Towards the integration of cellular scales in individual plant models

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

anne.goelzer@inrae.fr

Systems Biology team
Applied Mathematics and Computer Science,
from Genomes to the Environment
INRAE Jouy-en-Josas, FRANCE

INRAe

MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES
MaiAGE
DU GÉNOME À L'ENVIRONNEMENT

30/11/2022

$$\sum_k C_k = C_T$$

$$\sum_k N_k = N_T$$

Plant models at the individual scale

Biomass allocation models by source-sink empirical relations

Ecophysiological models of water flow and thermic budgets

When properly calibrated on a dataset, such models show a high capability of prediction

But the range of validity of these models is restricted to the conditions where the datasets were acquired.
For instance, the plant adaptation to complex conditions (e.g. combined stress conditions) is not well predicted

Multiscale digital *Arabidopsis* predicts individual organ and whole-organism growth

Yin Hoon Chew[a], Bénédicte Wenden[b], Anna Flis[c], Virginie Mengin[c], Jasper Taylor[d], Christopher L. Davey[e], Christopher Tindal[a], Howard Thomas[e], Helen J. Ougham[e], Philippe de Reffye[f], Mark Stitt[c], Mathew Williams[g], Robert Muetzelfeldt[d], Karen J. Halliday[a], and Andrew J. Millar[a,1]

[a]SynthSys and School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JD, United Kingdom; [b]Institut National de la Recherche Agronomique and Université Bordeaux, Unité Mixte de Recherche 1332 de Biologie du Fruit et Pathologie, F-33140 Villenave d'Ornon, France; [c]Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany; [d]Simulistics Ltd., Loanhead EH20 9PA, United Kingdom; [e]Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth SY23 2FG, United Kingdom; [f]Cirad-Amis, Unité Mixte de Recherche, botAnique et bioinforMatique de l'Architecture des Plantes, F-34398 Montpellier Cedex 5, France; and [g]School of GeoSciences, University of Edinburgh, Edinburgh EH9 3JN, United Kingdom

Four interconnected (existing) models
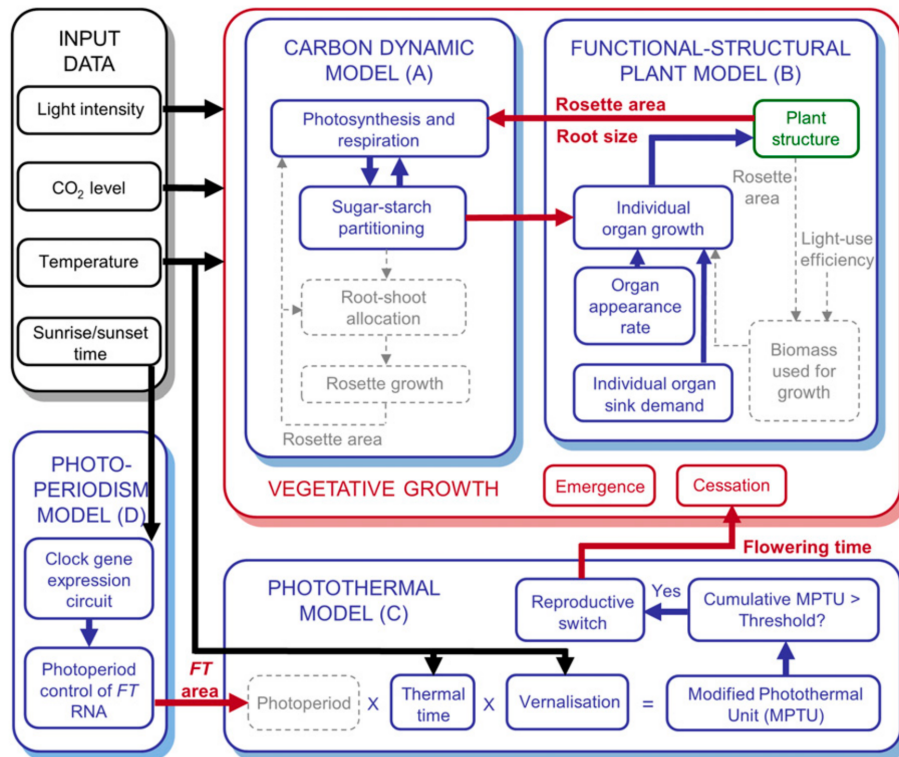
Carbon dynamic model : based on the model of Farquhar
- account for C fixation only
- scaled by the rubisco content and respiratory rates

FSPM: based on Greenlab

→ Need to develop more autonomous models at cellular scales to cope with single and combined abiotic stress

Rasse DP, Tocquin P (2006) New Phytol 172(3):500–513.
Christophe et al. (2008). Functional Plant Biology 35(11) 1147-1162

Adaptation to complex conditions occurs at the cellular scale

$$\sum_k C_k = C_T$$

$$\sum_k N_k = N_T$$
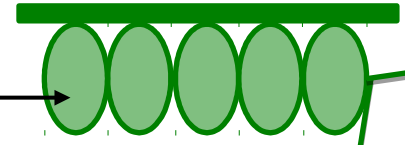
Refine the description of cellular scales in individual models

Resource balance analysis (RBA) models are particularly promising ...

C_flow, N_flow
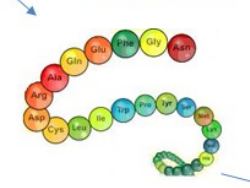
C_leaf, N_leaf

C_root, N_root

Mesophyll cells

Transport
De novo synthesis by metabolic network
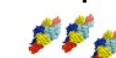
Amino acids

Trp

Asn

Asp

Ribosomes

ATP, GTP

Polypeptide

Transport
De novo synthesis by metabolic network

Ions, vitamins

Chaperones

Active protein
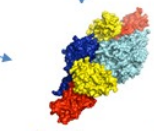
En régime permanent
(constant growth rate)

Resource Balance Analysis in a nutshell

Metabolic processes

$$Sv = 0$$

Stoichiometric matrix S is first extracted from the genome-scale metabolic model.

$$Sv - \mu v_P - \mu v_E + \mu v_B = 0$$

Metabolism additionally needs to be able to produce sufficient **P**recursors and **E**nergy carriers, and to recycle the metabolic **B**yproducts of different cellular processes.

Cellular processes

Mass conservation constraint C_1

Macromolecular processes

Ribosome   Secretion apparatus   Chaperone

Process machineries

Structure and efficiency are defined in the same way as for an enzyme.

Each macromolecule can be a substrate of a number of processes. New processes can easily be added to the model.

Proteins   Processes
Translation
Secretion
Chaperoning

Process substrates

Process capacity constraint C_2

For a particular growth rate, abundance of the processing machinery needs to be sufficient to process all of its substrates:

$$\mu P_T \le k_T R$$
$$\mu P_S \le k_S S$$
$$\mu P_C \le k_C C$$

Metabolic machineries

Enzymes / Transporters

Macromolecular Composition

2 x protein P1   1 x protein P2   1 ion

Enzyme can be composed of macromolecules (protein, RNA, ...), ions and cofactors.

$$k_{app} = C$$
$$k_{app} = a\mu + b$$
$$k_{app} = \frac{V_{max}[S]}{K_M + [S]}$$

Apparent catalytic rate can be either a constant, linear function of the growth rate, or a Michaelis-Menten function of an external metabolite concentration.
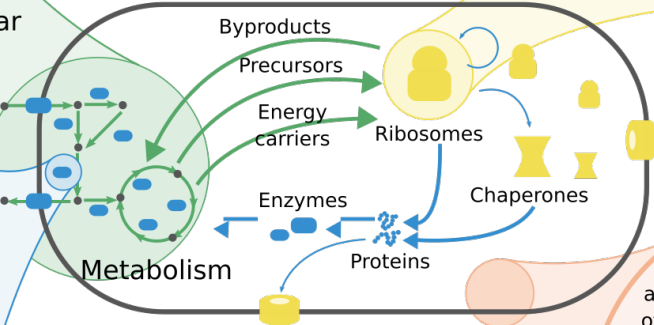
$$-k_{app_{i,\leftarrow}} E_i \le v_i \le k_{app_{i,\rightarrow}} E_i$$

Flux through a reaction is limited by the efficiency and abundance of an enzyme.

Enzyme capacity constraint C_2

Cell growing at rate μ

Byproducts
Precursors
Energy carriers
Ribosomes

Metabolism

Enzymes
Chaperones

Proteins

Compartments

Each compartment has an upper limit on the amount of **M**acromolecules that it can accomodate.

Periplasm

Cytoplasm

Compartment density constraint C_3

For each compartment X:
$$\sum_{i \in X} M_i \le D_X(\mu)$$

The RBA optimization problem

Concentration of housekeeping proteins

Growth rate

Concentration of molecular machines

Metabolic fluxes

For fixed $P_G \geq 0$, $\mu \geq 0$,

$$P_{rba}(\mu): \quad \text{find} \quad Y \in \mathbb{R}^{m+p}_{\geq 0}, \nu \in \mathbb{R}^m,$$

subject to

$$(C_1) \quad -\Omega\nu + \mu(C_Y^S Y + C_B^S \bar{B} + C_G^S P_G) = 0$$
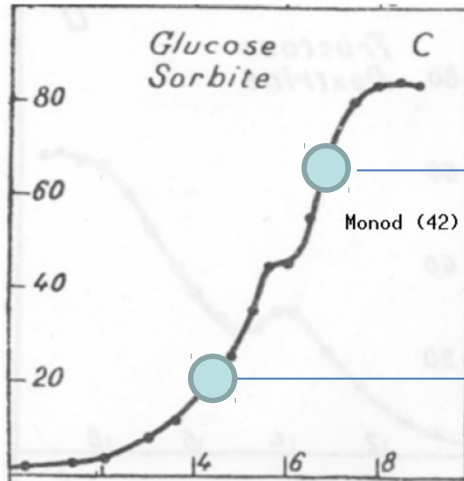
$$(C_{2a}) \quad \mu(C_Y^M Y + C_G^M P_G) - K_T Y \leq 0$$

$$(C_{2b}) \quad -K'_E Y \leq \nu \leq K_E Y$$

$$(C_3) \quad C_Y^D Y + C_G^D P_G - \bar{D} \leq 0$$

A constraint-based modeling method
- One constraint of mass conservation per intracellular metabolite
- One capacity constraint per molecular machine
- One density constraint per cellular compartment

The RBA optimization problem

Concentration of housekeeping proteins

Growth rate

Concentration of molecular machines

Metabolic fluxes

For fixed $P_G \geq 0$, $\mu \geq 0$,

$$\boldsymbol{P}_{rba}(\boldsymbol{\mu}): \quad \text{find} \qquad Y \in \mathbb{R}^{m+p}_{\geq 0}, \nu \in \mathbb{R}^m,$$

subject to

$(C_1) \qquad -\Omega\nu + \mu(C^S_Y Y + C^S_B \bar{B} + C^S_G P_G) = 0$

$(C_{2a}) \qquad \mu(C^M_Y Y + C^M_G P_G) - K_T Y \leq 0$

$(C_{2b}) \qquad -K'_E Y \leq \nu \leq K_E Y$

$$K_E(S_{ext}) = \frac{k_{max} S_{ext}}{K_m + S_{ext}}$$

$(C_3) \qquad C^D_Y Y + C^D_G P_G - \bar{D} \leq 0$

The extracellular medium is taken into account through the efficiencies of nutrient transporters located in the plasmic membrane (extracellular concentration of nutrients)

The RBA optimization problem

Concentration of housekeeping proteins

Growth rate

Concentration of molecular machines

Metabolic fluxes

For fixed $P_G \geq 0$, $\mu \geq 0$,

$P_{rba}(\mu)$:   find $Y \in \mathbb{R}^{m+p}_{\geq 0}, \nu \in \mathbb{R}^m,$
        subject to
(C_1) $-\Omega\nu + \mu(C_Y^S Y + C_B^S \bar{B} + C_G^S P_G) = 0$

(C_{2a}) $\mu(C_Y^M Y + C_G^M P_G) - K_T Y \leq 0$

(C_{2b}) $-K_E' Y \leq \nu \leq K_E Y$

(C_3) $C_Y^D Y + C_G^D P_G - \bar{D} \leq 0$

- At fixed μ, the RBA problem is a linear programming (LP) optimization problem
- There exists a maximal growth rate μ* such that for all μ lower than μ*, P$_{rba}$(μ) is feasible, and unfeasible for higher values
- The maximum μ* is computed by solving a series of LP optimization problem for different μ values
- At μ*, we compute the concentration of molecular machines and flux distribution

A. Goelzer , V. Fromion and G. Scorletti. *48th IEEE Conference on Decision and Control*, China, 4517 -22. **2009**.
A. Goelzer , V. Fromion and G. Scorletti. *Automatica*, 47(6):1210-1218. **2011**

Monod experiments on bacteria

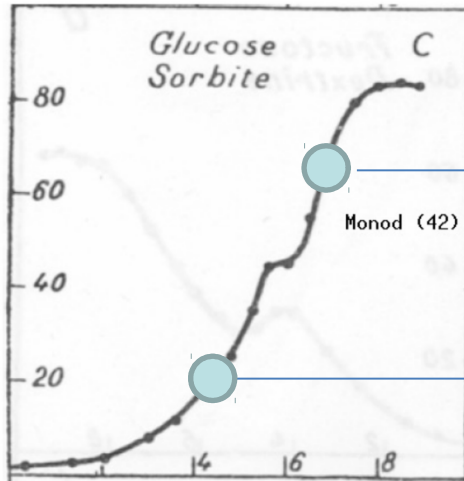| Predicted use of carbon sources by RBA	Predicted use of carbon sources by other constraint-based modeling methods (FBA)	
Free config.	Free config.	Fixed config.
Sorbitol	Sorbitol + Glucose	Sorbitol
Glucose	Sorbitol + Glucose	Glucose

RBA predicts a hierarchy in the use of carbon sources in agreement with Monod's experiments

Autonomous prediction : the cell configuration is not fixed *a priori*

Monod experiments on bacteria

	Predicted use of carbon sources by RBA	Predicted use of carbon sources by other constraint-based modeling methods (FBA)	
	Free config.	Free config.	Fixed config.
Sorbitol	Sorbitol	Sorbitol + Glucose	Sorbitol
Glucose	Glucose	Sorbitol + Glucose	Glucose

RBA predicts a hierarchy in the use of carbon sources in agreement with Monod's experiments

Autonomous prediction **:** the cell configuration is not fixed *a priori*

Promising to predict the cell phenotype in combined stress conditions

L. Tournier et al. *Journal of mathematical biology*, 75(6), 1349-1380. **2017**

The RBA framework
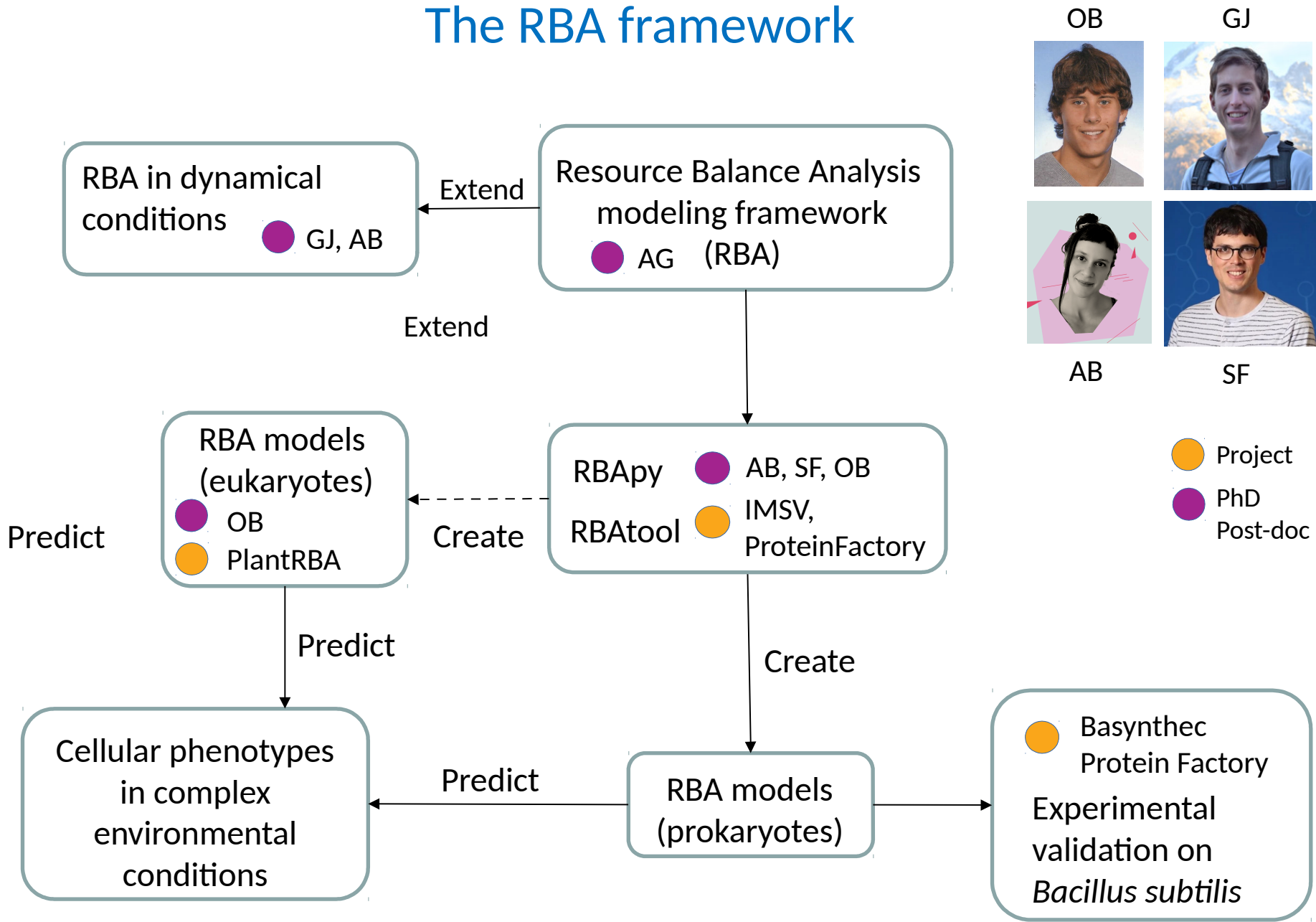
How to calibrate such kind of model ?

The RBA optimization problem

Concentration of housekeeping proteins

Growth rate

Concentration of molecular machines

Metabolic fluxes

For fixed $P_G \geq 0$, $\mu \geq 0$,

$P_{rba}(\mu)$: find $\qquad Y \in \mathbb{R}_{\geq 0}^{m+p}, \nu \in \mathbb{R}^m,$

subject to

$(C_1) \qquad -\Omega\nu + \mu(C_Y^S Y + C_B^S \bar{B} + C_G^S P_G) = 0$

$(C_{2a}) \qquad \mu(C_Y^M Y + C_G^M P_G) - K_T Y \leq 0$

$(C_{2b}) \qquad -K_E' Y \leq \nu \leq K_E Y$

$(C_3) \qquad C_Y^D Y + C_G^D P_G - \bar{D} \leq 0$

Stoichiometry of chemical reactions

Protein sequence Annotation

Literature
Biomass composition
Metabolomics

Estimation on omics datasets

Minimal medium Pyruvate (PYR)	Minimal medium Glucose, Citrate (S)	Medium S + Glutamate (TS)	Complex medium 18 amino acids (CH)	Medium CH + Glucose (CHG)
μ: 0,3 h^{-1}	μ: 0,6 h^{-1}	μ:0,9 h^{-1}	μ: 1,1 h^{-1}	μ:1,5 h^{-1}
797 proteins	1127 proteins	1158 proteins	984 proteins	1103 proteins
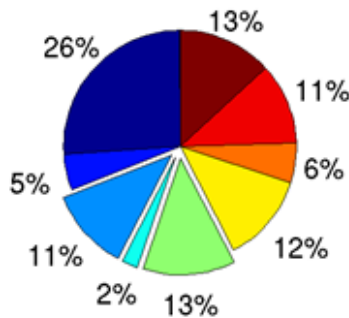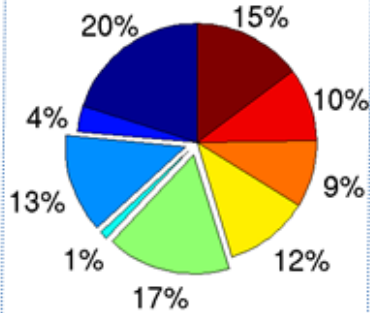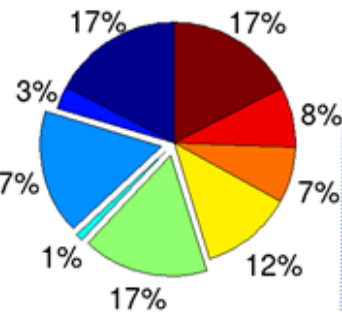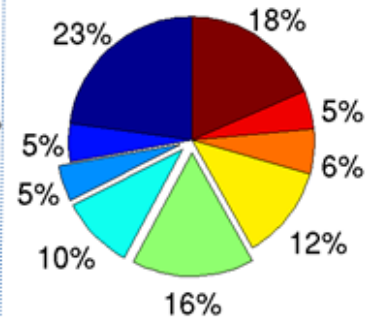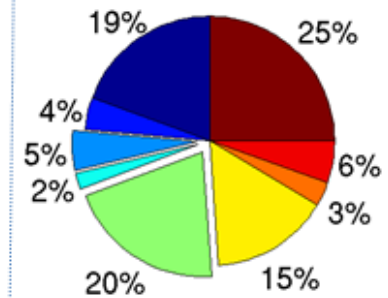
Legend:
- ■ Central Carbon Metabolism
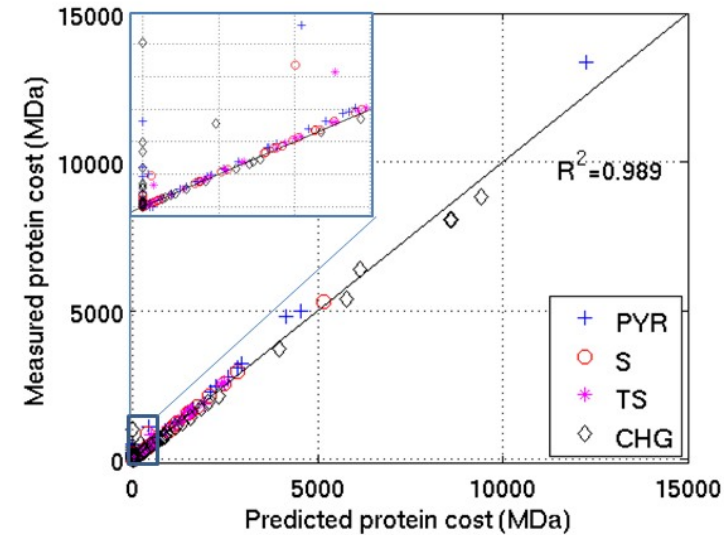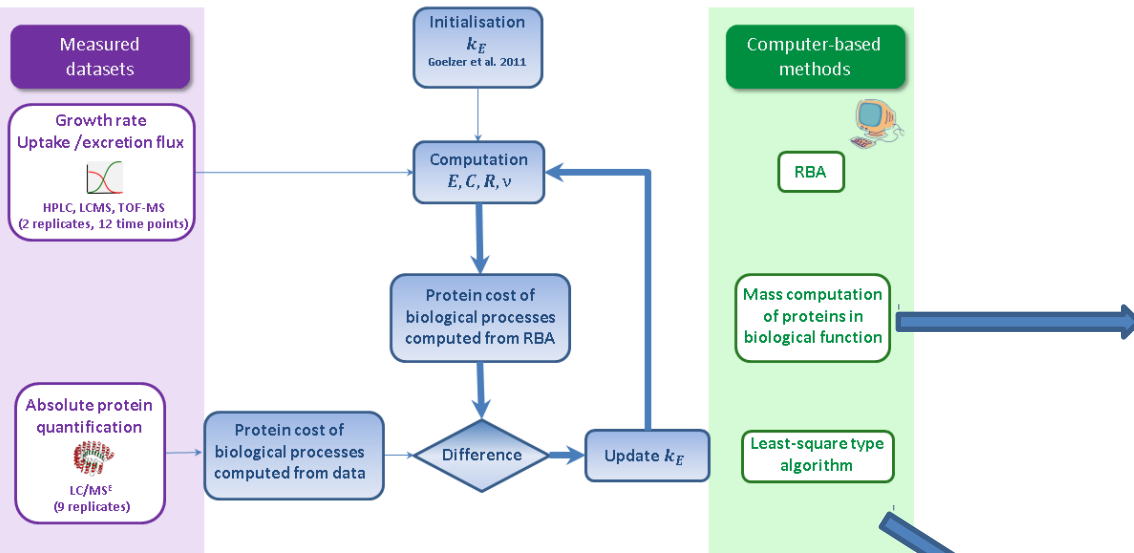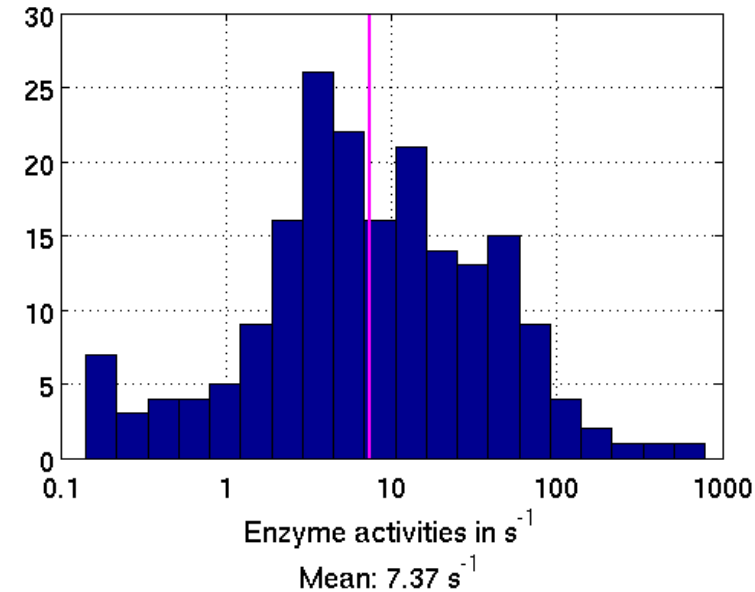- ■ Respiration + ATPase
- ■ Amino-acids synthesis
- ■ Amino-acids degradation
- ■ Other metabolic pathways
- ■ Neither translational nor metabolic
- ■ Motility/chemotaxis/flagella
- ■ Unclassified proteins
- ■ Translation Apparatus
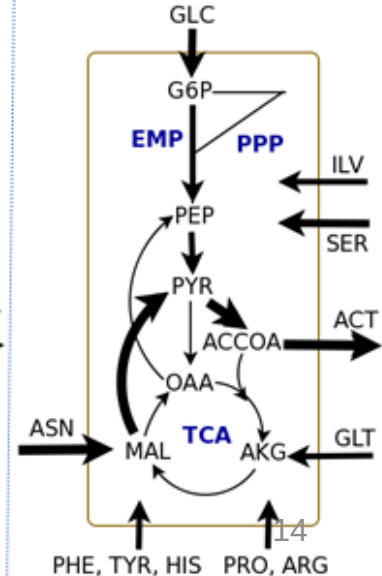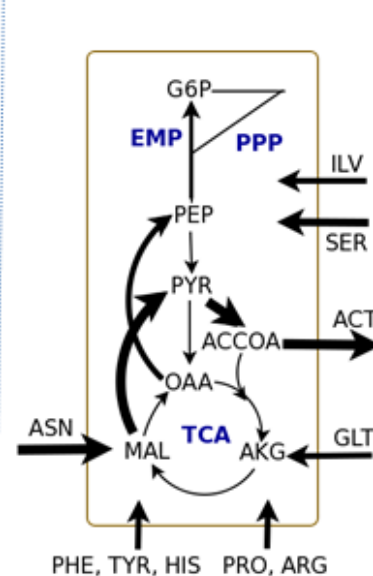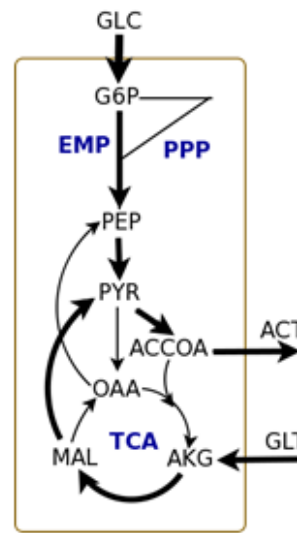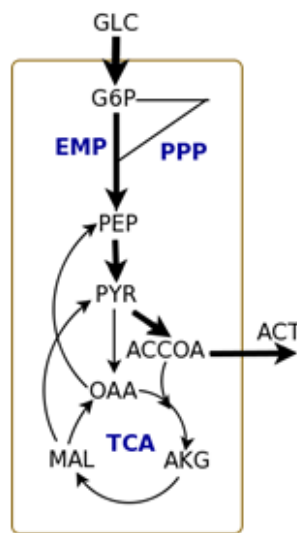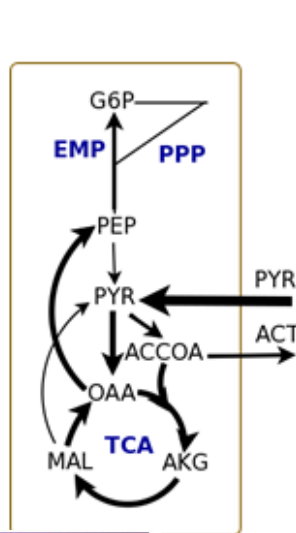
Estimation of parameters on omics datasets
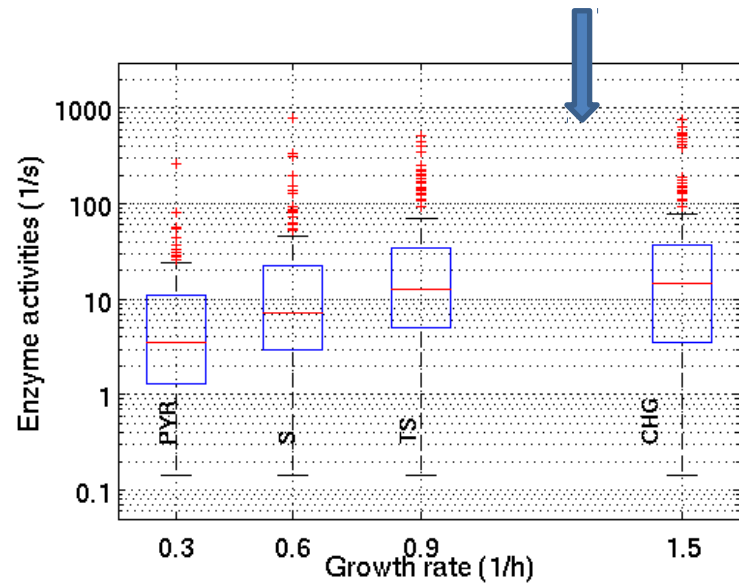
Identification of in vivo efficiencies of 600 enzymes
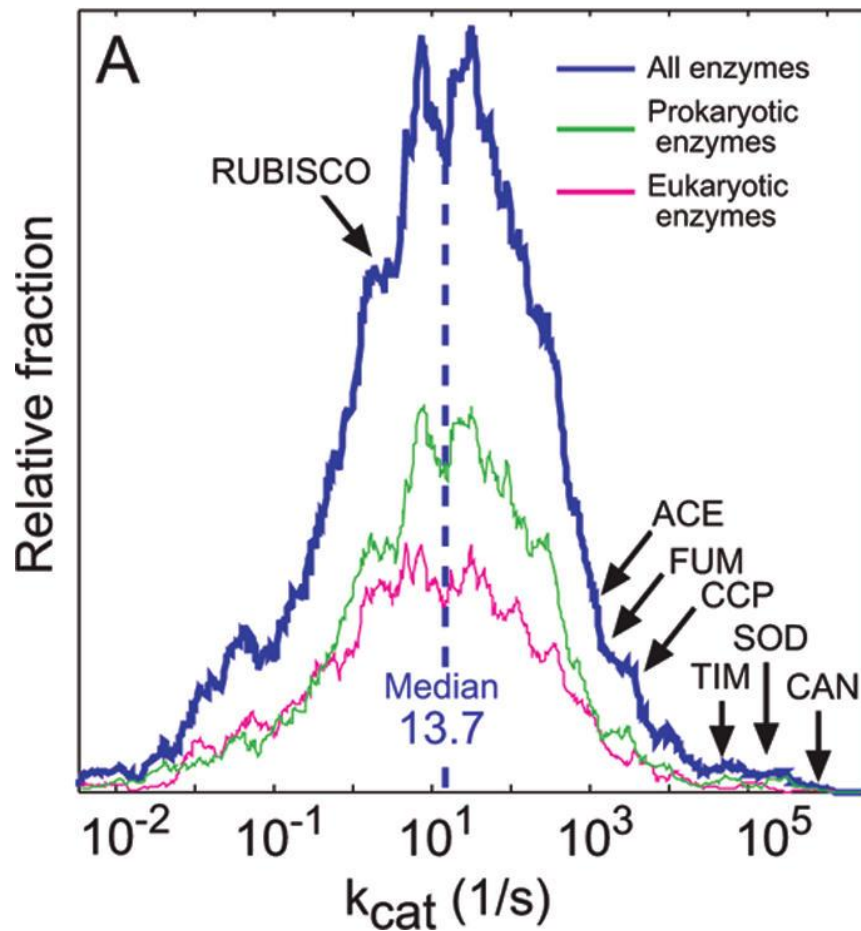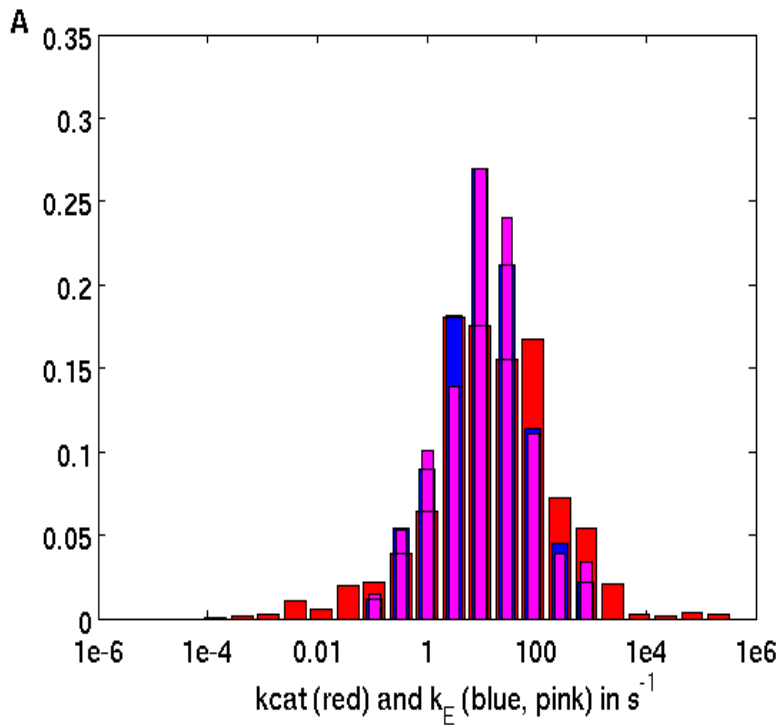
A. Bar-Even, et al. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters, Biochemistry, 2011, 50 (21), pp. 4402–4410

Prediction of resource allocation for the cellular processes of *Bacillus subtilis*

And for the plant cell ?

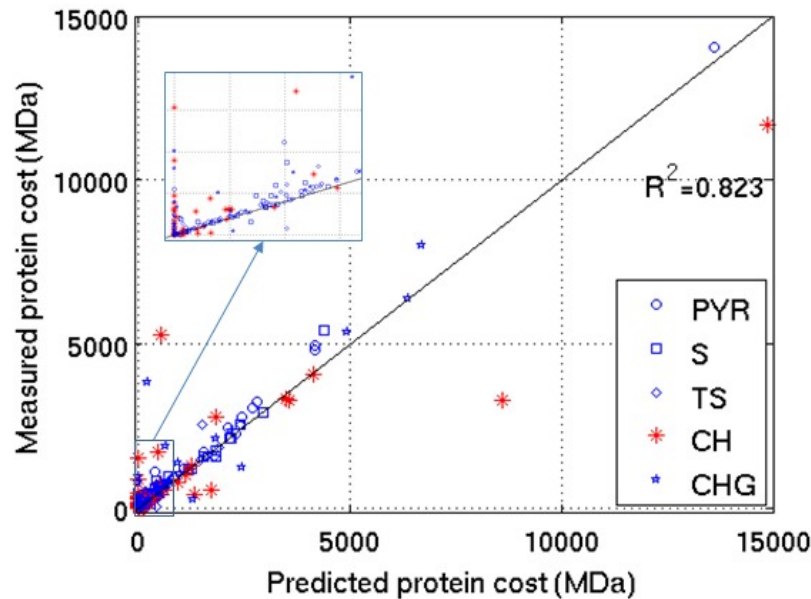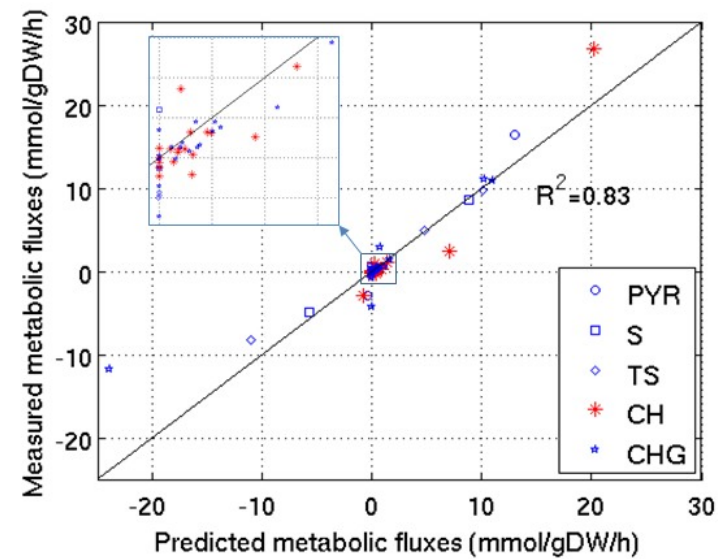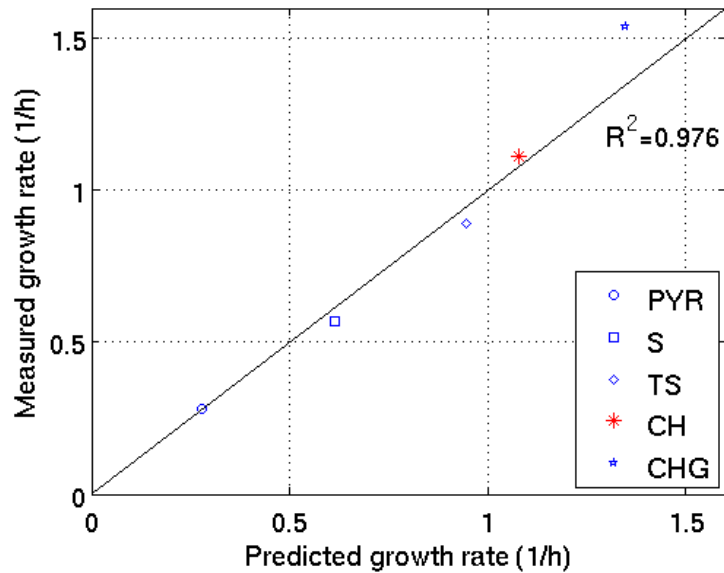# A RBA model for the mesophyll cell

<u>Objective :</u> Investigate if the principle of parcimonious resource allocation between cellular processes (i.e. maximize the relative growth rate) leads to realistic cellular phenotypes.

B. Rosette RBA model

Systemic description
Linear constraints

A. Knowledge & data

Public repositories
Literature
Expertise
Phenotypic data
Omics data
...
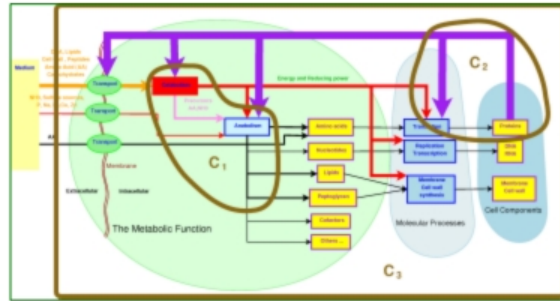
C. Model simulation

Assumption : parcimonious resource allocation between cellular processes
→ relative growth rate maximisation

$$\underset{Y\in\mathbb{R}^{N_y}_{\geq0},\,\nu\in\mathbb{R}^{N_m},\,f\in\mathbb{R}^{N_c}_{\geq0}}{\text{maximize}} \quad \mu\in\mathbb{R}_{\geq0},$$

$$(C_1) \qquad -\Omega\nu + \mu(C_Y^S Y + C_G^S P_G + C_B^S \bar{B} + C_F^S f\hat{B}) = 0$$

$$(C_2) \qquad \mu(C_Y^M Y + C_G^M P_G) - K_T Y \leq 0$$

$$-K_E' Y \leq \nu \leq K_E Y$$

$$(C_3) \qquad C_Y^D Y + C_G^D P_G - C_F^D f \leq 0$$

$$C_F^F f - \bar{C} = 0$$

$$\underline{f}_V \leq I_V f \leq \bar{f}_V$$
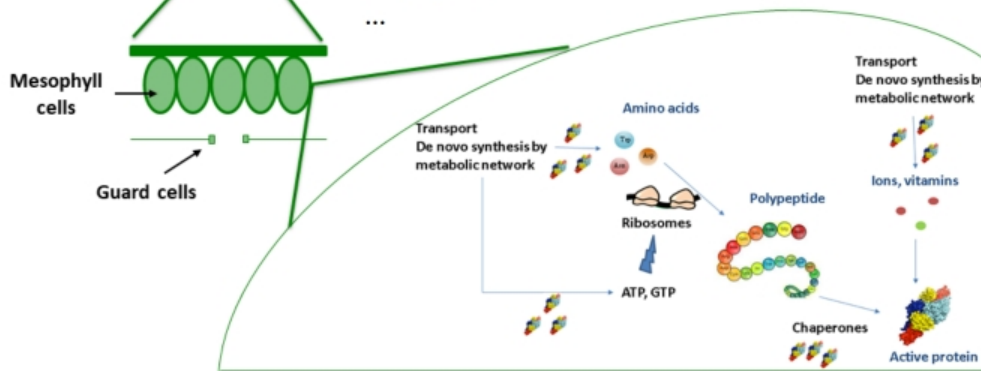
D. Biological prediction for an environmental condition

❖ Relative growth rate
❖ Quantitative traits as the C:N ratio
❖ Protein distribution among cellular compartments
❖ Protein abundances including ion and vitamin content for activity
❖ Metabolic fluxes including
    ✓ Rubisco activity
    ✓ Exchange fluxes (CO_2 assimilation rate, ...)

Cellular processes in the RBA model

Cellular compartments. cytoplasm, nucleus, mitochondrion, chloroplast, thylakoid, vacuole, peroxisome, golgi, endoplasmic reticulum, extracellular matrix. Fine description integrating organelle membranes.

Non-metabolic processes (NMP)

- Replication : nucleus, mitochondrion, chloroplast
- Transcription : nucleus, mitochondrion, chloroplast
- Translation: cytoplasm, mitochondrion, chloroplast
- Folding : cytoplasm, mitochondrion, chloroplast
- Protein, RNA translocation : nucleus, mitochondrion, chloroplast
- RNA degradation : cytoplasm
- Protein degradation : cytoplasm

⎫ To be refined

➡ Major processes already integrated

Metabolic processes (or reactions) from AraCore

- Photosynthesis
- Cavin cycle, photorespiration, gluconeogenesis/ glycolysis, PPP
- TCA cycle, oxidative phosphorylation
- Starch synthesis/degradation, sucrose synthesis/ degradation
- Amino acid, (deoxy)-nucleotide synthesis, nucleotide salvage pathways
- Chlorophyll A and B, Riboflavin, NAD(P), Pyridoxal-5P syntheses
- Transport of metabolites in/out of the cell and of organelles
- Fatty-acids : aggregated on Malonyl-coA
- Cell wall : cellulose, aggregated on dimers of glucose

Total: 728 metabolites, 1594 metabolic reactions, 28 NMP, 1330 proteins
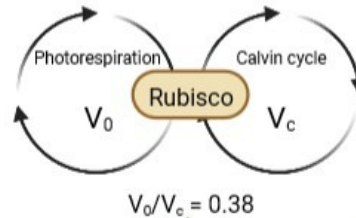
AraCore : Arnold et al. Plant Physiology. 2014; 165:1380-1391

Parameters only from literature and published datasets

Input/Output flows

Species	Flux (mM.gDW-1.day-1)
Photons	78.8
CO2	4.70
H2O	2.88
Pi	0.003
NO3	0.19
NH4	0.44
H2S	0.002
Mg2	0.0043
Fe2	1.22e-04
Mn2	5.52e-07
Cob2	1.36e-05
Zn2	2.63e-06
Cu	7.03e-06
Ca2	8.69e-09
K	6.41e-09
O2	-5.02

Rubisco rate

$V_0/V_c = 0.38$

Chlorophyll content

Measured (Sulpice et al. 2014) Predicted

Protein weight

Measured (Li et al. 2017) Predicted
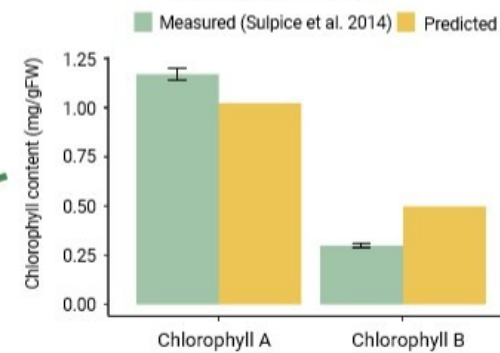
Compartment occupancy in plant cells

Cell membrane
Cytoplasm
Mitochondrion outer membrane
Mitochondrion matrix
Mitochondrion interspace
Mitochondrion inner membranne
Nucleus
Nucleus membrane
Peroxisome
Peroxisome membrane
Chloroplast stroma
Chloroplast membrane
Thylakoid membrane
Thylakoid lumen
Secreted

Ribosome content

Cytoplasmic ribosome   Mitochondrial ribosome   Chloroplastic ribosome

28.1%
25.5%
2.2%
2.7%
72.3%
69.2%

measured   predicted
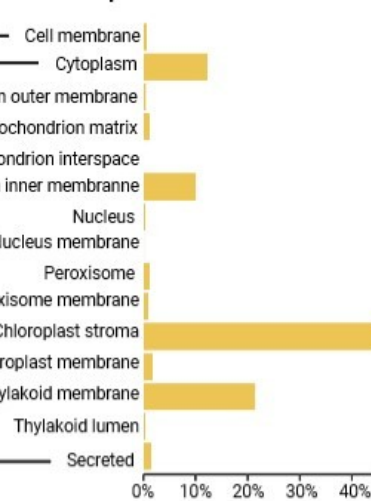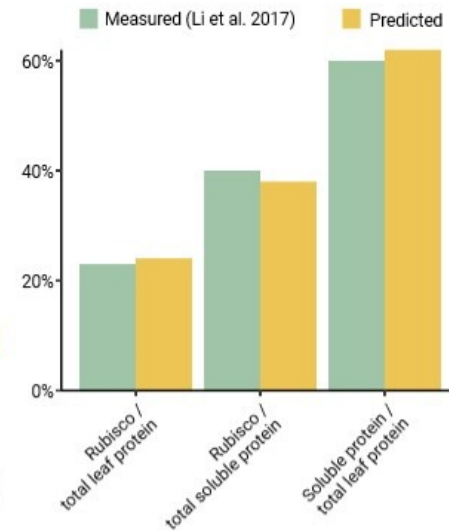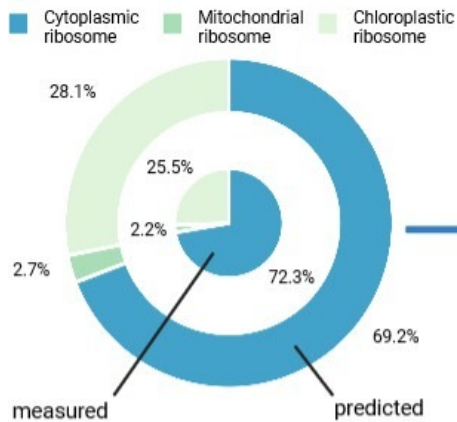
Quantitative trait
RGR: 0.2 day⁻¹
C:N ratio: 7.49

Predictions and available data are quantitatively consistent

21

Sensitivity analysis in non-limiting growth conditions

Relative variation of predicted growth rate (%) when the parameter is two-times increased (right) or decreased (left)
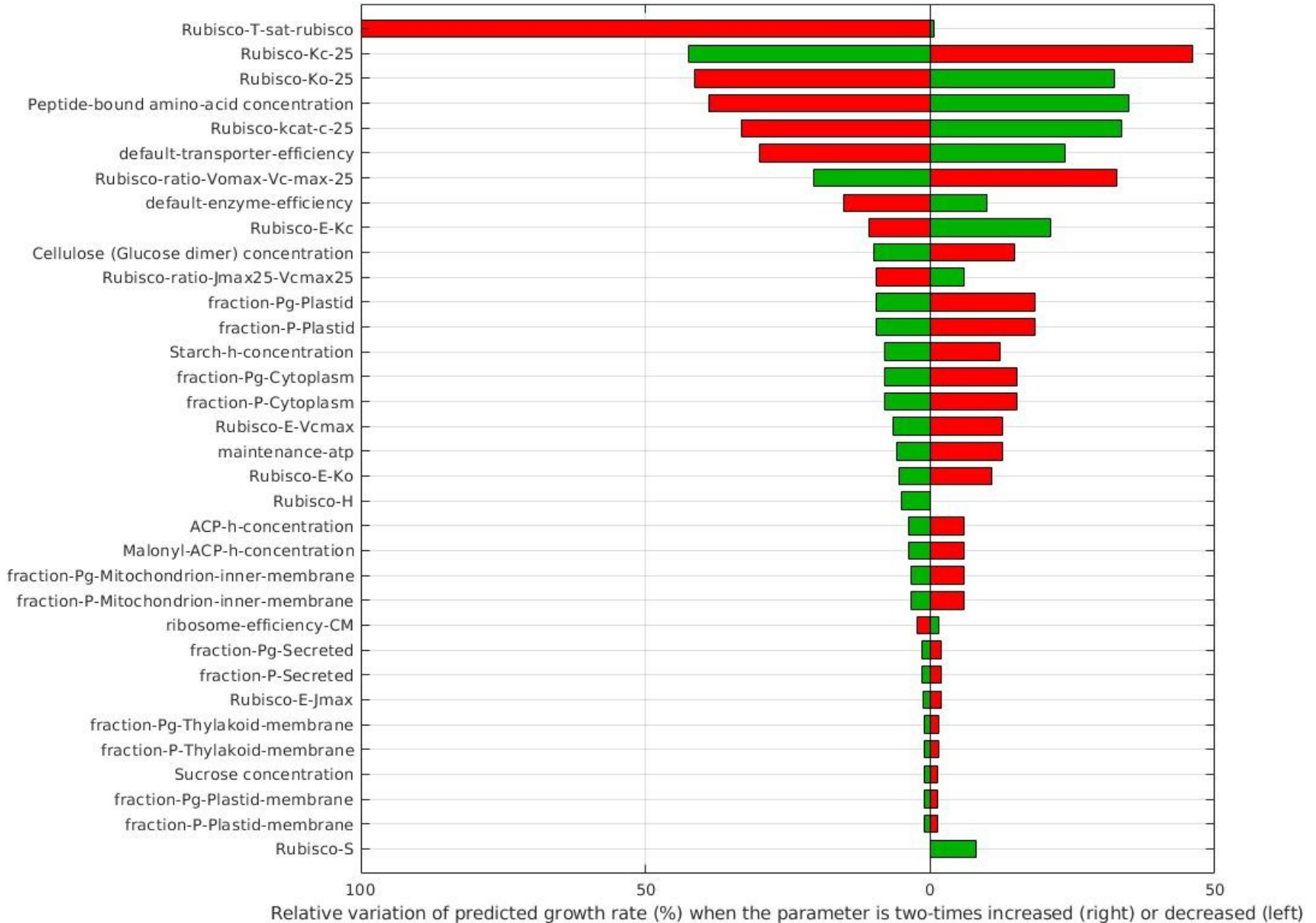
Comparison with the well-established Farquhar model of photosynthesis and carbon fixation

Simulations for different combinations of temperature et CO2

Variation of environmental conditions:
- Partial CO2 pressure
- Partial O2 pressure
- Irradiance
- Temperature

➡ Qualitatively correct

➡ Quantitative values are in the same range

1 point = 1 simulation

Farquhar et al. Planta 1980; 149, 78-90
Walker et al. Plant Cell Env 2013; 36.12:2108-2119

A. Goelzer et al. *In preparation*. 2022

23

Conclusion

A mature and versatile modeling framework

- Formulation into a convex optimization problem (LP problem) in steady-state regimen
- Experimentally validated under complex environmental conditions on bacteria
- Already use to model various bacteria (*B. subtilis, E. coli, C. necator*)
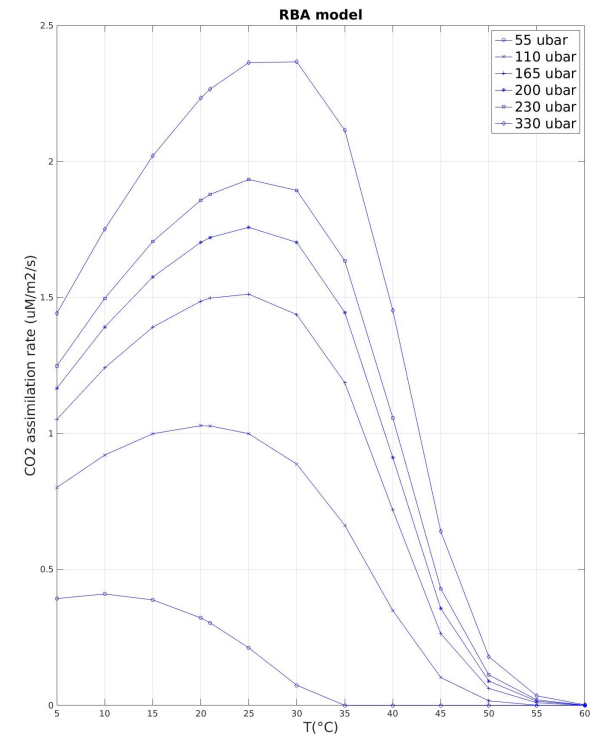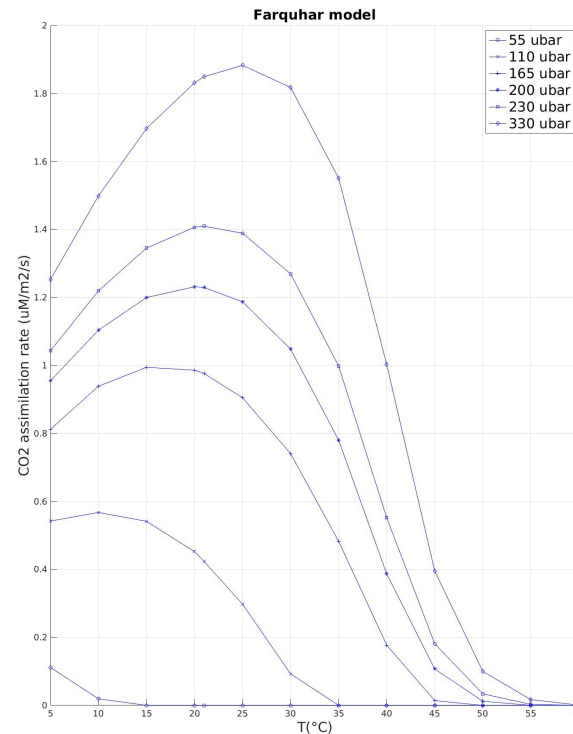- Extended in dynamical conditions for prokaryotes (dRBA)
- Allow for **autonomous** predictions of cellular phenotypes (in constrast to other constraint-based modeling methods)
- Intrinsically based on the <u>systemic decomposition of cells</u>

A set of open-source computational tools for the creation, analysis, calibration and simulation of RBA models available for the scientific community.
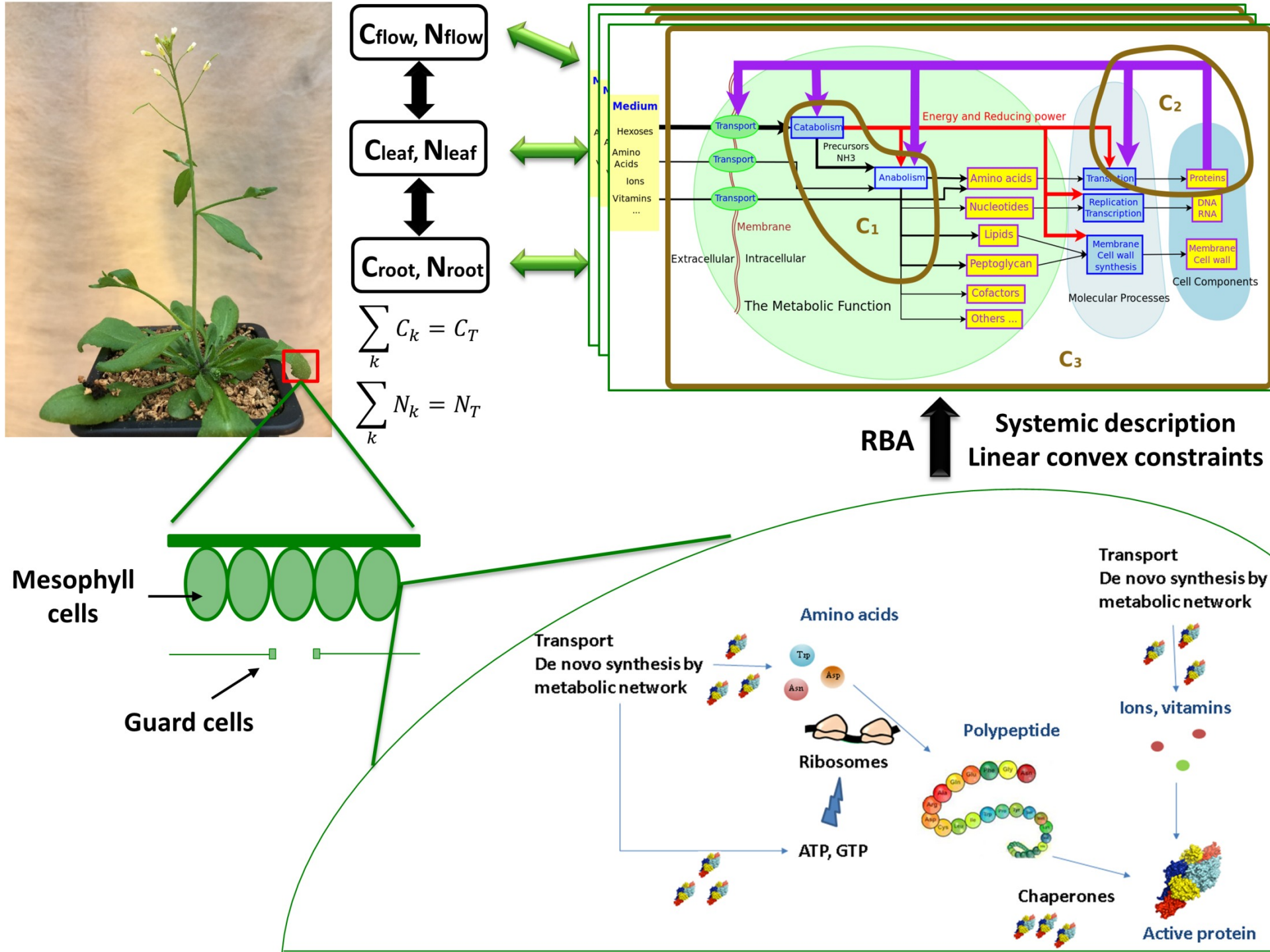
More information at
- https//rba.inrae.fr
- https://github.com/SysBioInra

Perspectives

Towards a multi-scale plant model of *A.thaliana*

$$\sum_k C_k = C_T$$

$$\sum_k N_k = N_T$$

Mesophyll cells

Guard cells

C_{flow}, N_{flow}

C_{leaf}, N_{leaf}

C_{root}, N_{root}

RBA

Systemic description
Linear convex constraints

Medium

Hexoses

Amino Acids

Ions

Vitamins

...

Transport

Catabolism

Energy and Reducing power

Precursors NH3

Anabolism

Membrane

Extracellular | Intracellular

C_1

C_2

C_3

Amino acids

Nucleotides

Lipids

Peptoglycan

Cofactors

Others ...

The Metabolic Function

Transcription

Replication Transcription

Membrane Cell wall synthesis

Molecular Processes

Proteins

DNA RNA

Membrane Cell wall

Cell Components

Transport
De novo synthesis by metabolic network

Amino acids

Trp

Asn

Asp

Ribosomes

ATP, GTP

Polypeptide

Transport
De novo synthesis by metabolic network

Ions, vitamins

Chaperones

Active protein

The next steps….
the plant model

- Organ-specialized RBA model (collab. IJPB, INRAE)
  - ✓ With biological expertise
  - ✓ Specialization of the molecular machines for organs using proteomics/transcriptomics
  - ✓ Developments of biocomputational pipelines for extracting information from public repositories
  - ✓ etc.
- Update RBApy for eukaryotic model creation and simulation

- Calibration and validation of the rosette model
  - ✓ In normal and combined stress conditions

  ➡ Work in progress.

- Use the model to revisit several biological questions on the mobilization of plant resources in different conditions
  - ✓ Cost of cell resource production
  - ✓ etc.

- Extend the model to other Arabidopsis accessions, other plants

Acknowledgments

V. Fromion

A. Bulovic, O. Bodeit, G. Jeanne

S. Fischer

L. Tournier, W. Liebermeister

E. Kuhn

O. Loudet   A. Hulot
L. Rajjou   D. Charif
F. Chardon   N. Bessoltane

V. Letort

All teams involved in