabolic state at different times (see Moulin et al. 2021 for a review). The current RBA model could already be simulated under dynamical conditions using the same type of algorithm. However, these simulations would neglect the inertia of macromolecule concentrations and of organelle volumes over time (also called memory effect in dynamical systems), since the configuration of the cell would be replayed entirely between time points. To include the memory effects, the RBA model would need to be extended for dynamical conditions as it was done in Jeanne et al 2018, or in Waldherr et al. 2015 for prokaryotic cells.

Model specialization to different cell types - the use of omics data. Beyond model calibration, the use of omics (quantitative proteomics, quantitative metabolomics and transcriptomics) technology contributes to the model specialization to different types of cells of organic tissues (Mintz-Oron et al. 2012, Aurich et al. 2016). RBA model specialization covers mainly three aspects: (i) the

metabolic coverage, and more broadly the coverage of cell processes; (ii) the major metabolic reserves; (iii) the specialization of the molecular machines (such as several isoenzymes catalyzing the same enzymatic reaction). The use of proteomics and transcriptomics data could help both to identify the cell processes accounting for a large part of resources in a specific cell type, and to determine the composition of the molecular machines, while metabolomics can be used straightforwardly to set metabolic reserves. RBA model specialization will first be an important step towards whole-plant modeling. Our RBA model currently describes a generic green leaf cell, representative of the rosette behavior with respect to its environment, i.e. the rosette is described as an homogenous population of our generic leaf cell type. Using cell-specialized RBA models, more complex organ or tissue composition could be described, by considering heterogeneous cell populations interacting and sharing resources. Cell-specialized models could further be interfaced with individual models to refine the description of organ functioning, such as (i) ecophysiological models to account for water and solute transport in the whole-plant and following the pioneer works for tomato (Chen et al. 2021) or CAM species (Töpfer et al. 2020) or (ii) functional-structural plant models such as Greenlab (Christophe et al. 2008) to account for the plant architecture and organogenesis. Additionally, cell-specialized models could be extremely helpful for studying plant organs of high agronomic interest, such as the seed, in order to define and forecast quality determinants under diverse environmental conditions. These insights will also be valuable in fine-tuning plant breeding programs.

Data availability

The RBA leaf model (encoded in XML files) and the PlantCellRBA software for running simulations are available at https://forgemia.inra.fr/anne.goelzer/rba-plant-cell-model.

Acknowledgements

We thank Wolfram Liebermeister, Ana Bulovic, Sophie Colombié and Jean-Denis Faure for critical comments on the manuscript and the Métaprogramme Digitbio of INRAE for funding.

Author Contributions

AG and VF conceived the study. AG developed, implemented and simulated the different models (RBA, FBA, Farquhar) and performed the model analysis (sensitivity, uniqueness of the solution, etc.). AG, FC, LR and OL analyzed the results. All authors wrote and approved the paper.

Conflict of interest

The authors declare no competing interest.

Supplementary material

Supplementary file 1. The RBA leaf model (encoded in XML files) and the PlantCellRBA software for running simulations are available at https://forgemia.inra.fr/anne.goelzer/rba-plant-cell-model.

Supplementary file 2. The updated AraCore metabolic model in SBML format, which was used as an input file to generate the RBA model and to perform FBA simulations.

Supplementary Table 1. (A) Assignment of functional categories for each metabolic reaction/enzymatic complex. **(B)** Predicted flux distribution for a leaf cell grown in standard environmental conditions at 21°C. **(C)** Uniqueness of the predicted flux distribution following the numerical test of Appa 2002. **(D)** FBA simulations for various constraints and optimality criteria. **(E)**

Description of model parameters (type, unit, references). **(F)** Comparison of the predicted flux distribution (B) with the fluxomics dataset of (Ma, et al. (2014)).

Supplementary figures. The Supplementary figures and their legends are compiled in a supplementary note.

References

Appa, G. (2002). On the uniqueness of solutions to linear programs. *Journal of the Operational Research Society*, *53*(10), 1127-1132.

Arnold, A., and Nikolovski, Z. (2014). Bottom-up metabolic reconstruction of Arabidopsis and its application to determining the metabolic costs of enzyme production. *Plant physiology*; 165.3: 1380-1391.

Arrivault, S., et al. (2019). Metabolite profiles reveal interspecific variation in operation of the Calvin–Benson cycle in both C4 and C3 plants. *Journal of Experimental Botany*, *70*(6), 1843-1858.

Aurich, M. K., et al. (2016). MetaboTools: a comprehensive toolbox for analysis of genome-scale metabolic models. *Frontiers in physiology*, 327.

Bar-Even, A., et al. (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, *50*(21), 4402-4410.

Beg, Q. K., et al. (2007). Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences*, *104*(31), 12663-12668.

Bodeit, O., et al. (2023). RBAtools: a programming interface for Resource Balance Analysis models. *Bioinformatics Advances*, vbad056.

Bouchabke, et al. (2008). Natural variation in Arabidopsis thaliana as a tool for highlighting differential drought responses. *PloS one*, *3*(2), e1705.

Bulovic, A., et al. (2019). Automated generation of bacterial resource allocation models. *Metabolic engineering*, *55*, 12-22.

Caemmerer, S., et al. (2009). Biochemical model of C_3 photosynthesis. In *Photosynthesis in silico* (pp. 209-230). Springer, Dordrecht.

Chen, J., et al. (2021). Modelling predicts tomatoes can be bigger and sweeter if biophysical factors and transmembrane transports are fine-tuned during fruit development. *New Phytologist*, *230*(4), 1489-1502.

Cheung, et al. (2013). A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. *The plant journal*, *75*(6), 1050-1061

Cheung, et al. (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*(2), 917-929..

Chew, et al. (2014). Multiscale digital Arabidopsis predicts individual organ and whole-organism growth. *Proceedings of the National Academy of Sciences*. 111(39):E4127-E4136.

Chietera, G., et al. (2018). Impact of the genetic–environment interaction on the dynamic of nitrogen pools in arabidopsis. *Agriculture* 8.2: 28.

Christophe, A., et al. (2008). A model-based analysis of the dynamics of carbon balance at the whole-plant level in Arabidopsis thaliana. *Functional plant biology*, *35*(11), 1147-1162.

de Reffye, P., et al. (2021). Two decades of research with the GreenLab model in Agronomy. *Annals of botany*, *127*(3), 281-295.

de Oliveira Dal'Molin, C. G., et al. (2010). AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant physiology*, *152*(2), 579-589.

de Oliveira Dal'Molin, et al. (2015). A multi-tissue genome-scale metabolic modeling framework for the analysis of whole plant systems. *Frontiers in plant science*, *6*, 4.

Farquhar G, et al. (1980). A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. *Planta*, 149.1: 78-90.

Feist, A., and Palsson, B. (2010). The biomass objective function. *Current opinion in microbiology*, *13*(3), 344-349.

Flügge, U. I., et al. (2011). The role of transporters in supplying energy to plant plastids. *Journal of experimental botany*, *62*(7), 2381-2392.

Gerlin, L., et al. (2022). A multi-organ metabolic model of tomato predicts plant responses to nutritional and genetic perturbations. *Plant Physiology*, *188*(3), 1709-1723.

Ghifari, A. S., et al. (2018). Plant mitochondrial protein import: the ins and outs. *Biochemical Journal*, *475*(13), 2191-2208.

Gloaguen, P., et al. (2017). ChloroKB: a web application for the integration of knowledge related to chloroplast metabolic network. *Plant physiology*, *174*(2), 922-934

Goelzer, A., et al. (2009). Cell design in bacteria as a convex optimization problem controller. *Proceedings of the 48h IEEE Conference on Decision and Control (CDC),* 4517-4522.

Goelzer, A., et al. (2011). Cell design in bacteria as a convex optimization problem. *Automatica*, *47*(6), 1210-1218.

Goelzer, A. et al. (2015). Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic engineering*, *32*, 232-243.

Goelzer A, and Fromion, V. (2017). Resource allocation in living organisms. *Biochemical Society Transactions*; 45 (4): 945-952

Goelzer, A., and Fromion, V. (2019). RBA for eukaryotic cells: foundations and theoretical developments. *bioRxiv*, 750182.

Grafahrend-Belau, E., et al. (2013). Multiscale metabolic modeling: dynamic flux balance analysis on a whole-plant scale. *Plant physiology*, *163*(2), 637-647.

Heinemann, B., et al. (2021). Estimating the number of protein molecules in a plant cell: protein and amino acid homeostasis during drought. *Plant physiology*, *185*(2), 385-404.

Henriet, C., et al. (2021). Proteomics of developing pea seeds reveals a complex antioxidant network underlying the response to sulfur deficiency and water stress. *Journal of Experimental Botany*, *72*(7), 2611-2626.

Hucka, M., et al. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, *19*(4), 524-531.

Ikram, S., et al. (2012). Natural variation of Arabidopsis response to nitrogen availability. *Journal of Experimental Botany*, *63*(1), 91-105.

Jeanne, G., et al. (2018). Dynamical resource allocation models for bioreactor optimization. *IFAC-PapersOnLine*, *51*(19), 20-23.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27-30.

Kölling, K., et al. (2015). A whole-plant chamber system for parallel gas exchange measurements of Arabidopsis and other herbaceous species. *Plant Methods*, *11*, 1-12.

Kurek, I., et al. (2007). Enhanced thermostability of Arabidopsis RuBisCO activase improves photosynthesis and growth rates under moderate heat stress. *The Plant Cell*, *19*(10), 3230-3241.

Li, L., et al. (2017). Protein degradation rate in Arabidopsis thaliana leaf growth and development. *The Plant Cell*; 29.2: 207-228

Lim, S. L., et al (2022). Arabidopsis guard cell chloroplasts import cytosolic ATP for starch turnover and stomatal opening. *Nature Communications*, *13*(1), 652.

Ma, F., et al. (2014). Isotopically nonstationary 13C flux analysis of changes in Arabidopsis thaliana leaf metabolism due to high light acclimation. *Proceedings of the National Academy of Sciences*, *111*(47), 16967-16972.

Menz, J., et al. (2018). Molecular basis of differential nitrogen use efficiencies and nitrogen source preferences in contrasting Arabidopsis accessions. *Scientific Reports*, *8*(1), 1-11.

Mintz-Oron, S., et al. (2012). Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proceedings of the National Academy of Sciences*, *109*(1), 339-344.

Moulin, C. et al. (2021). Combining kinetic and constraint-based modelling to better understand metabolism dynamics. *Processes*, *9*(10), 1701.

Nobel, P. S. (1999). *Physicochemical & environmental plant physiology*. Academic press.

Nesterov, Y. (2012). Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM J Optimiz*; 22(2):341-362. 2012.

Nesterov, Y., et al. (1994). Interior-point polynomial algorithms in convex programming. *SIAM*.

O'Brien, E. J., et al. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular systems biology*, *9*(1), 693.

Piques, M., et al. (2009). Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. *Molecular systems biology*, *5*(1), 314.

Poolman, M.G., et al. (2009). Plant Physiol. 2009; 151, 1570–1581. https://doi.org/10.1104/pp.109.141267

Pyl, E.-T, et al. (2012). Metabolism and growth in Arabidopsis depend on the daytime temperature but are temperature-compensated against cool nights. *The Plant Cell*; 24.6: 2443-2469.

Sánchez, B. J., et al. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular systems biology*, *13*(8), 935.

Schomburg, I., et al. (2002). BRENDA: a resource for enzyme data and metabolic information. *Trends in biochemical sciences*, *27*(1), 54-56.

Schulz, E., et al. (2021). Natural variation among Arabidopsis accessions in the regulation of flavonoid metabolism and stress gene expression by combined UV radiation and cold. *Plant and Cell Physiology*, *62*(3), 502-514.

Seaver, S. M., et al. (2014). High-throughput comparison, functional annotation, and metabolic modeling of plant genomes using the PlantSEED resource. *Proceedings of the National Academy of Sciences*, *111*(26), 9645-9650.

Schulze, W., et al. (1991). A quantification of the significance of assimilatory starch for growth of Arabidopsis thaliana L. Heynh. *Plant Physiology*, *95*(3), 890-895.

Scott, M., et al. (2010). Interdependence of cell growth and gene expression: origins and consequences. *Science*, *330*(6007), 1099-1102.

Shameer, S., et al. (2018). Computational analysis of the productivity potential of CAM. *Nature Plants*, *4*(3), 165-171.

Shameer, S., et al. (2020). Flux balance analysis of metabolism during growth by osmotic cell expansion and its application to tomato fruits. *The Plant Journal*, *103*(1), 68-82

Shaw, R., Cheung, C. M. (2018). A dynamic multi-tissue flux balance model captures carbon and nitrogen metabolism and optimal resource partitioning during Arabidopsis growth. *Frontiers in plant science*, *9*, 884.

Sulpice, R., et al. (2014). Arabidopsis coordinates the diurnal regulation of carbon allocation and growth across a wide range of photoperiods. *Molecular Plant*, *7*(1), 137-155.

Tan, X., and Cheung, C. (2020). A multiphase flux balance model reveals flexibility of central carbon metabolism in guard cells of C3 plants. *The Plant Journal*, *104*(6), 1648-1656.

Thiele, I., and Palsson, B. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, *5*(1), 93-121.

Thomson, S. M., el al. (2020). Protein import into chloroplasts and its regulation by the ubiquitin-proteasome system. *Biochemical Society Transactions*, *48*(1), 71-82.

Tisné S, et al. (2013). Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity. *The Plant Journal*; 74(3): 534-544.

Töpfer, N., et al. (2020). Alternative crassulacean acid metabolism modes provide environment-specific water-saving benefits in a leaf metabolic model. *The Plant Cell*, *32*(12), 3689-3705.

Tournier, L., et al. (2017). Optimal resource allocation enables mathematical exploration of microbial metabolic configurations. *Journal of mathematical biology*, *75*(6), 1349-1380.

Tuzet, A., et al. (2003). A coupled model of stomatal conductance, photosynthesis and transpiration. *Plant, Cell & Environment*, 26(7):1097-1116.

UniProt consortium (2017). UniProt: the universal protein knowledgebase. *Nucleic acids research*, vol. 45, no D1, p. D158-D169

Varma, A., and Palsson, B. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Applied and environmental microbiology*. 60(10): 3724-3731.

Voon, C. P., and Lim, B. L. (2019). ATP translocation and chloroplast biology. *National Science Review*, *6*(6), 1073-1076.

Waldherr, S., et al. (2015). Dynamic optimization of metabolic networks coupled with gene expression. *Journal of theoretical biology*, *365*, 469-485.

Walker, B., et al. (2013). Temperature response of in vivo Rubisco kinetics and mesophyll conductance in Arabidopsis thaliana: comparisons to Nicotiana tabacum. *Plant, Cell & Environment*; 36.12: 2108-2119.

Yin, X., et al. (2021). On the needs for combining physiological principles and mathematics to improve crop models. *Field Crops Research*, *271*, 108254.

Zandalinas, S. I., et al. (2021). The impact of multifactorial stress combination on plant growth and survival. *New Phytologist*, *230*(3), 1034-1048.