LES PROGRÈS RÉCENTS DANS LA MODÉLISATION SYSTÉMIQUE DES CELLULES BACTÉRIENNES OUVENT DES PERSPECTIVES IMPORTANTES EN BIOLOGIE DE SYNTHÈSE

Vincent Fromion, MalAGE, INRA-Jouy

Vincent.fromion@jouy.inra.fr
How a rational design of industrial strains can be achieved?

Biomass production

Metabolite/Protein production
Cell Factory

MODEL

UNDERSTAND
PREDICT

TEST

GENOMICS
TRANSCRIPTOMICS
METABOLOMICS
FLUXOMICS…

BUILD
ENGINEER
TRANSFORM
OPTIMIZE

PRODUCE
The pipeline for Cell Factory design

A computational pipeline

RBA

Optimal profile of gene expression

Modular decomposition of the cell

Minimal set of target genes to be retuned

Selection of biobricks (reverse engineering)
Based on a bank of promoter/TIR sequences

Biobricks/genes

Strain implementation

Optimized strain


Borkowski et al. Translation elicits a growth-rate-dependent and genome-wide, differential production of proteins in Bacillus subtilis. Submitted in Molecular Systems Biology

A (new) computational method predicting growth rate, protein abundance and metabolic fluxes
Resources (especially proteins) have to be shared by all biological processes (implicit feedback).
Three (main) structural constraints

The mass conservation and the stoichiometry of the metabolic network.
Resource Balance Analysis (RBA)

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

\[
\begin{align*}
\text{Find} & \quad R \geq 0, C \geq 0, \nu^x \in \mathbb{R}^m, \\
\text{subject to} & \quad \text{For all } i \in I_p,
- \sum_{j=1}^{m} S_{pj} \nu^x_j + \mu \left( \sum_{j=1}^{m} C_{M_{ij}}^{R} |\nu^x_j| + C_{R_i}^{M_p} R + C_{C_i}^{M_p} C + C_{G_i}^{M_p} P_{G}^{x,T} \right) - \nu_Y = 0 \\
& \quad \text{For all } i \in I_c,
- \sum_{j=1}^{m} S_{ci} \nu^x_j + \mu \bar{X}_c = 0 \\
& \quad \text{For all } i \in I_r,
\sum_{j=1}^{m} S_{ri} \nu^x_j + \mu \left( \sum_{j=1}^{m} C_{M_{ij}}^{R} |\nu^x_j| + C_{R_i}^{R} R + C_{C_i}^{M_r} C + C_{G_i}^{M_r} P_{G}^{x,T} \right) = 0 \\
& \quad \text{For all } i \in I_a,
\sum_{j=1}^{m} S_{ai} \nu^x_j = 0 \\
& \quad \mu \left( \sum_{j=1}^{m} C_{M_j}^{R} |\nu^x_j| + C_{R}^{R} R + C_{C}^{R} C + C_{G}^{R} P_{G}^{x,T} \right) - k_T R = 0 \\
& \quad \alpha_c \mu \left( \sum_{j=1}^{m} C_{M_j}^{R} |\nu^x_j| + C_{R}^{R} R + C_{C}^{R} C + C_{G}^{R} P_{G}^{x,T} \right) - k_C C = 0 \\
& \quad \sum_{j=1}^{m} C_{M_j}^{D} \nu^x_j + C_{R}^{D} R + C_{C}^{D} C + C_{G}^{D} P_{G}^{x,T} - D_c \leq 0 \\
& \quad \sum_{j=1}^{m} C_{M_j}^{S} \nu^x_j + C_{G}^{S} P_{G}^{x,T} - D_s \leq 0
\end{align*}
\]

Biological validation of RBA

EU–BaSynthec project
## 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)</td>
<td>Greifswald</td>
</tr>
<tr>
<td>[quantification some key membrane proteins]</td>
<td></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>
<tr>
<td>Medium</td>
<td>Protein Count</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Minimal medium Pyruvate (PYR)</td>
<td>797 proteins</td>
</tr>
<tr>
<td>Minimal medium Glucose, Citrate (S)</td>
<td>1127 proteins</td>
</tr>
<tr>
<td>Medium S + Glutamate (TS)</td>
<td>1158 proteins</td>
</tr>
<tr>
<td>Complex medium 18 amino acids (CH)</td>
<td>984 proteins</td>
</tr>
<tr>
<td>Medium CH + Glucose (CHG)</td>
<td>1103 proteins</td>
</tr>
</tbody>
</table>

**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
### RBA calibration

#### Training datasets

<table>
<thead>
<tr>
<th>Minimal medium</th>
<th>Minimal medium</th>
<th>Medium S</th>
<th>Complex medium</th>
<th>Medium CH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyruvate (PYR)</td>
<td>Glucose, Citrate (S)</td>
<td>S + Glutamate (TS)</td>
<td>18 amino acids (CH)</td>
<td>CH + Glucose (CHG)</td>
</tr>
<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>
</tbody>
</table>

- **797 proteins**
- **1127 proteins**
- **1158 proteins**
- **984 proteins**
- **1103 proteins**

#### Metabolic Pathways

- **Central Carbon Metabolism**
- **Neither**

---

**Diagram:**

- **G6P EMP PPP**
- **PEP PYR TCA TCA AKG**
- **PYR ACCOA ACT**
- **OAA MAL GLT**
- **GLC ILV SER ACT**
- **PHE, TYR, HIS PRO, ARG**
<table>
<thead>
<tr>
<th>Condition</th>
<th>µ (h⁻¹)</th>
<th>Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal medium Pyruvate (PYR)</td>
<td>0.3 h⁻¹</td>
<td>797</td>
</tr>
<tr>
<td>Minimal medium Glucose, Citrate (S)</td>
<td>0.6 h⁻¹</td>
<td>1127</td>
</tr>
<tr>
<td>Medium S + Glutamate (TS)</td>
<td>0.9 h⁻¹</td>
<td>1158</td>
</tr>
<tr>
<td>Complex medium 18 amino acids (CH)</td>
<td>1.1 h⁻¹</td>
<td>984</td>
</tr>
<tr>
<td>Medium CH + Glucose (CHG)</td>
<td>1.5 h⁻¹</td>
<td>1103</td>
</tr>
</tbody>
</table>

**RBA validation**

- Central Carbon Metabolism
- Neither

**Chemical Pathways:**
- GLC
- EMP
- PPP
- G6P
- GLT
- PHE, TYR, HIS, PRO, ARG

**Other metabolic pathways:**
- Basal Apparatus
RBA predictions for the 5 conditions

Towards a prediction of “optimal gene expression” profiles at genome scale...

Predicted consumption of glucose (+) and sorbitol (α)

Predicted enzyme abundances

Growth rate (1/θ)

Biomass (g)
The pipeline for Cell Factory design

RBA

Optimal profile of gene expression

Modular decomposition of the cell

Minimal set of target genes to be retuned


What happens in the cell?

- Metabolic network (flux/pool)
- States used for the genetic control
- Genetic regulation (Transcription Factor)
- Modification of enzyme concentrations
- Perturbations

Growth rate management

(RNA polymerases, Ribosomes, ppGpp, etc.)
A first knowledge-based model

Goelzer et al., *BMC Systems Biology, 2*:20, 2008

Each metabolic pathway integrates
- the kinetic reactions and their known enzymatic regulations,
- the known transcriptional, translational and post-translational regulations and their metabolite effectors, that have been experimentally validated,
- the organization of genes in operon,
- the Boolean conditions of transcription and translation for each gene.

The model integrated (2008)
- 622 reactions,
- 587 genes and 67 transcription factors, 21 other genetic regulations,
- more than 400 references,

The current model is now at the genome scale and integrates more than 200 genetic regulations in particular the regulatory networks of various stress (oxidative, heat, iron, etc.), the growth rate management loop, etc.
The regulations of the metabolic network are organized!

Functional modules can be identified
Two local control structures
Coordination of modules by global regulation
The pipeline for Cell Factory design


Modular decomposition of the cell

RBA

Optimal profile of gene expression

Minimal set of target genes to be retuned

The regulatory network enables to break down the metabolic network in functional modules

But it is not sufficient to the dynamics of all proteins/genes with respect to growth rate (e.g. constitutive genes)
Hardware implementation of the growth rate regulation for each gene

Where? in promoter/TIR sequence of each gene
Hardware implementation of growth rate management
A simplified model of the whole process

Dynamical models of transcription and translation
The steady-state for the transcription

\[ Poly_f + prom_{\text{fi}} \xrightleftharpoons[k_{a1}']{k_{-a1}'} OC_{ai} \rightarrow Poly_{ai} \rightarrow Poly_{ci} \rightarrow m_i + Poly_f \]

Transcription Initiation \quad Elongation

\[ m_i = \frac{k_{2i}' Poly_f n_{\text{promoter}}}{(1 + k_{1i}' Poly_f) k_d} \]

Messenger Degradation
The steady-state for the translation

\[ P_i = \frac{k_{2i} C_f m_i}{(1 + k_{1i} C_f) \mu} \]

with

\[ k_{1i} = \frac{k_{ai}}{k_{-ai} + k_{tai}} \]
\[ k_{2i} = \frac{k_{tai} k_{ai}}{k_{-ai} + k_{tai}} \]
# Identification of parameters from data

(Data acquisition in 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)</td>
<td>Greifswald</td>
</tr>
<tr>
<td>[quantification some key membrane proteins]</td>
<td></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>
Identification of parameters from data

**Transcription:** parameters $k'_{1i}$, $k'_{2i}$ for each gene and $\text{Poly}_f$ can be estimated from data.

\[
\frac{k_d}{m_i} = \frac{k'_{2i} \text{Poly}_f}{(1 + k'_{1i} \text{Poly}_f)} \quad \text{All these quantities are measured (3 conditions)}
\]

A bank of promoters + model relating DNA sequence to mRNA abundance

**Translation:** parameters $k_{1i}$, $k_{2i}$ for each mRNA and $C_f$ can be estimated from data.

\[
\frac{\mu P_i}{m_i} = \frac{k_{2i} C_f}{(1 + k_{1i} C_f)} \quad \text{All these quantities are measured (5 conditions)}
\]

A bank of TIR sequences + model relating mRNA abundance to protein abundance
The pipeline for Cell Factory design

RBA

Optimal profile of gene expression

Modular decomposition of the cell

Minimal set of target genes to be retuned

Selection of biobricks (reverse engineering)
Based on a bank of promoter/TIR sequences

Biobricks/genes

Goelzer, A., Fromion and G. Scorletti *Cell design in bacteria as a convex optimization problem.*
48th IEEE Conference on Decision and Control, China, 4517 -22. 2009

Goelzer, A, V. Fromion and G. Scorletti *Cell design in bacteria as a convex optimization problem.*


Borkowski et al. *Translation elicits a growth-rate-dependent and genome-wide, differential production of proteins in Bacillus subtilis.* Submitted in Molecular Systems Biology
The pipeline for Cell Factory design

A computational pipeline

The pipeline for Cell Factory design

- **RBA**
  - Optimal profile of gene expression
  - Modular decomposition of the cell
    - Minimal set of target genes to be retuned
      - Selection of biobricks (reverse engineering)
        - Based on a bank of promoter/TIR sequences
        - Biobricks/genes
          - Strain implementation
            - Optimized strain

References:

Transfer to other bacteria

Requirements to build the pipeline components for other bacteria

- Annotated sequence, metabolic network reconstruction
- The known genetic and metabolic regulatory network
- A training dataset

RBA model
Module decomposition of the cell
Bank of promoter and TIR seq.

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

INRA-MaIAGE
V. Fromion
P. Nicolas
O. Borkowski
M. Mariadassou, L. Tournier

Ecole Centrale de Lyon
G. Scorletti

INRA-Micalis
M. Jules, V. Sauveplane, H. Bouraoui, H. Bautista, S. Aymerich
E. Prestel, E. Dervyn P. Noirot

The BaSyntheC partners
J. Muntel R. Nölker, D. Becher M. Hecker, U. Mäder (Greifswald),
V. Chubukov, U. Sauer (ETHZ)