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

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 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

- Transport
- De novo synthesis by metabolic network

Amino acids

Ribosomes + tRNAs

ATP, GTP

Polypeptide

Chaperones

Ions, vitamins

Protein
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

(C1a) For all $i \in I_p$,

$$- \sum_{j=1}^{m} S_{p_{ij}} \nu^x_j + \mu \left( \sum_{j=1}^{m} C_{M_{ij}}^{M_p} |\nu^x_j| + C_{R_i}^{M_p} R + C_{C_i}^{M_p} C + C_{G_i}^{M_p} P_G^{x,T} \right) - \nu_Y = 0$$

(C1b) For all $i \in I_c$,

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

(C1c) For all $i \in I_r$,

$$\sum_{j=1}^{m} S_{r_{ij}} \nu^x_j + \mu \left( \sum_{j=1}^{m} C_{M_{ij}}^{M_r} |\nu^x_j| + C_{R_i}^{M_r} R + C_{C_i}^{M_r} C + C_{G_i}^{M_r} P_G^{x,T} \right) = 0$$

(C1d) For all $i \in I_i$,

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

(C2a) $\mu \left( \sum_{j=1}^{m} C_{M_{ij}}^{R} |\nu^r_j| + C_{R_i}^{R} R + C_{C_i}^{R} C + C_{G_i}^{R} P_G^{r,T} \right) - k_TR = 0$

(C2b) $\alpha_c \mu \left( \sum_{j=1}^{m} C_{M_{ij}}^{C} |\nu^c_j| + C_{R_i}^{C} R + C_{C_i}^{C} C + C_{G_i}^{C} P_G^{c,T} \right) - k_CC = 0$

(C3a) $\sum_{j=1}^{m} C_{M_{ij}}^{D} |\nu^c_j| + C_{R_i}^{D} R + C_{C_i}^{D} C + C_{G_i}^{D} P_G^{c,T} - \bar{D}_c \leq 0$

(C3b) $\sum_{j=1}^{m} C_{M_{ij}}^{S} |\nu^s_j| + C_{G_i}^{S} P_G^{s,T} - \bar{D}_s \leq 0$

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 $R \geq 0, C \geq 0, \nu^x \in \mathbb{R}^m, E^x \in \mathbb{R}_+^m$.

subject to

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

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

(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 $i \in I_r$,

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

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

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

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

$$\mu \left( \sum_{j=1}^{N_M} C_{M_j}^R E^x_j + C^R R + C^R C + C^R P_{G}^{x,T} \right) - k_T R = 0$$

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

$$\alpha_c \mu \left( \sum_{j=1}^{N_M} C_{M_j}^R E^x_j + C^R R + C^R C + C^R P_{G}^{x,T} \right) - k_C C = 0$$

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

$$\sum_{j=1}^{N_M} C_{M_j}^D E^c_j + C^D R + C^D C + C^D P_{G}^{c,T} - \bar{D}_c \leq 0$$

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

$$\sum_{j=1}^{N_M} C_{M_j}^S E^s_j + C^S P_{G}^{s,T} - \bar{D}_s \leq 0$$

(C$^{lp}_{4}$) for all $j \in I_m$,

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

$|\nu^x_j| \leq k_{E_j} E_j$
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)

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

<table>
<thead>
<tr>
<th>Minimal medium Pyruvate (PYR)</th>
<th>Minimal medium Glucose, Citrate (S)</th>
<th>Medium S + Glutamate (TS)</th>
<th>Complex medium 18 amino acids (CH)</th>
<th>Medium CH + Glucose (CHG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>μ: 0.3 h⁻¹</td>
<td>μ: 0.6 h⁻¹</td>
<td>μ: 0.9 h⁻¹</td>
<td>μ: 1.1 h⁻¹</td>
<td>μ: 1.5 h⁻¹</td>
</tr>
<tr>
<td>797 proteins</td>
<td>1127 proteins</td>
<td>1158 proteins</td>
<td>984 proteins</td>
<td>1103 proteins</td>
</tr>
</tbody>
</table>

![Pie charts representing protein distribution](chart.png)

**Legend for Pie Charts:**
- 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

**Diagram Flowcharts:**
1. GLC → EMP → G6P → PPP
2. GLC → EMP → G6P → PPP
3. GLC → EMP → G6P → PPP
4. GLC → EMP → G6P → PPP

**Pathways:**
- EMP
- PPP
- TCA
- GLT
- ILV
- SER
- ACT
- ASN
- PHE, TYR, HIS, PRO, ARG
- MAL
- AKG
- ACCOA
- OAA
- PYR
- GLT
- G6P
- ACT
“Consistency” with the expected distribution

RBA predictions for the 5 conditions

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

- 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)