Approche hybride de modélisation explicable du métabolisme des écosystèmes microbiens

Hybrid approach for explainable metabolic modelling of microbial ecosystems'

PhD defense of Maxime LECOMTE

December 5, 2023

Membres du jury

Président: SIMON Laurent

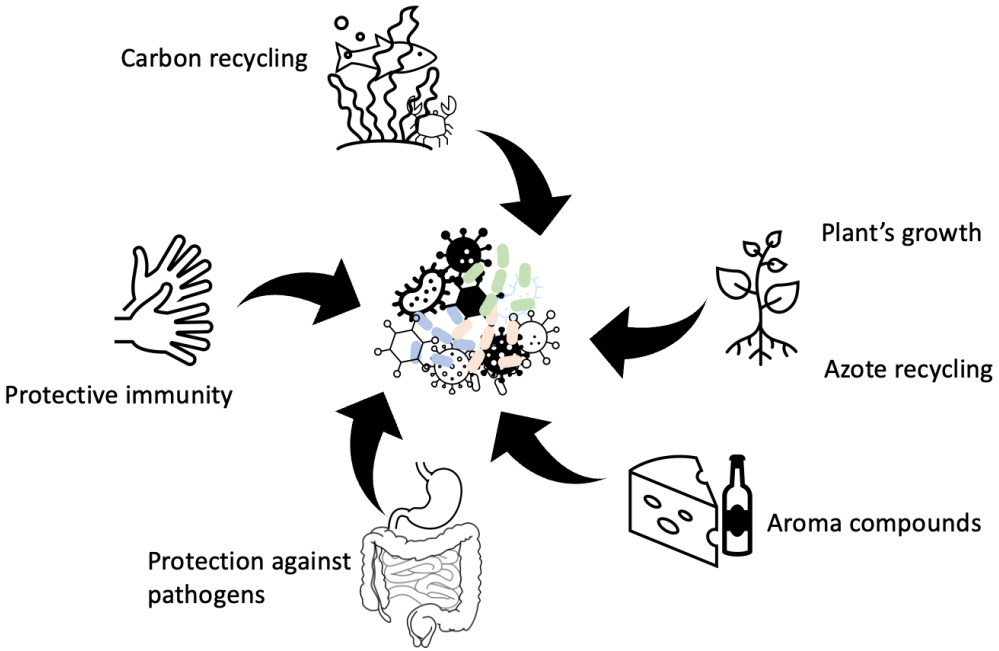
Rapportrices: BAROUKH Caroline
COCAIGN-BOUSQUET Muriel

Examinateurs: COTTRET Ludovic
LAROCHE Béatrice
MARKOV Gabriel

Co-direction: David SHERMAN et Hélène FALENTIN
Encadrement: Clémence FRIOUX

École doctorale Mathématiques et informatique — Université de Bordeaux — INRAE — Inria — LaBRI

Microbial ecosystems are complex



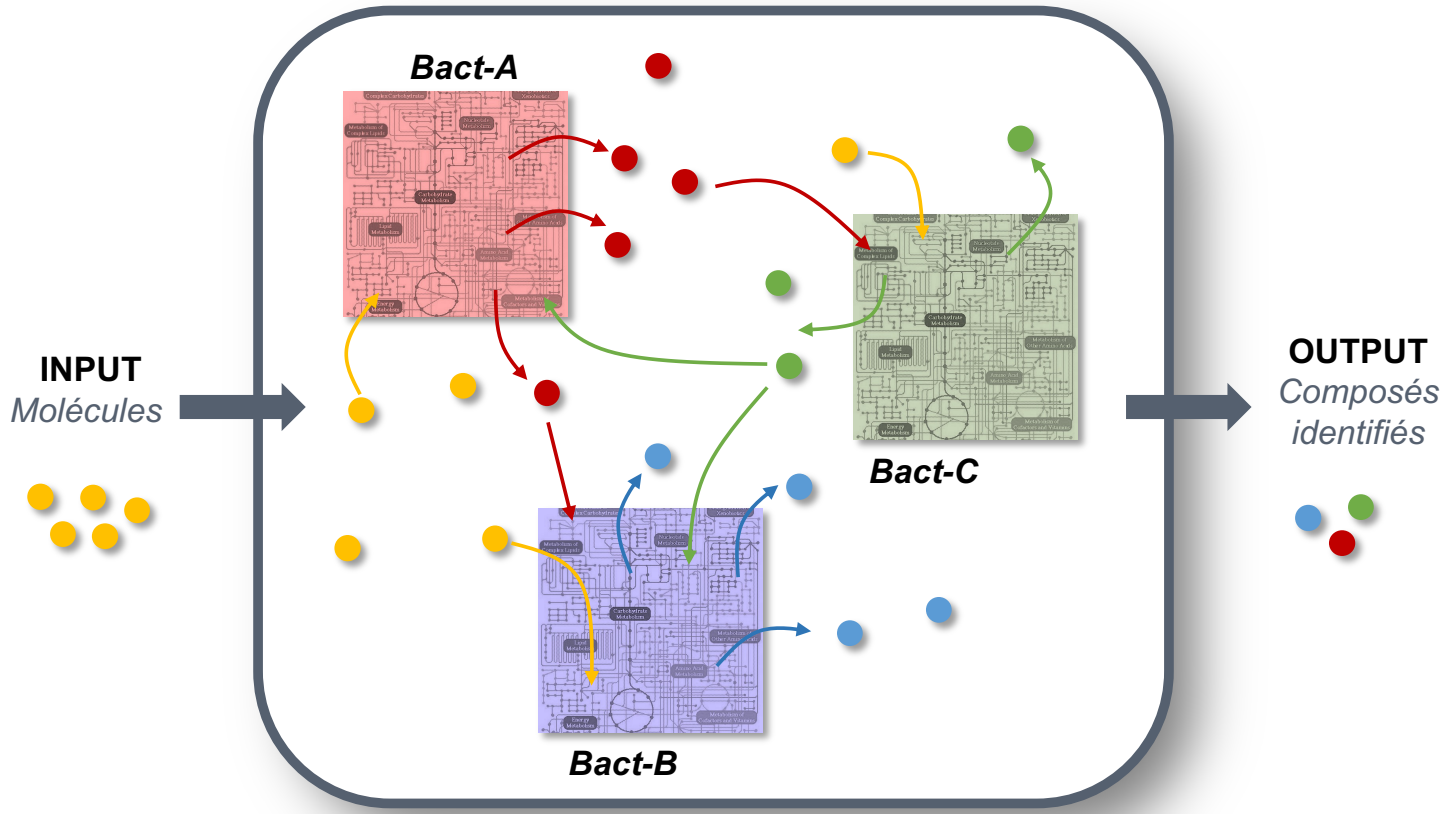
In the environment

- Taxonomic high diversity
- Functional high diversity
- Bacterial interactions are complex

Taxonomic diversity impact bacterial interactions

Royet and Plailly, 2004; Belkaid and Hand, 2014; Zhang et al., 2015; Hoorman, 2011; McSweeney and Sousa, 2000.

Bacterial interaction process



- The direction, which species , ecosystem consequence

Computational methods are required

Modeling as a computational method to study microbial communities

Modeling

Mathematical object used to simulate or predict

	Numerical	Discrete
Mathematical formalism	ODE or constraints	Knowledge base and logical rules
Solver	Numeric	Logic
Data	Multi-omics data (genomic, transcriptomics, metabolomics, growth measures…)	Genomes annotations
Scope	Controlled ecosystems	Screening bacterial communities
Mathematical object	Metabolic networks	Metabolic networks

Metabolism modeling for understanding complex ecosystems

Metabolism explains observable phenotype

Metabolism

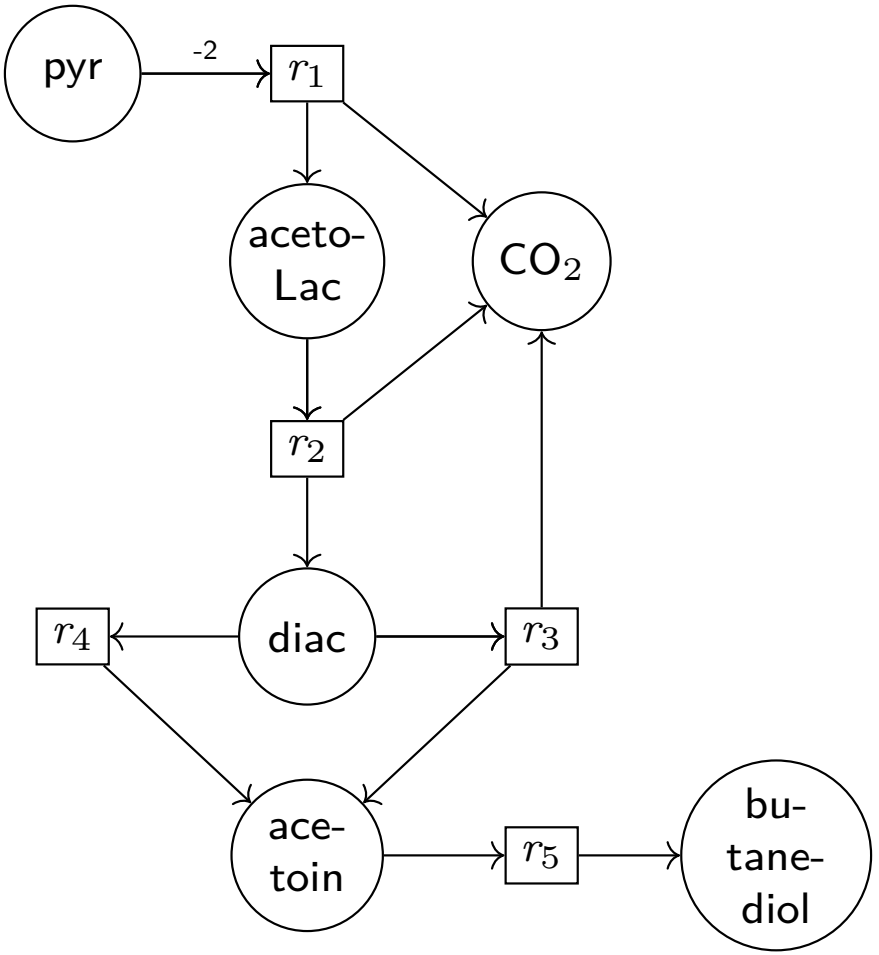
Set of all biochemical reactions occurring in the cell of an organism that permit the production of energy and metabolic goods.

r_1 :2 pyr $\rightarrow$ 1 aceto-Lac + 1 CO_2
r_2 :1 aceto-Lac $\rightarrow$ 1 diac + 1 CO_2
r_3 :1 aceto-Lac $\rightarrow$ 1 acetoin + 1 CO_2
r_4 :1 diac $\rightarrow$ 1 acetoin
r_5 :1 acetoin $\rightarrow$1 butanediol

Stoichiometry matrix

	r_1	r_2	r_3	r_4	r_5
pyr	-2	0	0	0	0
aceto-Lac	1	-1	-1	0	0
diac	0	1	0	-1	0
CO_2	1	1	1	0	0
acetoin	0	0	1	1	-1
butanediol	0	0	0	0	1

Bipartite graph relationship

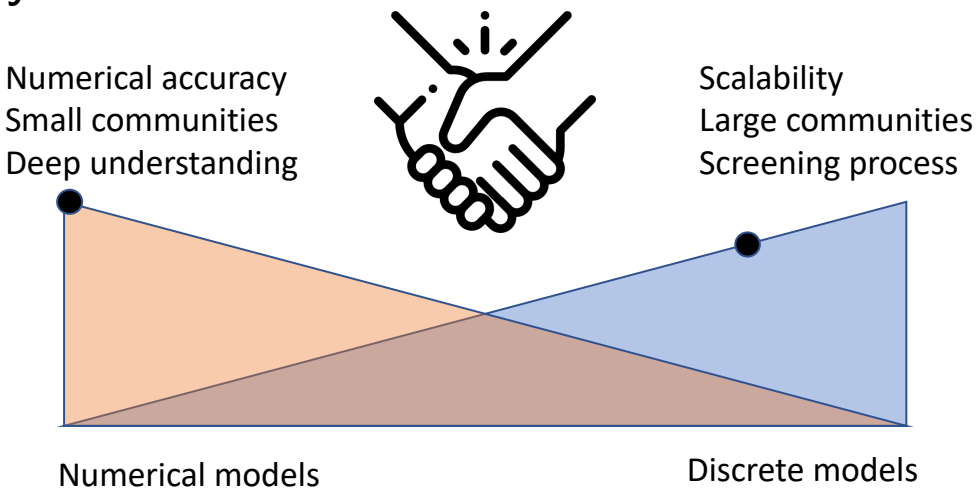


Sánchez López de Nava A, 2023.

Hybrid approach of the metabolism

The approach is hybrid

- Temporal aspect
- Quantitative information



- Snapshot
- Qualitative information

I am hybrid

→ INRAE-Inria PhD student

→ Biology and computation background

Hybrid approach

Mathematical formalism choice is determined from the biological issue

Objective and contributions

Objective

Contribute to analyzing metabolic interactions of bacterial communities

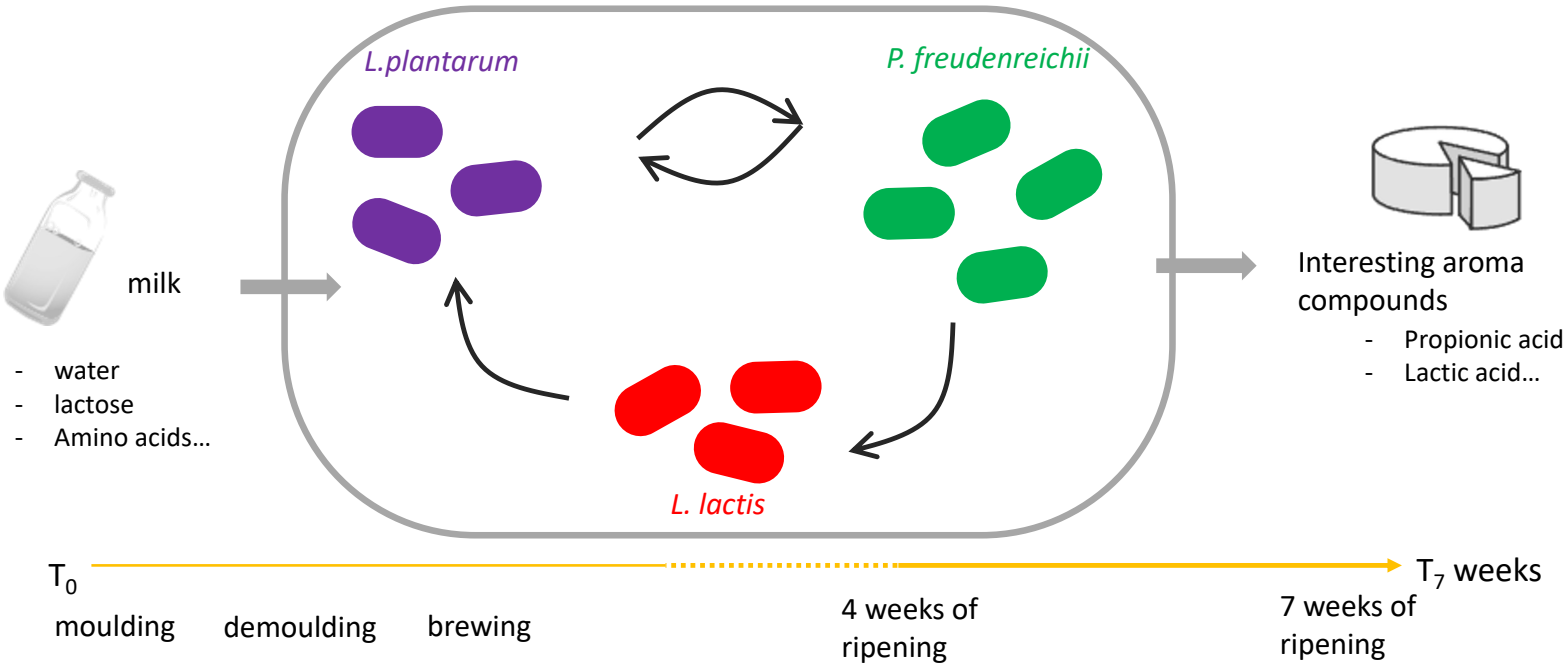
1. Numerical method applied to controlled community involved in cheese production → TANGO implementation (*in revision in Metabolic Engineering*)
   - Multi-omics data integration
   - Dynamic model → **iterative process**
2. Discrete model for screening metabolic bacterial potential in natural ecosystems → CoCoMiCo software (*in prep.*)
   - Inference of **interaction properties** from KB
3. Draft to improve discrete models
   - Inference of **temporal information** from KB
   - Selection of the best community based on **biological constraints**

---

https://forgemia.inra.fr/tango/tango_models

https://gitlab.inria.fr/ccmc/CoCoMiCo

Bacterial fermentation: a complex dynamic process



Challenge

- Build coherent metabolic networks (iterative refinement)
- Nutrient concentration over time
- The dynamic of bacterial density
- Resource sharing

Solution

- Create a dynamic and a numeric model of the metabolism (FBA, dFBA)
- TANGO implementation for characterizing bacterial interaction

Orth, Thiele, and Palsson, 2010; Mahadevan, Edwards, and Doyle, 2002.

Numerical model of the metabolism – FBA

definition

Metabolic model

From a GEM, a model metabolic has the capacity to simulate and to predict on the metabolic activity

Numerical model of the metabolism – FBA

Flux Balance Analysis (Orth, Thiele, and Palsson, 2010)

Metabolic model

From a GEM, a model metabolic has the capacity to simulate and to predict on the metabolic activity

Constraint-based approaches

$$\text{maximiser/minimiser } f_{obj}$$
$$\text{tel que } (S.v)_{int} = 0$$
$$\text{et } v_{i_{min}} \leq v_i \leq v_{i_{max}}$$

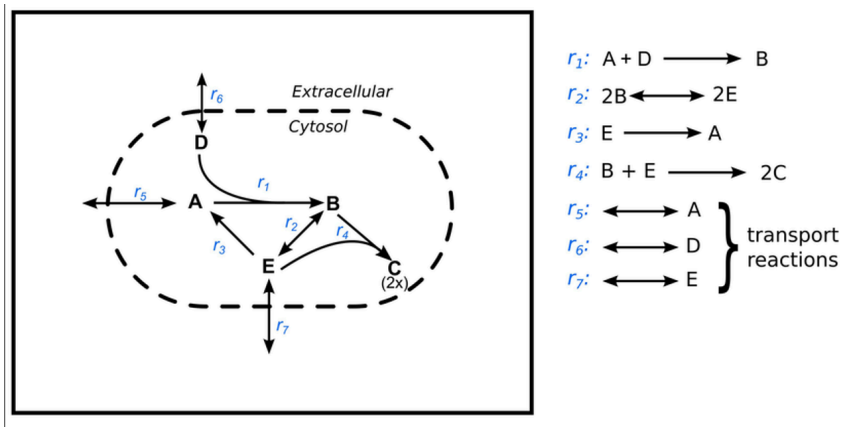


Figure 1: Example of metabolic network

$$S = \begin{array}{c|ccccccc} & r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 \\ A & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ B & 1 & -2 & 0 & -1 & 0 & 0 & 0 \\ C & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ D & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ E & 0 & 2 & -1 & -1 & 0 & 0 & 1 \end{array}$$

(Stoichiometric values)

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix}$$

(Metabolic flux values)

Figure 2: A. Stoichiometry matrix representation and the flux vector v

Numerical model of the metabolism – FBA

Flux Balance Analysis (Orth, Thiele, and Palsson, 2010)

Metabolic model

From a GEM, a model metabolic has the capacity to simulate and to predict on the metabolic content

Constraint-based approaches

maximiser/minimiser f_{obj}

tel que $(S.v)_{int} = 0$

et $v_{i_{min}} \leq v_i \leq v_{i_{max}}$

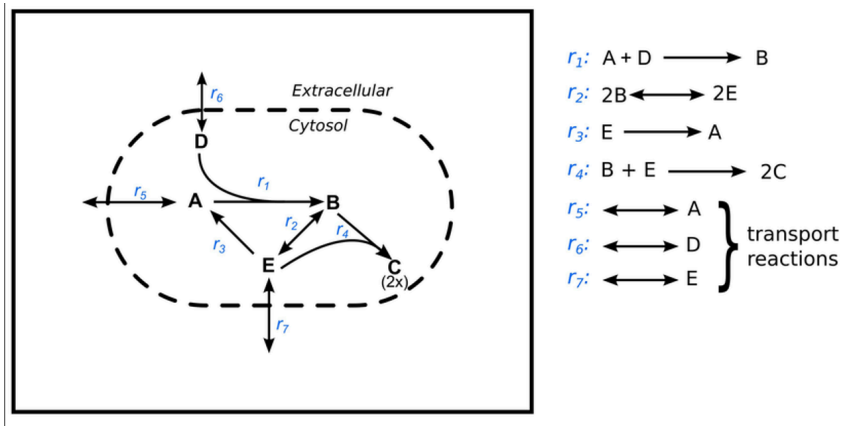


Figure 1: Example of metabolic network

Subject to

$$S\vec{v} = \vec{0} = \begin{cases} \frac{dA}{dt} = -v_1 + v_3 + v_5 \\ \frac{dB}{dt} = v_1 - 2v_2 - v_4 \\ \frac{dC}{dt} = 2v_4 \\ \frac{dD}{dt} = -v_1 + v_6 \\ \frac{dE}{dt} = 2v_2 - v_3 - v_4 + v_7 \end{cases}$$

(Steady state system)

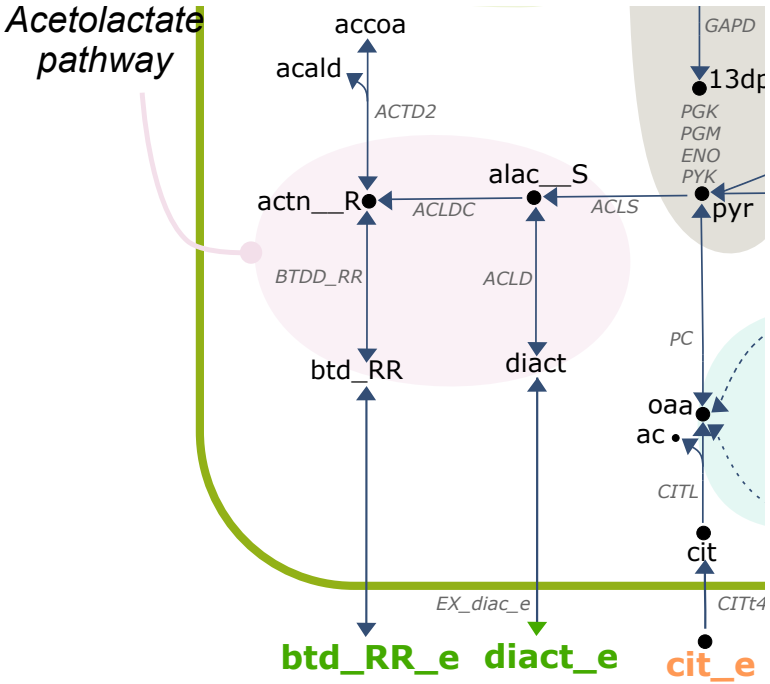
$$\begin{array}{l} 0 \leq v_1 < \infty \\ -\infty < v_2 < \infty \\ 0 \leq v_3 < \infty \\ 0 \leq v_4 < \infty \\ 0 \leq v_5 \leq \infty \\ -\infty < v_6 < \infty \\ 0 \leq v_7 \leq \infty \end{array}$$

(Reaction bounds)

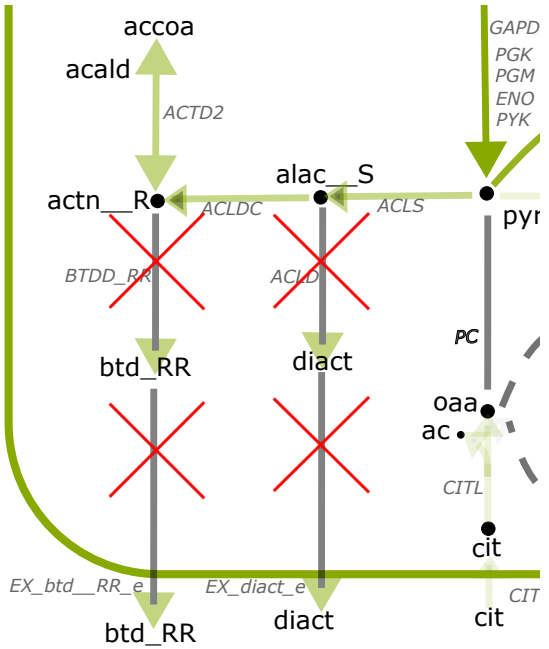
Figure 2: B. Linear programming problem.

FBA utilization to refine genome-scale metabolic model

- Reactions modifications (flux, add, remove)
- Qualitative check of existing pathway and metabolic goods



Initial metabolic map

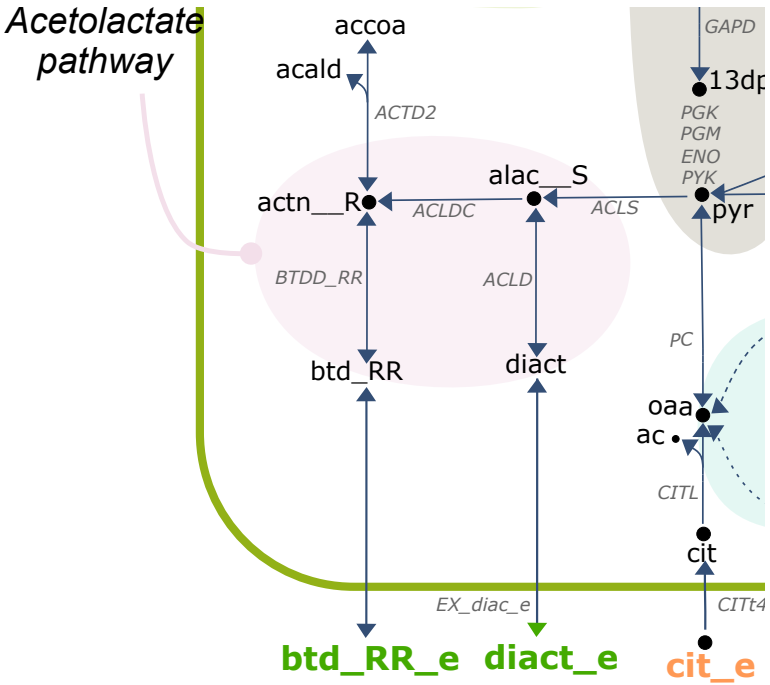


Flux representation **before** modifications

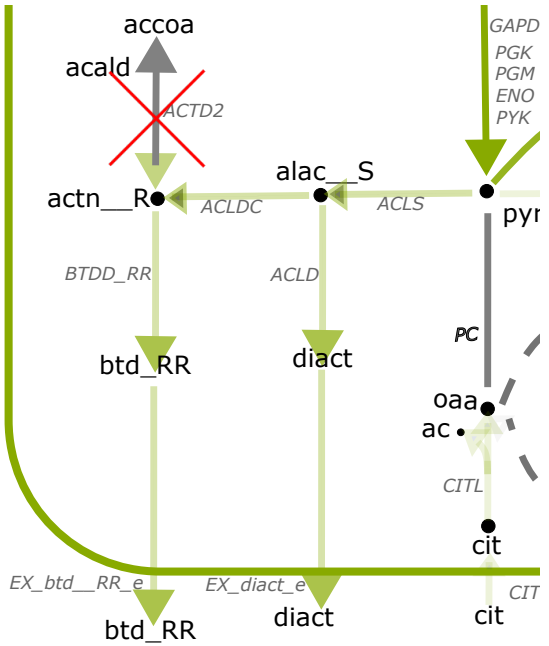
Carroll et al., 1999; Swindell et al., 1996; Makhlouf, 2006.

FBA utilization to refine genome-scale metabolic model

- Reactions modifications (flux, add, remove)
- Qualitative check of existing pathway and metabolic goods



Initial metabolic map



Flux representation **after** modifications

Improvement

Modification reaction bounds in the model and in the software

Carroll et al., 1999; Swindell et al., 1996; Makhlouf, 2006.

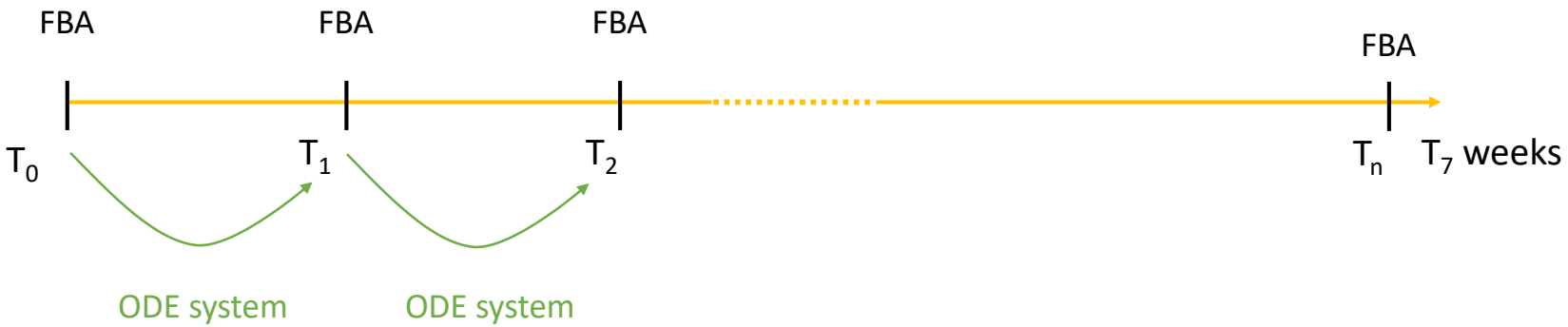
Dynamic model of the metabolism

dFBA

- FBA at each time point
- Previous time point output used as input for the next time point

FBA

$$F_j = \sum_{i \in \mathcal{B}} \mu_{i,j} \left((c^{ex}_{min,i}, c^{ex}_{max,i})(b^n, m^n) \right) b_i$$



ODE System

$$b_i^{n+1} = b_i^n + \Delta t * F_{b_i}$$

$$m_j^{n+1} = \begin{cases} m_i^n + \Delta t * F_j & \text{si } F_j > 0 \\ m_j^n / (1 - \Delta t * F_j / m_j^n) & \text{sinon} \end{cases}$$

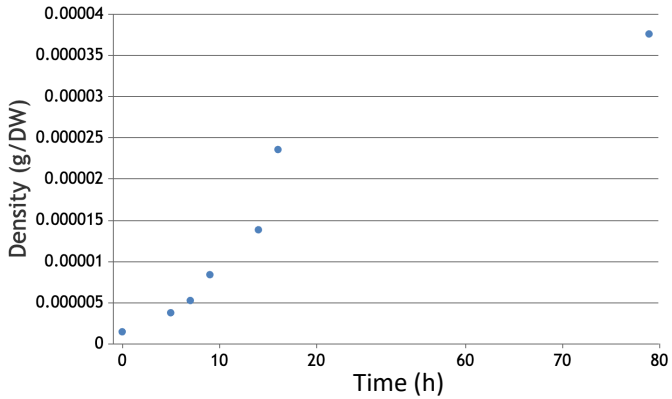
Mahadevan, Edwards, and Doyle, 2002.

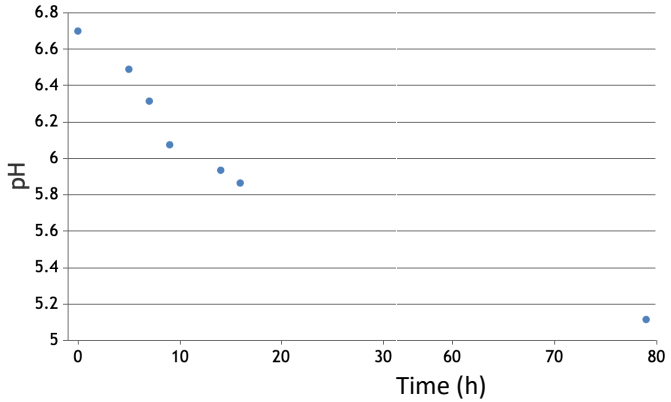
Dynamic individual calibration

- Make individual GSMN accurate for inferring mechanistic bacterial behavior
- Finding optimal parameters for explaining **individual** biological observations
- Quantitative check of metabolic goods and biomass density

LAB

$$J(b_i, pH|\theta_i, b_{i,exp}, pH_{exp}) = \left\| \frac{\log_{10}(b_i) - \log_{10}(b_{i,exp})}{\sigma_{log,i,exp}} \right\|^2 + \alpha \left\| \frac{pH - pH_{exp}}{\sigma_{pH,exp}} \right\|^2 \quad (1)$$





Milk condition

- Without adding extra nutrients

Dynamic individual calibration

- Make individual GSMN accurate for inferring mechanistic bacterial behavior
- Finding optimal parameters for explaining **individual** biological observations
- Quantitative check of metabolic goods and biomass density

P. freudenreichii

$$J(b, m|\theta_i, b_{exp}, m_{exp}) = \left\| \frac{\log_{10}(b) - \log_{10}(b_{exp})}{\sigma_{log,b,exp}} \right\|^2 + \alpha \left\| \frac{m - m_{exp}}{\sigma_{m,exp}} \right\|^2 \quad (1)$$

Acides	Concentration T_0	Concentration $T_f = 89$ h	Erreur standard
Lactate	16.5	7.88	0.08
Acetate	0	3.07	0.02
Succinate	0	0.371	0.063
Propionate	0	8.31	0.02

Table 1: Acids concentrations data in g/L

Milk condition

- Adding lactate and peptone in the nutritional environment

Dynamic modulations

Defining the usual consumption limitation in the model

$$c^{ex}_{min,i,j} = max(-\frac{m_j}{\Delta t * \sum_{i \in \mathcal{M}_|} b_i}, v^{int}_{i,j}) \quad (2)$$

- $\mathcal{M}_|$ Bacteria subset can metabolize $_j$
- $v^{int}_{i,j}$ close to the FBA value
- Balanced resource sharing

Lactose consumption case for LAB

1- Calcul of pH from lactate concentration

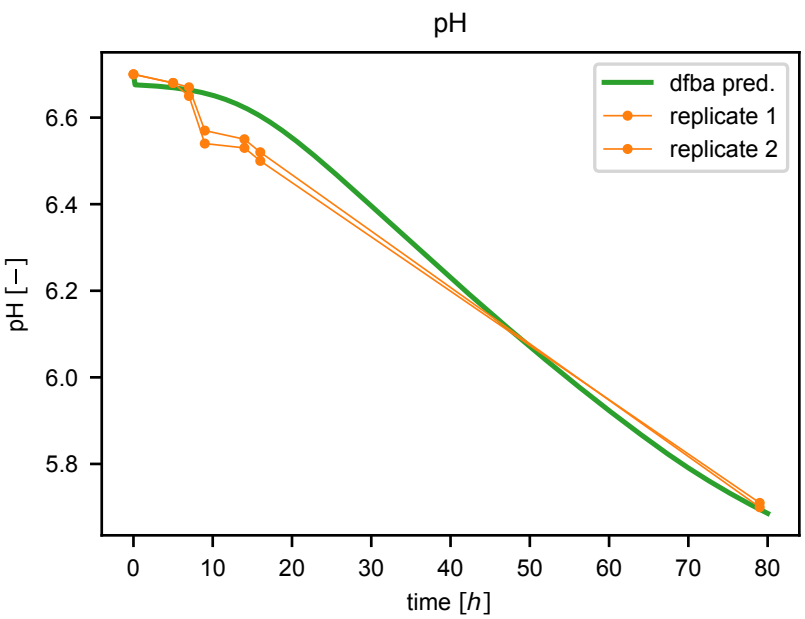
$$pH = pKa_{lac} c_1 * (m_{lac\_\_L\_e} + m_{lac\_\_D\_e}) + c_2$$

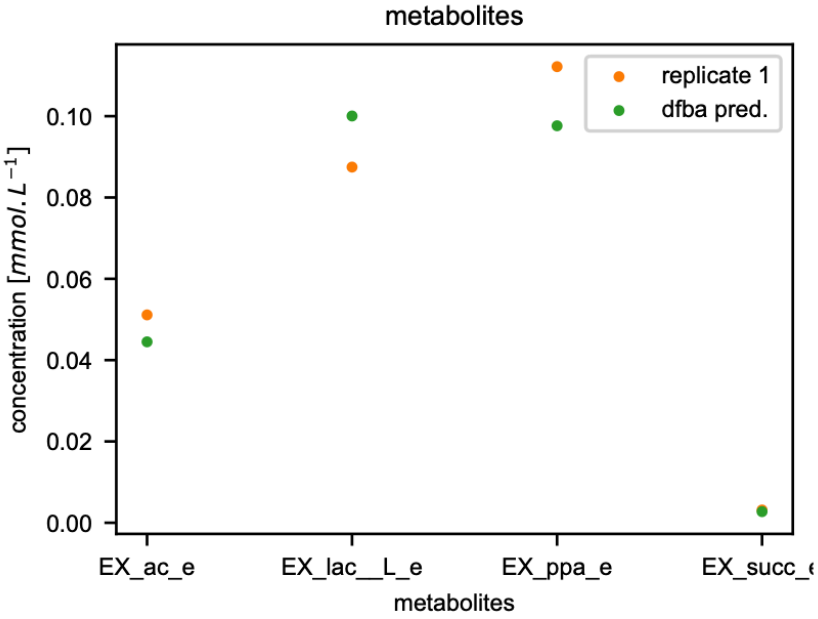
2- Regulate lactose consumption

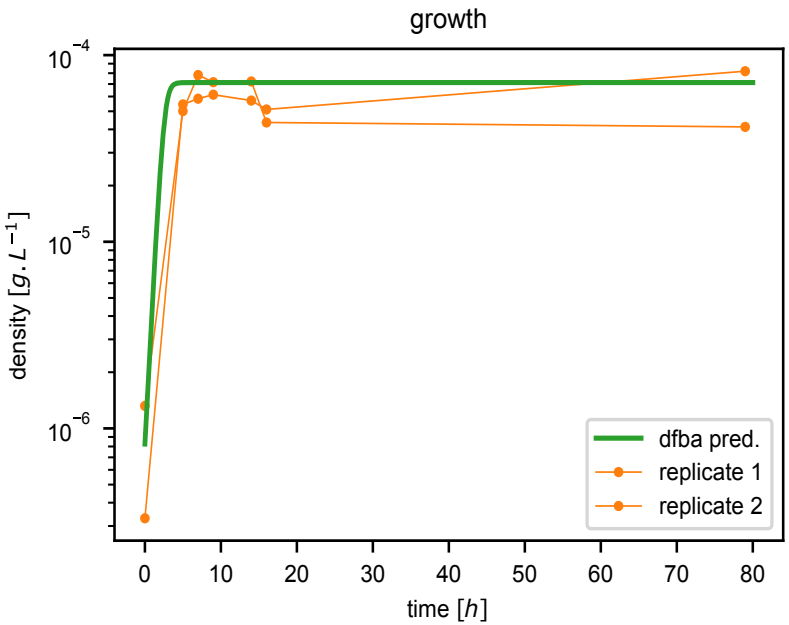
$$c^{ex}_{min,i,j} = max(-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, -\mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} - \mu_{min,lcts}) \quad (3)$$

Ozcan et al., 2020.

Model validation



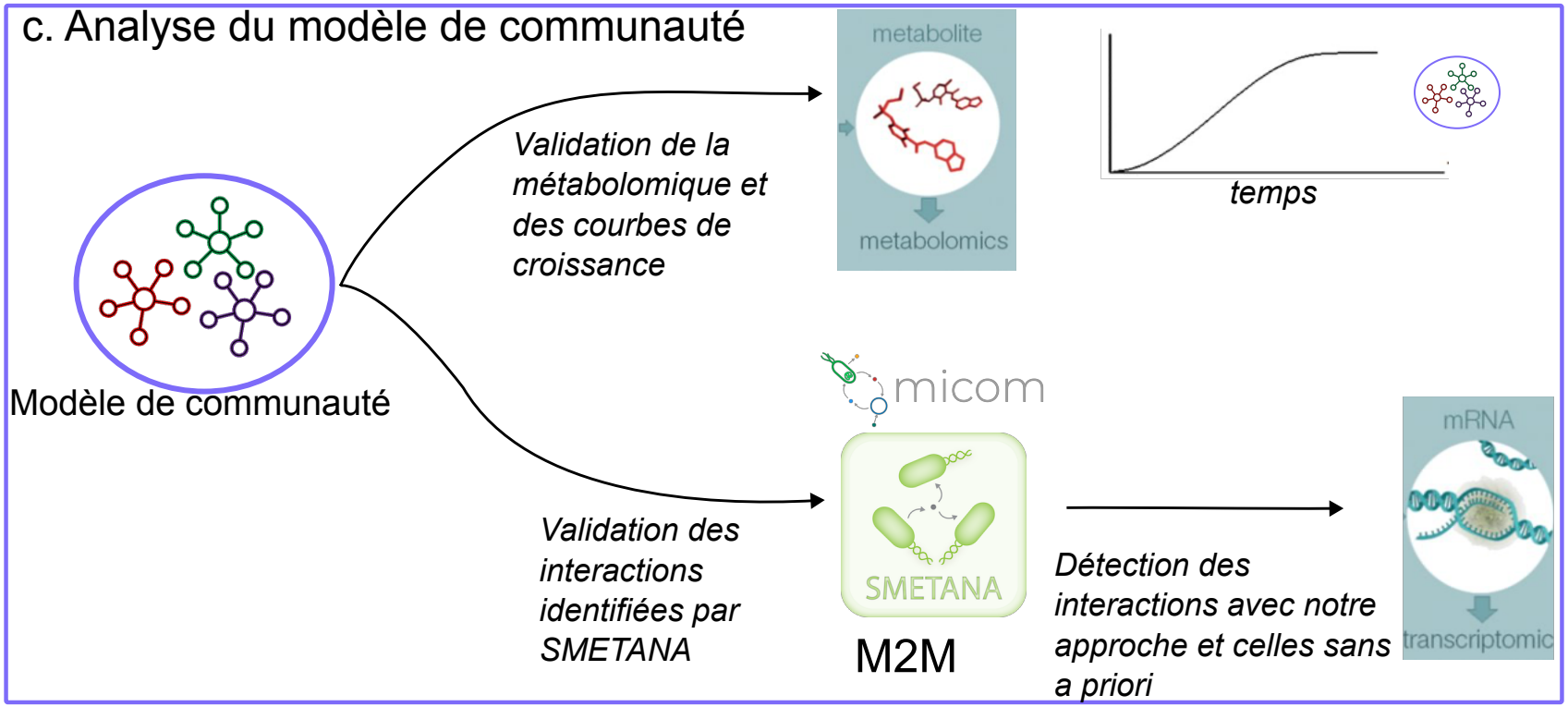




Take home message

Individual metabolic model well calibrated on know behavior → reproduce experimental data

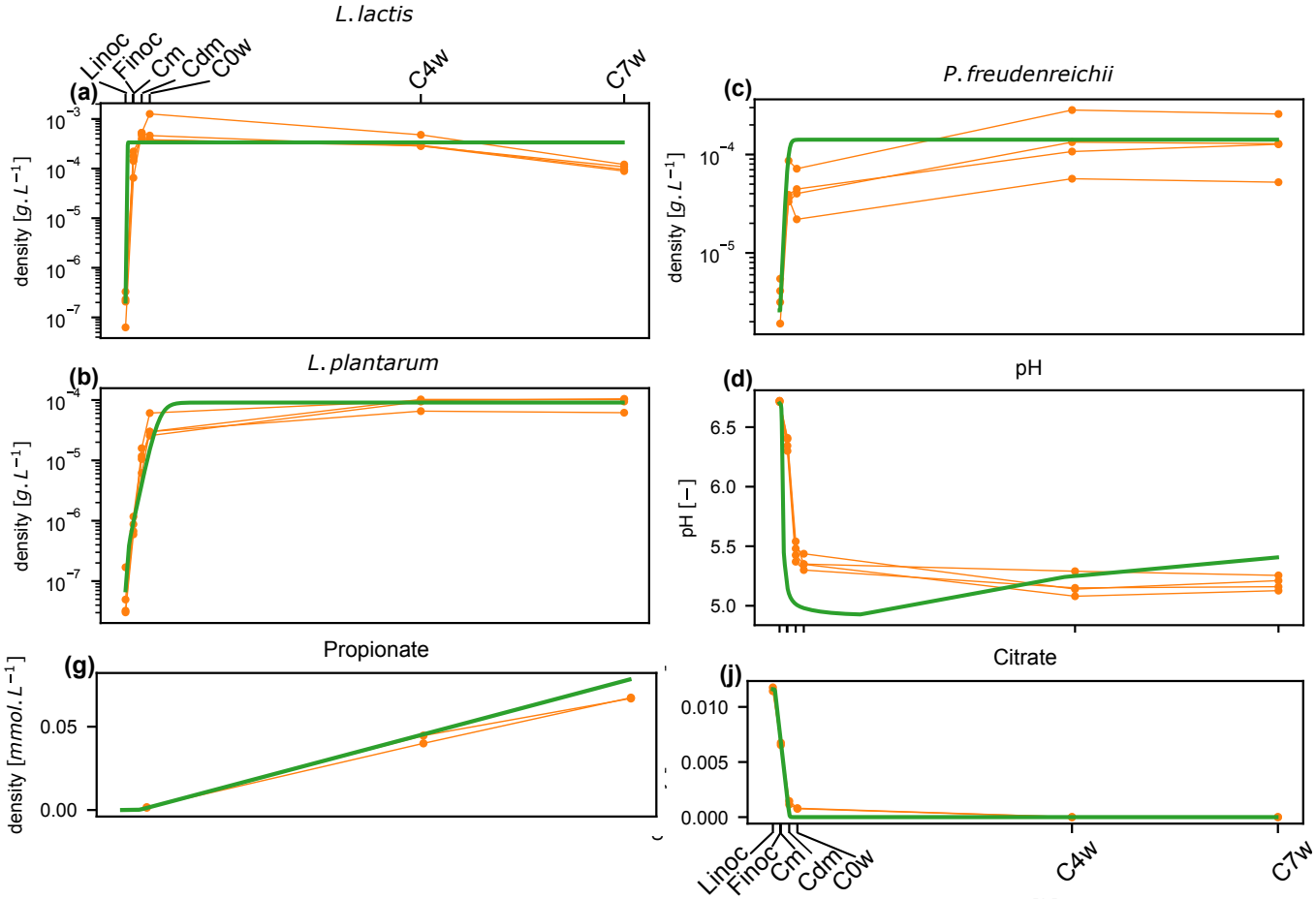
Community validation



- Bacterial interaction **prediction**
- Metabolic explanation of biological observations
- **No community calibration**

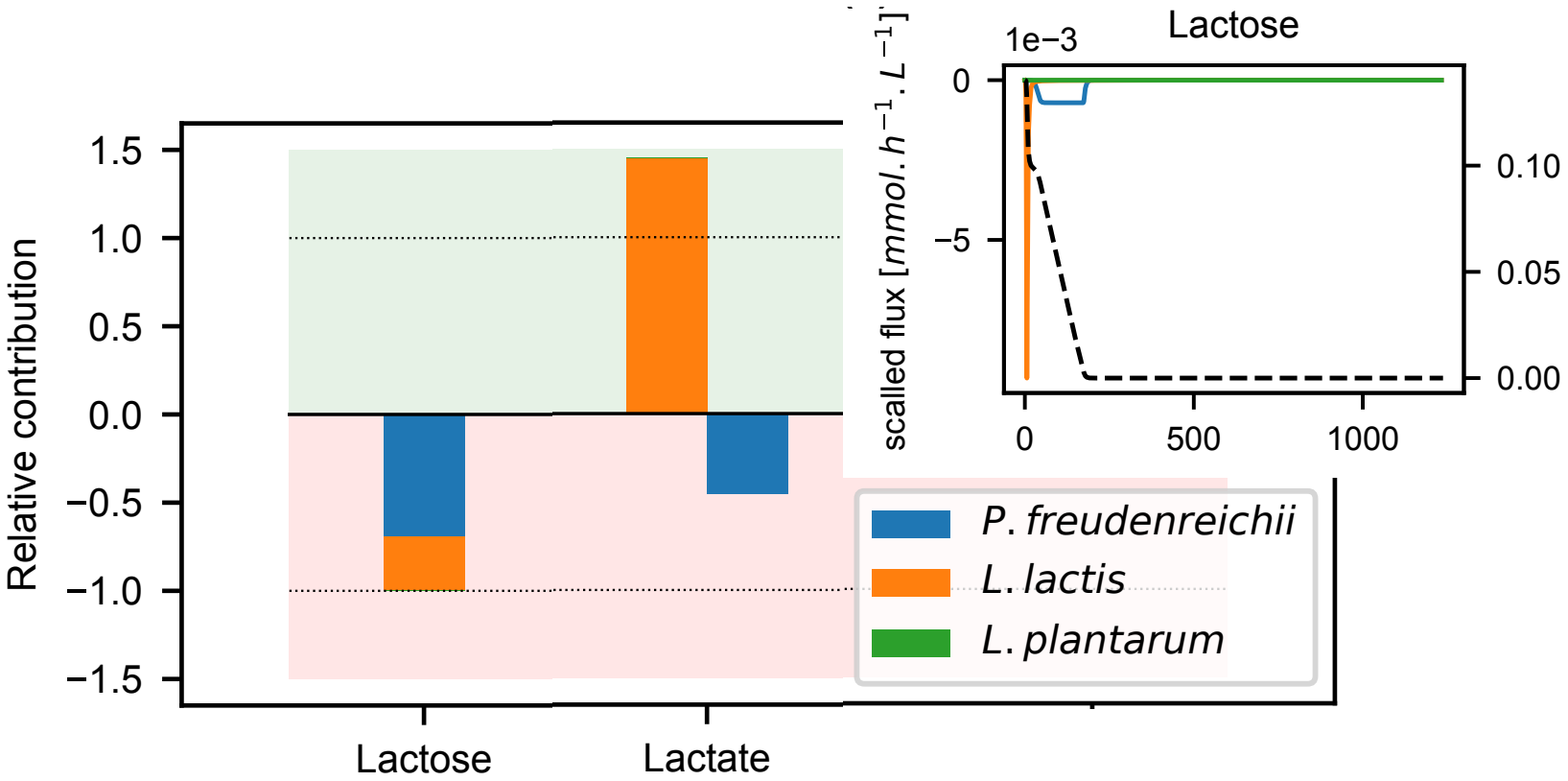
Community prediction

Growth and pH



- Growth well predicted for all bacteria
- Lactate proxy production can explain the observed pH
- Metabolomic is well predicted

Highlighted bacterial community behavior



- No competition for lactose
- *L. lactis* main lactate producer
- Cooperation between *L. lactis* and *P. freudenreichii*

- 11 shared metabolites predicted with SMETANA, MiCOM (phenylalanine, succinate, xanthine..)

Added-value and limitations

Take home message

Originality

- High quality of refinement and well calibrated individual GSMN from cheese
- Individual calibrations based on community behavior permit the prediction metabolites concentrations

Scalability issue

- Refinement process is time consuming
- The iterative methodology assume well documented GSMNs in literature
- Based on *a priori* knowledge for screening compounds

Solution

For screening large community, use of different formalism is required