INRA SCIENCE & IMPACT

Mathématiques et Informatique Appliquées
MaIAGE
du Génome à l'Environnement

First step towards the modeling of stochastic gene expression effect on resource allocation and bacterial growth

Marc Dinh, MaIAGE, INRA Jouy-en-Josas

Context

- Observations:
  - Bacteria has limited (external) resources to growth
  - In steady-state conditions, (colony) growth is exponential

- Main idea of resource allocation and growth rate modeling

Bacteria allocates their limited resources
optimally to **maximize** their growth rate

- RBA (Resource Balance Analysis) framework: idea successfully applied and validated [1]

[1] Goelzer et al., Quantitative prediction of genome-wide resource allocation in bacteria, Metabolic Engineering, 2015

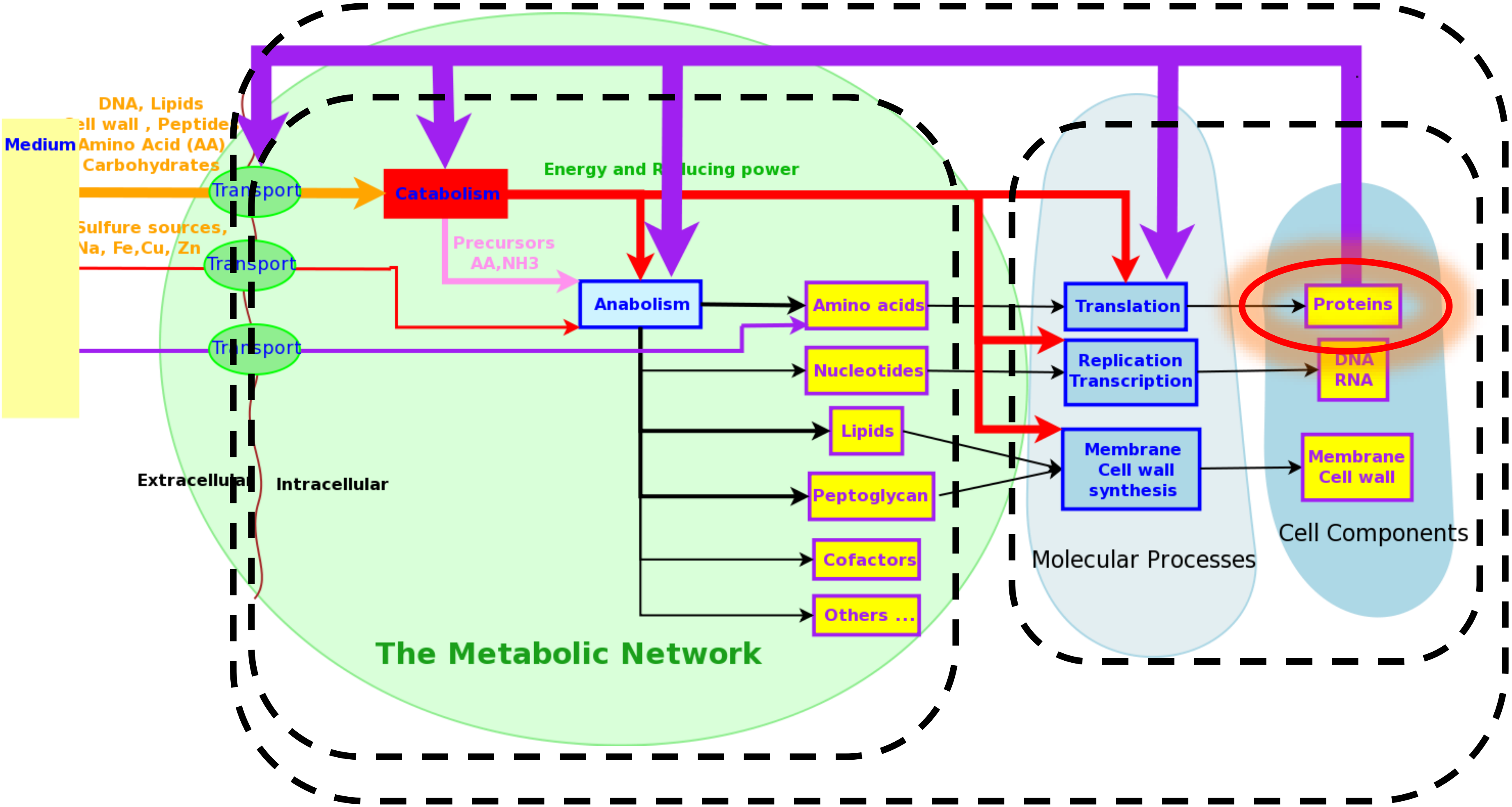
Objective

- RBA based on a deterministic assumption of gene expression: the bacteria is capable of producing an exact number (the optimal one) of each type of proteins
- Observation: protein production (gene expression) is stochastic
- Objective

Add **stochastic** behavior of gene expression in the modeling of resource allocation and growth rate

- Presentation of ongoing work (on metabolism)

Brief overview of RBA (1/2)



Brief overview of RBA (2/2)

find $R \geqslant 0, C \geqslant 0, \nu \in \mathcal{R}^m, E \in \mathcal{R}_+^m$

subject to

C1

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

$$-\sum_{j=1}^{N_M} S_{p_{ij}} \nu_j^x + \mu\left(\sum_{j=1}^{N_M} C_{M_{ij}}^{M_p} E_j^x + C_{R_i}^{M_p} R + C_{C_i}^{M_p} C + \sum_{j=1}^{N_G} C_{G_{ij}}^{M_p} P_{G_j}^x\right) - \nu_Y = 0$$

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

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

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

$$\sum_{j=1}^{N_M} S_{r_{ij}} \nu_j^x + \mu\left(\sum_{j=1}^{N_M} C_{M_{ij}}^{M_r} E_j^x + C_{R_i}^{M_r} R + C_{C_i}^{M_r} C + \sum_{j=1}^{N_G} C_{G_{ij}}^{M_r} P_{G_j}^x\right) = 0$$

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

$$\sum_{j=1}^{m} S_{I_{ij}} \nu_j^x = 0$$

C2

(C_{2a}^{lp}) $$\mu\left(\sum_{j=1}^{N_M} C_{M_j}^R E_j^x + C_R^R R + C_C^R C + \sum_{j=1}^{N_G} C_{G_j}^R P_{G_j}^x\right) - k_T(\mu) R = 0$$

(C_{2b}^{lp}) $$\alpha_c \mu\left(\sum_{j=1}^{N_M} C_{M_j}^R E_j^x + C_R^R R + C_C^R C + \sum_{j=1}^{N_G} C_{G_j}^R P_{G_j}^x\right) - k_C(\mu) C = 0$$

C3

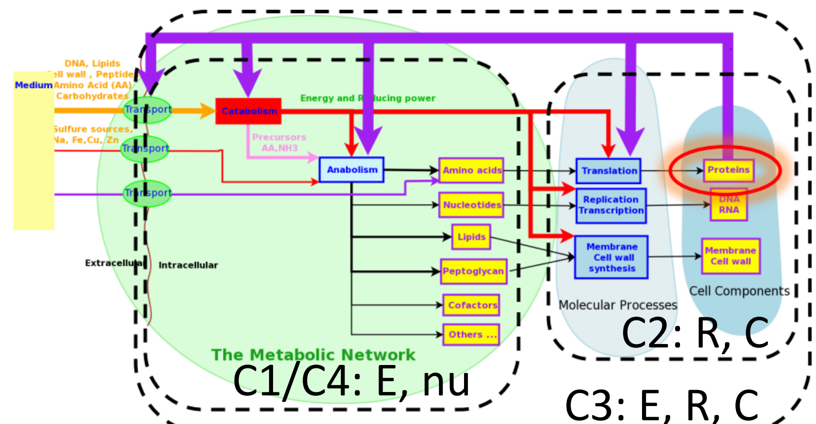
(C_{3a}^{lp}) $$\sum_{j=1}^{N_{M_c}} C_{M_j}^D E_j^c + C_R^D R + C_C^D C + \sum_{j=1}^{N_{G_c}} C_{G_j}^D P_{G_j}^c - \bar{D}_c \leqslant 0$$

(C_{3b}^{lp}) $$\sum_{j=1}^{N_{M_s}} C_{M_j}^S E_j^s + \sum_{j=1}^{N_{G_s}} C_{G_j}^S P_{G_j}^s - \bar{D}_s \leqslant 0$$

C4

(C_4^{lp}) *for all* $j \in I_m$,

$$|\nu_j^x| \leqslant k_{E_j} E_j^x$$



!Simplification to add stochastic behavior!
Metabolism only

Modeling of stochasticity

- Assumptions:
  - Biomass composition is known: c is known
  - Stochastic behavior follows an exponential distribution
- Modeling as a stochastic optimization problem

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

subject to

(C^{lp}_{1a}) *for all* $i \in I_p$,
$$-\sum_{j=1}^{N_M} S_{p_{ij}} \nu_j^x + \mu \left(\sum_{j=1}^{N_M} C^{M_p}_{M_{ij}} E_j^x + C^{M_p}_{R_i} R + C^{M_p}_{C_i} C + \sum_{j=1}^{N_G} C^{M_p}_{G_{ij}} P^x_{G_j} \right) - \nu_Y = 0$$

(C^{lp}_{1b}) *for all* $i \in I_c$,
$$-\sum_{j=1}^{N_m} S_{c_{ij}} \nu_j^x + \mu \bar{X}_{c_i} = 0$$

(C^{lp}_{1c}) *for all* $i \in I_r$,
$$\sum_{j=1}^{N_M} S_{r_{ij}} \nu_j^x + \mu \left(\sum_{j=1}^{N_M} C^{M_r}_{M_{ij}} E_j^x + C^{M_r}_{R_i} R + C^{M_r}_{C_i} C + \sum_{j=1}^{N_G} C^{M_r}_{G_{ij}} P^x_{G_j} \right) = 0$$

(C^{lp}_{1d}) *for all* $i \in I_i$,
$$\sum_{j=1}^{m} S_{I_{ij}} \nu_j^x = 0$$

(C^{lp}_{2a})
$$\mu \left(\sum_{j=1}^{N_M} C^R_{M_j} E_j^x + C^R_R R + C^R_C C + \sum_{j=1}^{N_G} C^R_{G_j} P^x_{G_j} \right) - k_T(\mu) R = 0$$

(C^{lp}_{2b})
$$\alpha_c \mu \left(\sum_{j=1}^{N_M} C^R_{M_j} E_j^x + C^R_R R + C^R_C C + \sum_{j=1}^{N_G} C^R_{G_j} P^x_{G_j} \right) - k_C(\mu) C = 0$$

(C^{lp}_{3a})
$$\sum_{j=1}^{N_{M_c}} C^D_{M_j} E_j^c + C^D_R R + C^D_C C + \sum_{j=1}^{N_{G_c}} C^D_{G_j} P^c_{G_j} - \bar{D}_c \leqslant 0$$

(C^{lp}_{3b})
$$\sum_{j=1}^{N_{M_s}} C^S_{M_j} E_j^s + \sum_{j=1}^{N_{G_s}} C^S_{G_j} P^s_{G_j} - \bar{D}_s \leqslant 0$$

(C^{lp}_4) *for all* $j \in I_m$,
$$|\nu_j^x| \leqslant k_{E_j} E_j^x$$

$$\max_{E \in \mathbb{R}^n_+,\ \nu \in \mathbb{R}^n} \mathbb{E}_{\tilde{E} \sim \exp(E)} (\max(c^T \nu))$$

$$s.t. \begin{cases} S\nu = 0 \\ \sum \alpha_i E_i \leq cost \\ |\nu| \leq k\tilde{E} \end{cases}$$

Modeling as a two stage problem

$\tilde{E} \sim \exp(E) \Leftrightarrow \tilde{E} = E\xi \; with \; \xi \sim \exp(1)$: same distribution

$$\max_{E\in\mathbb{R}^n_+,\, \nu\in\mathbb{R}^n} \mathbb{E}_{\tilde{E} \sim \exp(E)}(\max(c^T\nu))$$

$$s.t. \begin{cases} S\nu = 0 \\ \Sigma\, \alpha_i E_i \leq cost \\ |\nu| \leq k\tilde{E} \end{cases}$$

$$\max_{E\in\mathbb{R}^n_+} \left\{ f(E) = \mathbb{E}_{\xi \sim \exp(1)}\big(g(E,\xi)\big) \right\}$$

$$s.t. \; \Sigma\, \alpha_i E_i \leq cost$$

$$g(E,\xi) = \max_{\nu\in\mathbb{R}^n}(c^T\nu)$$

$$s.t. \begin{cases} S\nu = 0 \\ |\nu| \leq kE\xi \end{cases}$$

- Two-stage with recourse problem [2]
- Algorithm exists if a stochastic sub-gradient of f can be computed [3]

[2] R.J.B. Wets, Programming under uncertainty: the equivalent convex problem, SIAM Applied Mathematics, 1966

[3] Nemirovski et al., Robust stochastic approximation approach to stochastic programming, SIAM Journal Optimization, 2009

Algorithm

- $Definition$: $G(E,\xi)$ $is\ a\ stochastic\ subgradient\ of\ f$
  $$if\ \mathbb{E}_{\xi} G(E,\xi) \in \partial f$$
- Sketch of algorithm used [3]
  - Subgradient type with iteration of the form $E(k+1) = projection\ onto\ first\ stage\ feasible\ set\ of \left(E(k) - \gamma(k) G\left(E(k), \xi(k)\right)\right)$ where $\xi(k)$ is a realization
  - Upper bound on absolute optimal value error of the from
  $$\frac{a + b\Sigma_k \gamma^2(k)}{c\Sigma_k \gamma(k)}$$
  - Converges if $\frac{\Sigma_k \gamma^2(k)}{\Sigma_k \gamma(k)} \longrightarrow 0$
- Exponential distribution: problem is convex and computation of stochastic subgradient is possible

Results

- "Full rank" network: it is possible to parameterize all the network fluxes by one flux
  $Sv = 0 \iff v = Mv_0$
- Analytical solution can be obtained

$$v_0^{sto} = \frac{cost}{\left(\sum_{i=1,\dots,n}\sqrt{\alpha_i/k_i}\right)^2} < v_0^{det} = \frac{cost}{\sum_{i=1,\dots,n}\alpha_i/k_i}$$

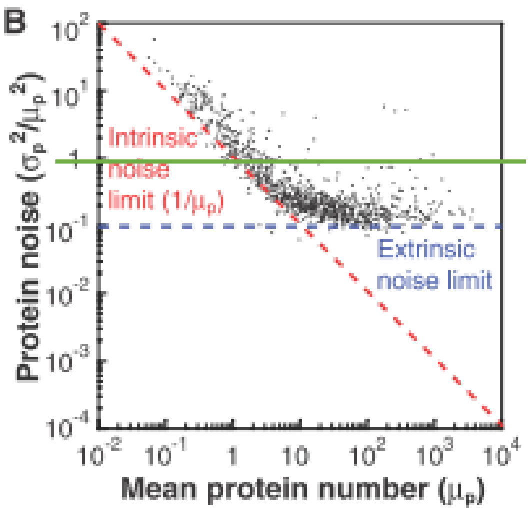
$$E_i^{sto} = \frac{\sqrt{\alpha_i/k_i}\,cost}{\sum_{i=1,\dots,n}\sqrt{\alpha_i/k_i}} \qquad E_i^{det} = \frac{(\alpha_i/k_i)cost}{\sum_{i=1,\dots,n}\alpha_i/k_i}$$

- Numerical experiments are consistent with analytical results on a "full rank" CCM network

Improvements

- Exponential distribution assumption: not consistent with observations

Taniguchi et al., Quatifying *E. Coli* proteome and transcriptome with single-molecule sensitivity in single cells, Science, 2011



- Behavior of algorithm on non "full rank" networks and results on whole metabolism
- How to include the other processes and how to compute a solution