Genome-wide prediction of resource allocation in bacteria

(Resource Balance Analysis)

Anne Goelzer

anne.goelzer@jouy.inra.fr

Mathématiques et Informatique Appliquées

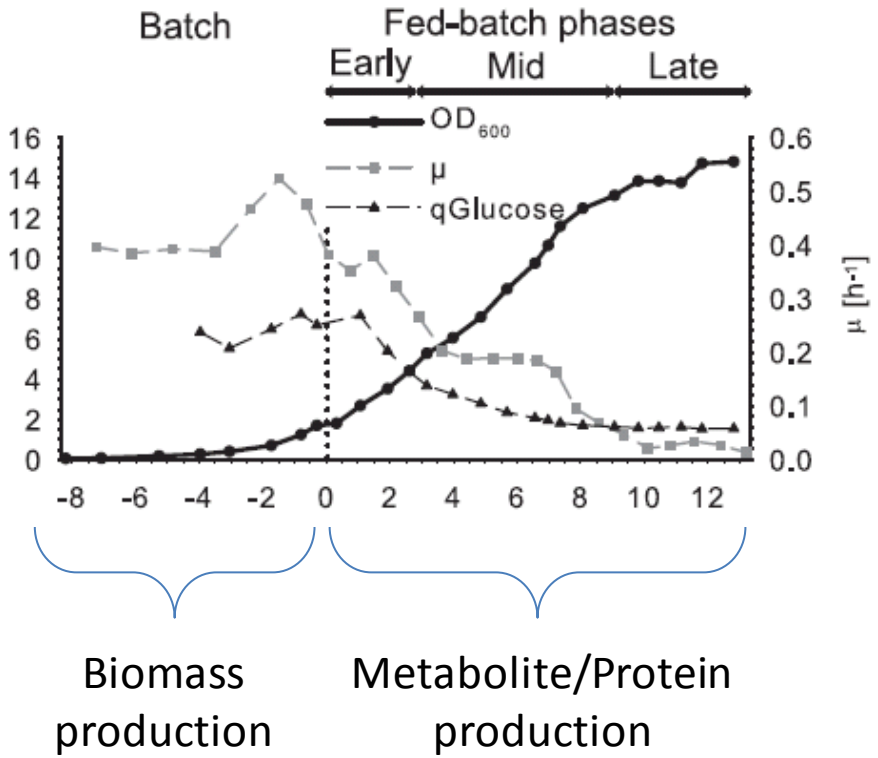
du Génome à l'Environnement

INRA Jouy-en-Josas, FRANCE

INRA SCIENCE & IMPACT

MOABI, 27 November 2015

How a rational design of industrial strains can be achieved ?

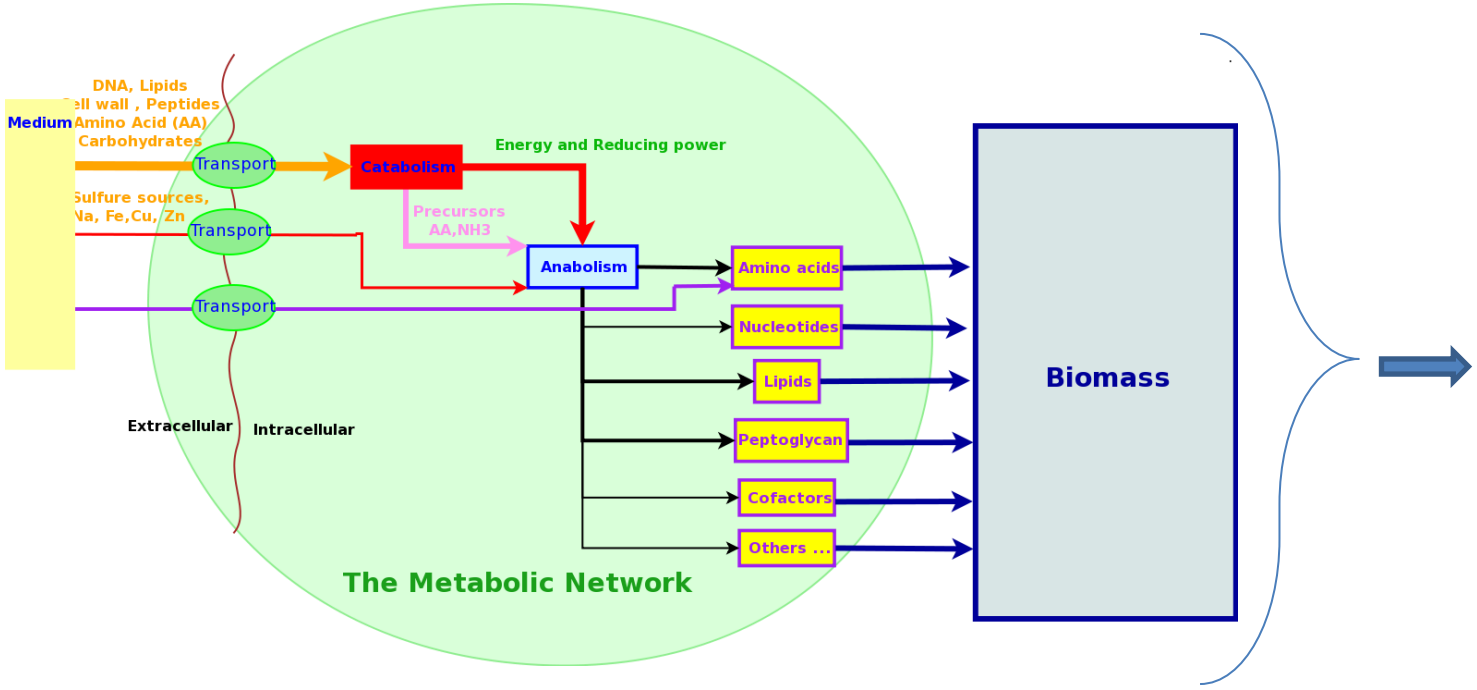


- What we would like: obtain the maximal capability of production in each phase (metabolic engineering)
- Problem: impact of genetic modification are difficult to anticipate/predict
- Can we understand/predict at the cell scale the impact of genetic modification ?

What kind of computational methods could be suitable to drive the strain design ?

The methods of type« Constraint-based modeling » have been widely used in Metabolic Engineering

For instance, the method Flux Balance Analysis

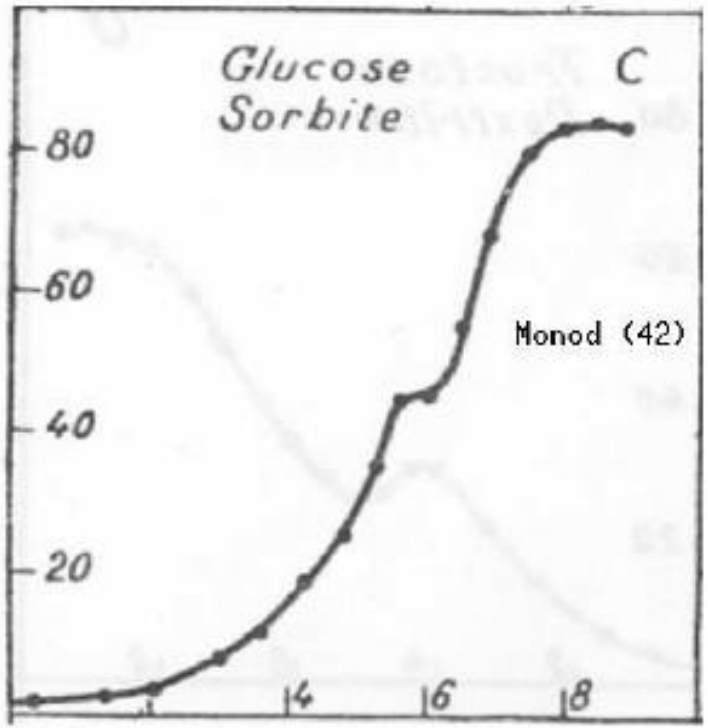


A LP optimization problem

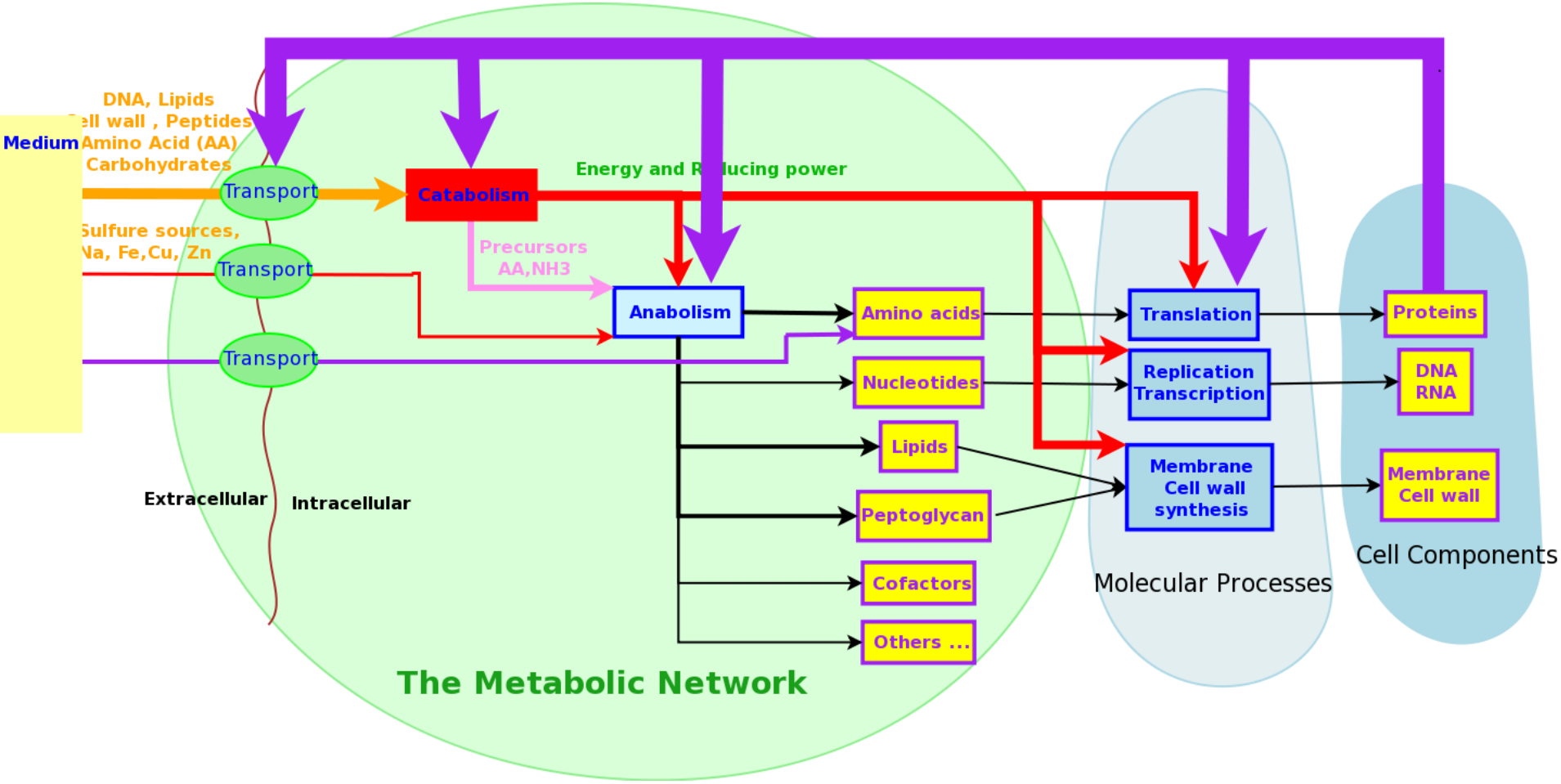
$$\begin{array}{ll} maximize & c^T v \\ v \in \mathbb{R}^n & \\ subject\ to & \\ & Sv = 0 \\ & \alpha_i \leq v_i \leq \beta_i \end{array}$$

Biomass production
Metabolic Fluxes at genome-scale

A major discrepancy: the catabolic repression

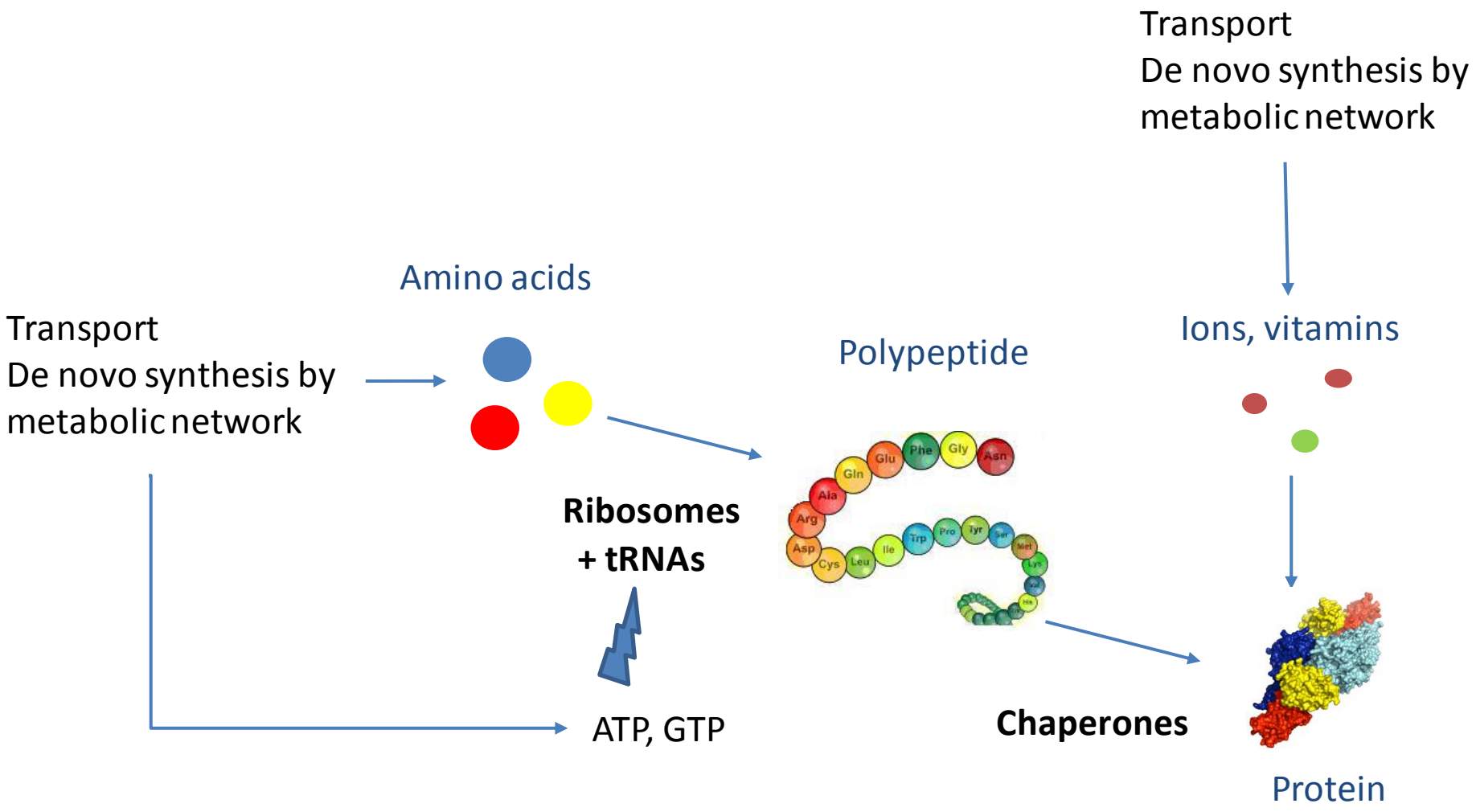


The glucose is favored over sorbitol in *Bacillus subtilis*
FBA predicts that both glucose and sorbitol are used simultaneously



Resources (especially proteins) have to be shared by all biological processes (implicit feedback).

Detailed integration of production costs for protein synthesis



Formalization into an optimization problem

Resource Balance Analysis (RBA)

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

Find $R \geq 0, C \geq 0, \nu^x \in \mathcal{R}^m$,

subject to

(C_{1a}) For all $i \in I_p$,
$$-\sum_{j=1}^{m} S_{p_{ij}} \nu_j^x + \mu \left(\sum_{j=1}^{m} C_{M_{ij}}^{M_p} |\nu_j^x| + C_{R_i}^{M_p} R + C_{C_i}^{M_p} C + C_{G_i}^{M_p} P_G^{x,T} \right) - \nu_Y = 0$$

(C_{1b}) For all $i \in I_c$,
$$-\sum_{j=1}^{m} S_{c_{ij}} \nu_j^x + \mu \bar{X}_{c_i} = 0$$

(C_{1c}) For all $i \in I_r$,
$$\sum_{j=1}^{m} S_{r_{ij}} \nu_j^x + \mu \left(\sum_{j=1}^{m} C_{M_{ij}}^{M_r} |\nu_j^x| + C_{R_i}^{M_r} R + C_{C_i}^{M_r} C + C_{G_i}^{M_r} P_G^{x,T} \right) = 0$$

(C_{1d}) For all $i \in I_i$,
$$\sum_{j=1}^{m} S_{I_{ij}} \nu_j^x = 0$$

(C_{2a}) $$\mu \left(\sum_{j=1}^{m} C_{M_j}^{R} |\nu_j^x| + C_R^R R + C_C^R C + C_G^R P_G^{x,T} \right) - k_T R = 0$$

(C_{2b}) $$\alpha_c \mu \left(\sum_{j=1}^{m} C_{M_j}^{R} |\nu_j^x| + C_R^R R + C_C^R C + C_G^R P_G^{x,T} \right) - k_C C = 0$$

(C_{3a}) $$\sum_{j=1}^{m} C_{M_j}^{D} |\nu_j^c| + C_R^D R + C_C^D C + C_G^D P_G^{c,T} - \bar{D}_c \leq 0$$

(C_{3b}) $$\sum_{j=1}^{m} C_{M_j}^{S} |\nu_j^s| + C_G^S P_G^{s,T} - \bar{D}_s \leq 0$$

Energy & precursors production

Protein production

Protein folding

Cytosol occupancy

Membrane occupancy

A. Goelzer , V. Fromion and G. Scorletti *Cell design in bacteria as a convex optimization problem*. 48th IEEE Conference on Decision and Control, China, 4517 -22. **2009**.
A. Goelzer , V. Fromion and G. Scorletti *Cell design in bacteria as a convex optimization problem*. Automatica,47(6):1210-1218. **2011**.

The RBA framework

- The feasibility problem is convex
- Equivalence with a Linear Programming (LP) optimization problem

  ⇨ same complexity as FBA, efficient resolution at genome scale!

Equivalence with a Linear Programming problem

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

find $\quad R \geq 0, C \geq 0, \nu^x \in \mathcal{R}^m, E^x \in \mathcal{R}^m_+$

subject to

(C^{lp}_{1a}) for all $i \in I_p$,

$$-\sum_{j=1}^{N_M} S_{p_{ij}} \nu^x_j + \mu \left(\sum_{j=1}^{N_M} C^{M_p}_{M_{ij}} E^x_j + C^{M_p}_{R_i} R + C^{M_p}_{C_i} C + C^{M_p}_{G_i} P^{x,T}_G \right) - \nu_Y = 0$$

(C^{lp}_{1b}) for all $i \in I_c$,

$$-\sum_{j=1}^{N_m} S_{c_{ij}} \nu^x_j + \mu \bar{X}_{c_i} = 0$$

(C^{lp}_{1c}) for all $i \in I_r$,

$$\sum_{j=1}^{N_M} S_{r_{ij}} \nu^x_j + \mu \left(\sum_{j=1}^{N_M} C^{M_r}_{M_{ij}} E^x_j + C^{M_r}_{R_i} R + C^{M_r}_{C_i} C + C^{M_r}_{G_i} P^{x,T}_G \right) = 0$$

(C^{lp}_{1d}) for all $i \in I_i$,

$$\sum_{j=1}^{m} S_{I_{ij}} \nu^x_j = 0$$

(C^{lp}_{2a})

$$\mu \left(\sum_{j=1}^{N_M} C^R_{M_j} E^x_j + C^R_R R + C^R_C C + C^R_G P^{x,T}_G \right) - k_T R = 0$$

(C^{lp}_{2b})

$$\alpha_c \mu \left(\sum_{j=1}^{N_M} C^R_{M_j} E^x_j + C^R_R R + C^R_C C + C^R_G P^{x,T}_G \right) - k_C C = 0$$

(C^{lp}_{3a})

$$\sum_{j=1}^{N_{Mc}} C^D_{M_j} E^c_j + C^D_R R + C^D_C C + C^D_G P^{c,T}_G - \bar{D}_c \leq 0$$

(C^{lp}_{3b})

$$\sum_{j=1}^{N_{Ms}} C^S_{M_j} E^s_j + C^S_G P^{s,T}_G - \bar{D}_s \leq 0$$

(C^{lp}_4) for all $j \in I_m$,

$$\nu^x_j - k_{E_j} E^x_j \leq \text{ and } -(\nu^x_j + k_{E_j} E^x_j) \leq 0$$

$$\left|\nu^x{}_j\right| \leq k_{E_j} E_j$$

The RBA framework

- The feasibility problem is convex
- Equivalence with a Linear Programming (LP) optimization problem

  ⇨ same complexity as FBA, efficient resolution at genome scale!

- For a set of **given extracellular nutrient concentrations**, we can prove that there exists a **maximal growth rate value**
  - **without setting an objective function** (contrary to FBA);
  - **defined by a trade-off on the resource allocation** (especially on proteins);
  - for which a resource distribution (enzyme/ribosomes) exists;
  - and can be efficiently computed through the iterative resolution of LP optimization problems;

- Every mechanism saving resources increases the growth rate

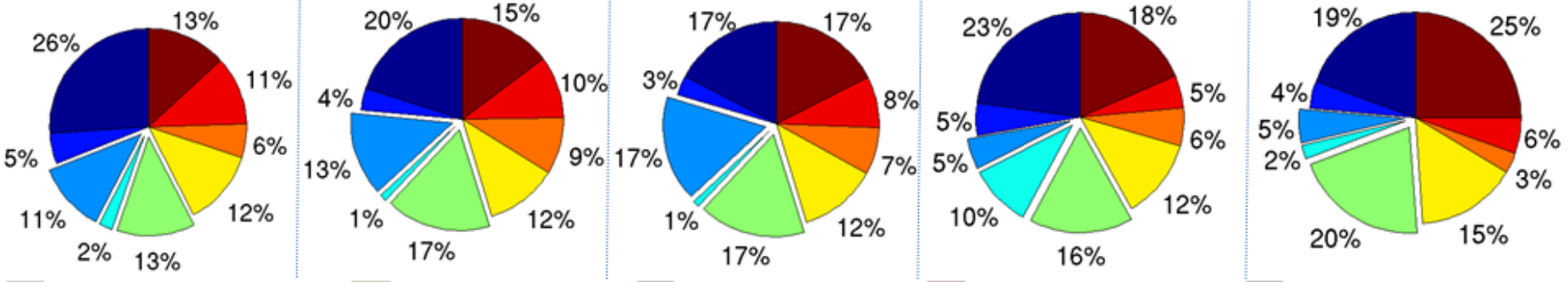
- Theoretical prediction of induced/repressed sub-systems in the metabolic network (towards the prediction of genetic regulations)

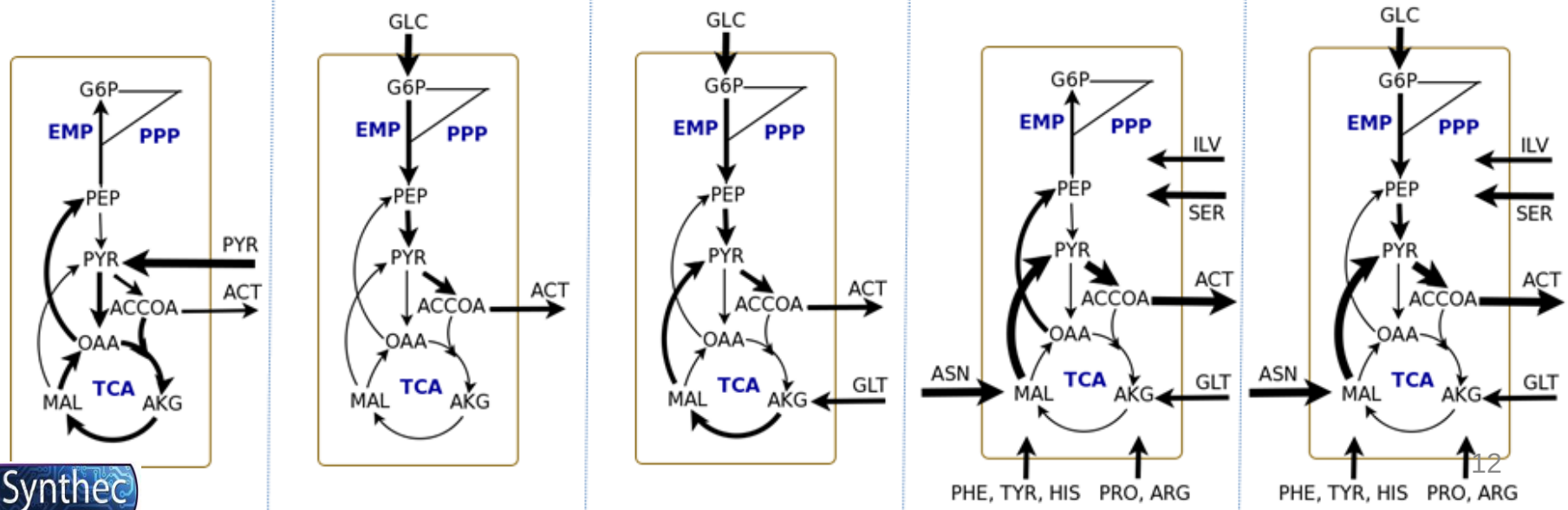
Data dedicated to RBA validation (5 conditions)

Data type	Group
Physiological data: length, width, volume	Inra - Jouy/Grignon
DNA concentration (5 conditions)	Inra – Grignon
Transcriptomic data	Inra – Grignon
Protein quantification (absolute) [quantification some key membrane proteins]	Greifswald
Concentration of ribosomes	Greifswald
Amount of total mRNAs	Greifswald/Inra - Grignon
mRNA half life	Greifswald
Polymerase activity (chip on chip)	Inra – Jouy
Metabolic data - external/internal	ETH – Zurich

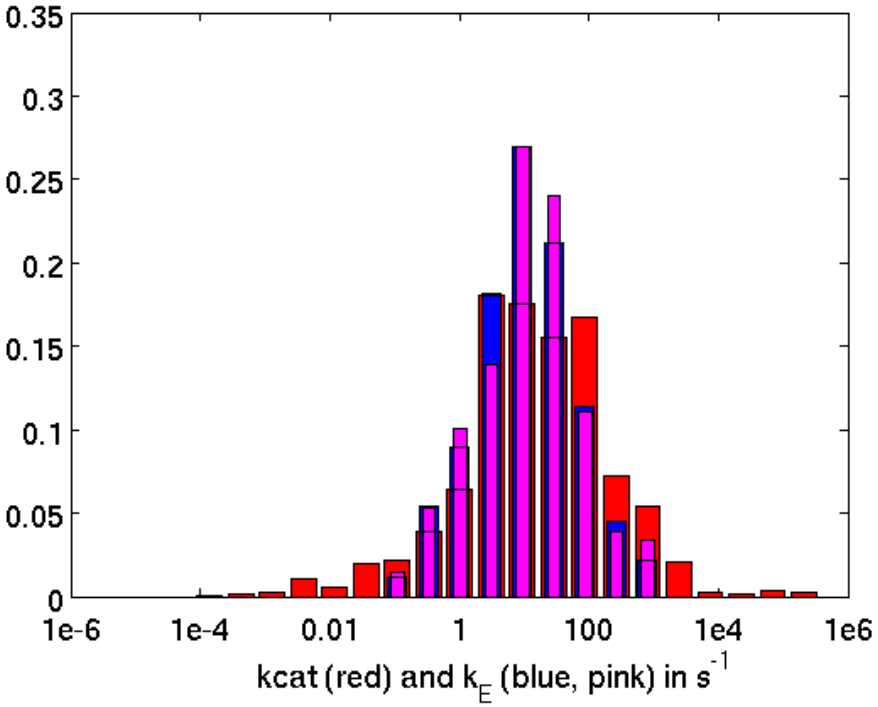
J. Muntel , et al. *Comprehensive absolute quantification of the cytosolic proteome of Bacillus subtilis by multiplexed LC/MS (LC/MSE).* Molecular and Cellular Proteomics, 13(4):1008-1019. **2014** .

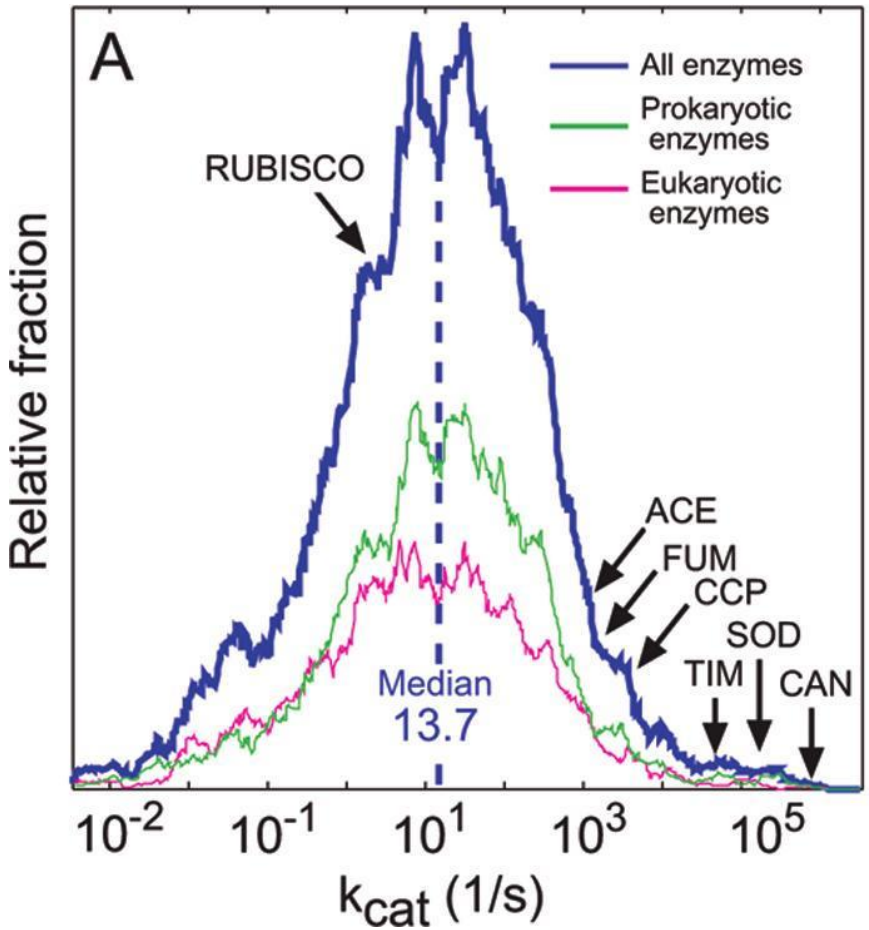
Minimal medium Pyruvate (PYR)	Minimal medium Glucose, Citrate (S)	Medium S + Glutamate (TS)	Complex medium 18 amino acids (CH)	Medium CH + Glucose (CHG)
μ: 0,3 h^{-1}	μ: 0,6 h^{-1}	μ:0,9 h^{-1}	μ: 1,1 h^{-1}	μ:1,5 h^{-1}
797 proteins	1127 proteins	1158 proteins	984 proteins	1103 proteins





“Consistency” with the expected distribution

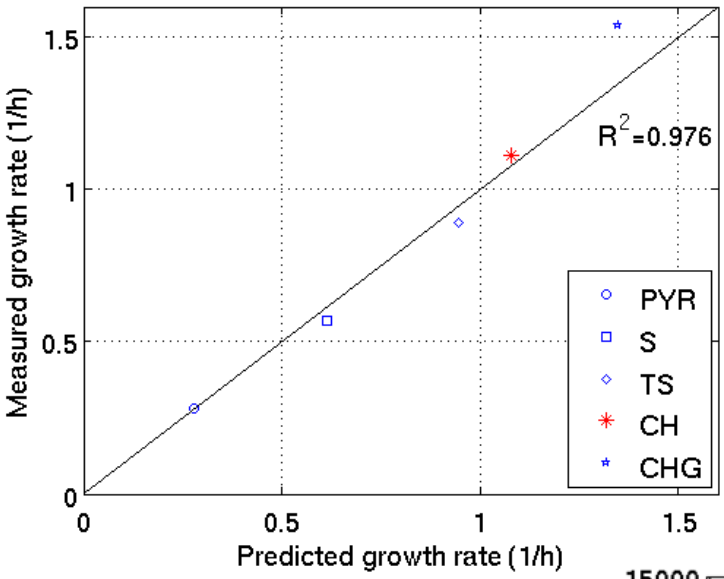


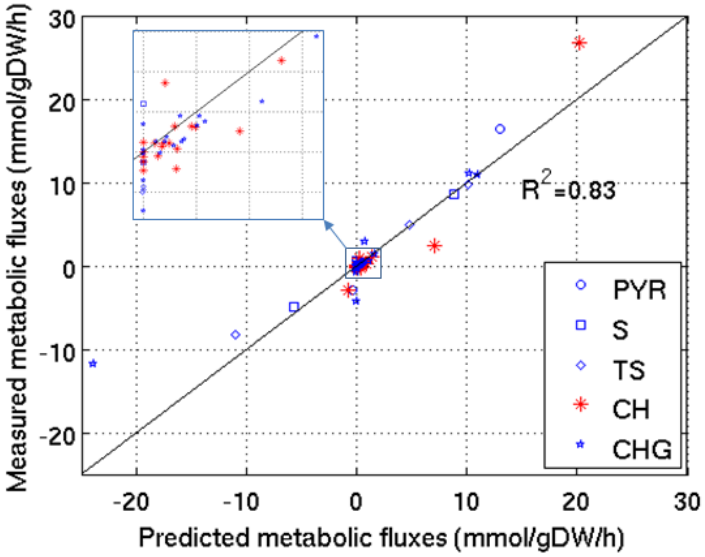


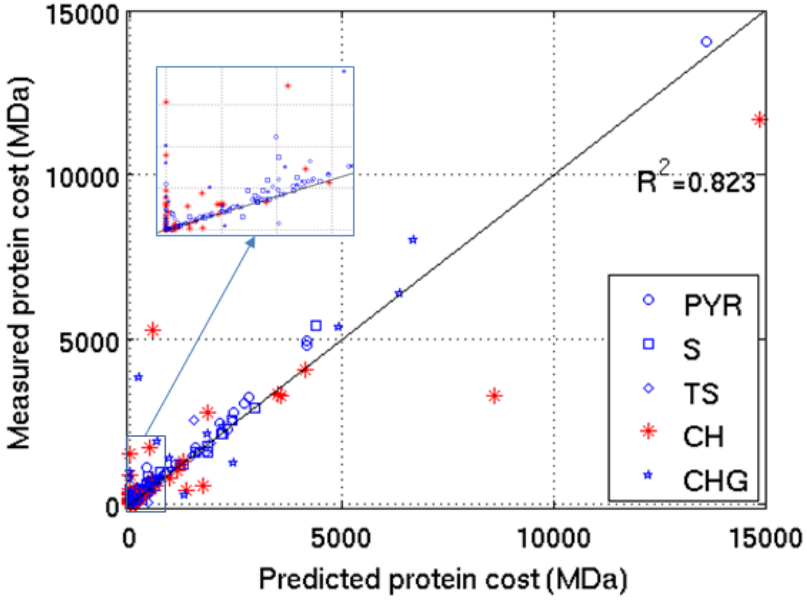
A. Bar-Even, et al. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters, Biochemistry, 2011, 50 (21), pp. 4402–4410

BaSynthec

RBA predictions for the 5 conditions







Goelzer et al. *Quantitative prediction of genome-wide resource allocation in bacteria. Metabolic Engineering. 32:232–243* ***2015***

Perspectives

- Extension of the RBA theoretical framework
  - ✓ To dynamical conditions (dRBA)
  - ✓ To stochastic fluctuations in gene expression
  - ✓ To include thermodynamics and kinetics constraints to predict the metabolite abundances
  - ✓ To predicts the emergence of regulatory networks
  - ✓ To handle multiple cells and multiple organisms

Exposé de Marc Dinh

Exposé de Laurent Tournier

- RBA for bioengineering and strain design

- Resource allocation for other prokaryotes and multi-cellular organisms
  - ✓Escherichia coli
  - ✓Synechocystis sp PCC6803
  - ✓Arabidopsis thaliana
  - ✓Maïze
  - ✓Animals like goat

Acknowledgments

INRA-MaIAGE

V. Fromion

M. Mariadassou, P. Nicolas, L. Tournier

Ecole Centrale de Lyon

G. Scorletti

INRA-Micalis

M. Jules, S. Aymerich (Grignon)

P. Noirot, E. Prestel, E. Dervyn (Jouy)

The BaSynthec partners

J. Muntel R. Nölker, D. Becher M. Hecker and U. Mader (Greifswald),

V. Chubukov and U. Sauer (ETHZ)