



Cell design in bacteria as a convex optimization problem

Anne Goelzer, Vincent Fromion, Gérard Scorletti

► **To cite this version:**

Anne Goelzer, Vincent Fromion, Gérard Scorletti. Cell design in bacteria as a convex optimization problem. 48. IEEE Conference on Decision and Control and 28th Chinese Control Conference, Dec 2009, Shanghai, China. hal-02758318

HAL Id: hal-02758318

<https://hal.inrae.fr/hal-02758318>

Submitted on 4 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cell design in bacteria as a convex optimization problem

Anne Goelzer, Vincent Fromion and Gérard Scorletti

Abstract—In this paper, we investigate the cell design of bacteria during the exponential growth. To this purpose, we propose to formulate the problem as a non differentiable convex optimization problem equivalent to a Linear Programming feasibility problem. Its resolution predicts for a specific medium not only the distribution of metabolic fluxes and the maximal growth rate, but also the concentrations of the ribosomes and the proteins involved in the metabolic network and thus the composition of the cell for different growth rates. Moreover, our model recovers the known modular structure of the regulation of metabolic pathways for the gram-positive model bacterium *Bacillus subtilis*.

I. INTRODUCTION

A challenging question in System Biology is to understand the organization of the regulations in the cell and to identify the rules that have led to the emergence of this organization.

In this paper, we present the second step of our investigations on the metabolic network regulations. Due to the high complexity of this biological network and to the interdisciplinary nature of the problem, our approach is based on a strong interaction between biological and automatic control concepts.

The first step of our approach was presented in [1]: a model of the metabolic network of *B. subtilis* was proposed by the two first authors with a qualitative (functional) analysis of the network. In [1], the metabolic network of *B. subtilis* was decomposed into elementary functional modules locally controlled. We revealed that these modules are further coordinated by so-called global regulations in response to physiological changes, leading to a strong modular organization of the regulation of the metabolic network. In consequence, we focus in this paper on the investigation of possible design rules/constraints that led to this modular organization through the evolution of organisms. To this purpose, we develop a quantitative mathematical model in order to analyze some cell behaviors and to identify these structural constraints if they exist.

Inside the cell, the metabolic network produces the metabolic precursors and energy necessary for the growth. For microorganisms, an emerging principle proposed for the design of the metabolic network is the maximization of the growth with respect to a given extracellular medium since this aspect is crucial in the context of the competition

between bacteria. This principle led to the development of the so-called Flux Balance Analysis (FBA) approach and the following Linear Programming (LP) problem [2]:

$$\begin{aligned} & \text{maximize} && c^T v \\ & \text{subject to} && \\ & && S \cdot v = 0 \\ & && \alpha_i \leq v_i \leq \beta_i. \end{aligned}$$

where the metabolic network is mathematically represented by the stoichiometric matrix S , linking all the metabolic fluxes (v) to the metabolites in steady-state ($S \cdot v = 0$). The objective function c is usually chosen as the mean composition of the cell at a given level. Although the FBA approach was experimentally validated on several organisms [3], [4], the cell design problem is simplified since the cell composition varies with the growth rate [5].

We then propose in this paper to extend the FBA approach by considering that the cell is composed of subsystems, whose role is to carry out some specific functions like the production of metabolites and proteins, or the DNA duplication, etc. and by looking at the resource allocation among all these subsystems. This resource allocation and in particular the sharing of proteins (*i.e.* the main building blocks of the cell) between these subsystems impose new and strong constraints which are not included in the FBA approach. We capture this problem of resource allocation into a non differentiable convex optimization problem, which can be transformed into a LP feasibility problem, efficiently solved even for large-scale problems. Interesting properties of the solution are investigated and discussed from a biological point of view. In particular, we show that the chosen formulation provides interesting insights on the strong modularity of the metabolic network in bacteria.

The paper is organized as follows. In section II, the cell design constraints are discussed and the mathematical model is proposed as a convex optimization problem. Section III presents the analysis of the optimization problem and the properties that can be deduced. Section IV focuses on the prediction of the modular structure of the metabolic network in bacteria. Finally, the resolution of the optimization problem for the gram-positive model bacterium *B. subtilis* is presented in section V. For the sake of clarity, all proofs are reported in the internal note [6].

II. CELL DESIGN CONSTRAINTS

A systemic view of the cell is displayed on Figure 1. The metabolic network, composed of proteins (*i.e.* enzymes), degrade the nutrients imported inside the cell in order to produce the metabolic precursors required for the synthesis

A. Goelzer is with Institut National de la Recherche en Agronomie, Unité de Mathématique, Informatique et Génome UR1077, F-78350 Jouy-en-Josas, France anne.goelzer@jouy.inra.fr

V. Fromion is with Institut National de la Recherche en Agronomie, Unité de Mathématique, Informatique et Génome UR1077, F-78350 Jouy-en-Josas, France vincent.fromion@jouy.inra.fr

G. Scorletti is with Laboratoire Ampère, École Centrale de Lyon - BP163 - 69131 Ecully, France gerard.scorletti@ec-lyon.fr

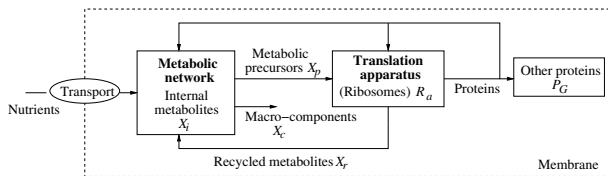


Fig. 1. A systemic view of the cell

of all the cell components, and for the protein synthesis in particular. Various assembling processes consume these precursors to produce all the cell components (proteins, the cell wall, DNA, the lipid membrane, etc) while producing metabolites that are recycled by the metabolic network. These assembling processes are also mainly composed of proteins and sometimes other cell components. For example, the ribosome, a key actor of the protein production, is composed of proteins and rRNA.

So the proteins produced by the ribosomes are a critical resource that need to be shared by all the cellular processes inside the cell: the metabolic network (*i.e.* enzymes), the translation apparatus itself (*i.e.* ribosomal proteins, elongation factors, etc.), and all the other proteins involved in the molecular assembling processes or in biological processes such as stress, or preparation to the stationary phase. During the exponential growth, in order to increase the growth rate, the cell has to increase the synthesis flux of metabolic precursors, by increasing the enzyme concentration involved in the metabolic network. In parallel, the capability of the protein production has also to increase to fulfill the increase of the enzyme concentration. So the synthesis flux of proteins has to be sufficient to satisfy the increase of both the enzyme and the ribosome concentration. This trade-off about the fate of proteins implies the existence of a bottleneck between the protein production and the metabolic network. In the next subsections, we will mathematically formalised this trade-off during the exponential phase.

Let us first introduce the notations used in the sequel and on Figure 1 describing all the cell components. Proteins can be assigned to three main classes of biological processes (metabolic network, translation apparatus and other proteins), so let us first introduce the notations for these sets.

The metabolic network is composed of (i) N_m enzymes $\mathcal{E}_i \triangleq (\mathcal{E}_1, \dots, \mathcal{E}_m)$ leading to a flux vector $\mathbf{v} \triangleq (v_1, \dots, v_m)$ of size N_m ; (ii) N_i internal metabolites $X_i \triangleq (X_{i_1}, \dots, X_{i_{N_i}})$; (iii) N_p metabolic precursors $X_p \triangleq (X_{p_1}, \dots, X_{p_{N_p}})$ consumed during the synthesis of proteins; (iv) N_r recycled metabolites $X_r \triangleq (X_{r_1}, \dots, X_{r_{N_r}})$ produced during the synthesis of proteins. The concentration of the i -th enzyme in mol/l is noted E_i and we assume that $E_i(t)$ is linked to the instantaneous flux v_i in mol/l/h by the following relation $v_i(t) = \pm k_E E_i(t)$, where $k_E > 0$ corresponds to the enzyme efficiency at a given temperature.

The concentration of ribosomes is noted R_a . The set of proteins of size N_G belonging neither to the metabolic network nor to the translation apparatus, is referred as \mathcal{P}_G from here on. We introduced a mean concentration P_G for the proteins in \mathcal{P}_{G_i} such that

$$P_G \triangleq \frac{P_{G_i}}{n_{P_{G_i}}} \text{ for any } i \in \{1, \dots, N_G\}$$

where P_{G_i} is the concentration of protein \mathcal{P}_{G_i} in \mathcal{P}_G and $n_{P_{G_i}}$ is the number of protein \mathcal{P}_{G_i} inside the cell. Finally the cell is composed of N_c macro-components $X_c \triangleq (X_{c_1}, \dots, X_{c_{N_c}})$ such as the cell wall or the lipid membrane, whose intracellular concentration is independent of the growth rate.

Let us now discuss the nature of the constraints at which the system displayed in Figure 1 is submitted.

A. Impact of the volume variation

During exponential growth phase, the volume of the bacteria population is increasing exponentially:

$$V(t) = V_0 e^{\mu(t-t_0)}, \quad \frac{dV(t)}{dt} = \mu V(t)$$

where V_0 corresponds to the initial volume of all bacteria at $t = t_0$ and μ corresponds to the growth rate. The concentration of a protein \mathcal{P} (in mol/l) corresponds by definition to $P(t) = \frac{p(t)}{V(t)}$ where $p(t)$ is the number of the cell component \mathcal{P} in moles. The variation of the component concentration is given by

$$\frac{dP(t)}{dt} = \underbrace{\frac{dp(t)}{dt} \frac{1}{V(t)}}_{\text{Production} \triangleq p(t)} - \underbrace{\frac{dV(t)}{dt} \frac{p(t)}{V^2(t)}}_{\text{Dilution effect}} = p(t) - \mu P(t)$$

During the exponential growth (in steady-state regime), $P(t)$ is constant. So $p(t) = \mu P(t)$ in order to maintain constant the concentration $P(t)$ inside the cell despite the volume variation due to the cell growth.

Hence if α_p moles of a specific metabolite X_k are consumed during the synthesis of the protein \mathcal{P} , $\mu \alpha_p P(t)$ mol/l/h are consumed to maintain the concentration $P(t)$ constant at the growth rate μ . If the protein \mathcal{P} is an enzyme, the flux v_k of metabolite X_{p_k} required to maintain the concentration $E(t)$ constant in steady-state corresponds to

$$v_k(t) = \mu \alpha_E E(t) = \mu \alpha_E \frac{|v_E(t)|}{k_E} \quad (1)$$

Based on these definitions, three design constraints allowing the bacterium to duplicate itself can be identified.

B. Three design constraints

The different subsets of proteins (metabolic, ribosomal and \mathcal{P}_G) has to respect different structural constraints to ensure their coordination at the growth rate $\mu > 0$.

(C₁), the “**Metabolic capability constraint**”: the metabolic network capability has to be sufficient

(a) to produce all metabolic precursors required for cell growth, including those consumed during the molecular assembling and translation processes. Basically, the synthesis flux of the N_p metabolic precursors by the metabolic network has to be more important than the one consumed during the synthesis of the cell components. (C_{1a}): for all $i \in \{1, \dots, N_p\}$,

$$-\sum_{j=1}^m S_{p_{ij}} v_j + \mu \left(\sum_{j=1}^m C_{M_{ij}}^{M_p} |v_j| + C_{R_i}^{M_p} R_a + C_{G_i}^{M_p} P_G \right) + v_Y \leq 0$$

where S_p is the sub-part of the stoichiometric matrix S for X_p and v_Y corresponds to free exchange fluxes with the environment such as the diffusion of metabolites through the membrane. $C_{M_{ij}}^{M_p}$, $C_{R_i}^{M_p}$ and $C_{G_i}^{M_p}$ are nonnegative and respectively correspond to the number of the X_{p_i} metabolite required for the synthesis of one ribosome, one protein in \mathcal{P}_G and the j -th protein involved in the metabolic network (referred as α_E in equation 1).

(b) to maintain the concentration of the set of macro-components X_c constant to $\bar{X}_c \triangleq (\bar{X}_{c_1}, \dots, \bar{X}_{c_{N_c}})$, where \bar{X}_{c_i} is the target concentration of X_{c_i} . (C_{1b}): for all $i \in \{1, \dots, N_c\}$,

$$-\sum_{j=1}^m S_{c_{ij}} v_j + \mu \bar{X}_c \leq 0$$

where S_c is the sub-part of S for X_c .

(c) to absorb all recycled metabolites produced during the synthesis of cell components. (C_{1c}): for all $i \in \{1, \dots, N_r\}$,

$$\sum_{j=1}^m S_{r_{ij}} v_j + \mu \left(\sum_{j=1}^m C_{M_{ij}}^{M_r} |v_j| + C_{R_i}^{M_r} R_a + C_{G_i}^{M_r} P_G \right) \leq 0$$

where S_r is the sub-part of S for X_r . $C_{R_i}^{M_r}$, $C_{G_i}^{M_r}$ and $C_{M_{ij}}^{M_r}$ are also nonnegative and respectively correspond to the number of the X_{r_i} metabolite produced during the synthesis of one ribosome, one protein in \mathcal{P}_G and the j -th protein involved in the metabolic network.

(d) Moreover, the metabolic network has also to satisfy the mass conservation law. (C_{1d}): for all $i \in \{1, \dots, N_i\}$,

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

where S_I is the sub-part of S for X_i .

(C_2), the “Resource management constraint”: the translation apparatus capability has to be sufficient to ensure the concentration maintenance of all the cell proteins at the growth rate μ .

$$\mu \left(\sum_{j=1}^m C_{M_j}^R |v_j| + C_R^R R_a + C_G^R P_G \right) - k_T R_a \leq 0$$

where k_T is the translation efficiency (around 12 to 20 amino acids per second at 37°C [5]). C_R^R , C_G^R and $C_{M_j}^R$ are positive and respectively correspond to the total number of amino acid residues per ribosome, per protein in \mathcal{P}_G and

per protein involved in the metabolic network.

(C_3), the “Density constraint”: The cell has also to manage its intracellular density to ensure the suitable diffusion of all cell components (proteins, metabolites, DNA, etc) inside the cell [7], [8].

$$\sum_{j=1}^m C_{M_j}^D |v_j| + C_R^D R_a + C_G^D P_G - \bar{D} \leq 0$$

where \bar{D} is the mean density of the cell components (usually in g/ml but can be converted in mol/l). C_R^D , C_G^D and $C_{M_j}^D$ are equal to C_R^R , C_G^R and $C_{M_j}^R$.

C. A non smooth optimization problem

The satisfaction of these previous constraints leads to the following feasibility problem $P_f(\mu)$. For fixed $P_G \geq 0$, $\mu \geq 0$,

$$\begin{aligned} &\text{find} && R_a \geq 0, v \in \mathcal{R}^m \\ &\text{subject to} && \end{aligned}$$

$$(C_{1a}), (C_{1b}), (C_{1c}), (C_{1d}), (C_2), (C_3).$$

and let us define the associated set of feasible solutions

$$\mathcal{C}_{\mu, P_G} = \{(R_a, v) \in \mathcal{R}^+ \times \mathcal{R}^m \mid (C_{1a}), (C_{1b}), (C_{1c}), (C_2), (C_3)\}.$$

$P_f(\mu)$ is a nonsmooth optimization problem due to the presence of the absolute value in the different constraints. However, some interesting properties with respect to μ can be underlined and interpreted from a biological point of view.

III. GENERAL PROPERTIES OF $P_f(\mu)$

Let us first define the following sets: $I_m \triangleq \{1, \dots, m\}$, $I_p \triangleq \{1, \dots, N_p\}$, $I_r \triangleq \{1, \dots, N_r\}$, $I_i \triangleq \{1, \dots, N_i\}$ and $I_c \triangleq \{1, \dots, N_c\}$.

Lemma 3.1: If for $\mu > 0$ and for $P_G \geq 0$ $P_f(\mu)$ is feasible then any $(\bar{R}_a, \bar{v}) \in \mathcal{C}_{\mu, P_G}$ is such that $\bar{R}_a > 0$ and $\bar{v} \neq 0$, i.e., there exists a nonempty subset U of I_m , such that for all $j \in U$, $\bar{v}_j \neq 0$, and for $k \in I_m/U$, $\bar{v}_k = 0$.

Lemma 3.1 points out that if $P_f(\mu)$ is feasible then any feasible solution is non null. Practically, we prove here that both the concentration of ribosomes and the concentration of a subset of enzymes/transporters have to be non null to allow the growth of the cell. Despite the biological obviousness of this result, it strongly emphasizes the validity of the formulation of the cell design.

Let us now investigate the properties of $P_f(\mu)$ with respect to μ .

A. Properties of the solution μ

Proposition 3.2: $P_f(\mu)$ has the following properties :

- For any $P_G \geq 0$ and for any $\mu \geq 0$, \mathcal{C}_{μ, P_G} is convex.
- If for $P_G \geq 0$ and for $\mu^+ > 0$, $P_f(\mu^+)$ is feasible then for any $\mu \in [0, \mu^+]$, $P_f(\mu)$ is also feasible and $\mathcal{C}_{\mu^+, P_G} \subseteq \mathcal{C}_{\mu, P_G}$.
- For any $P_G \geq 0$ there exists a finite $\mu^* \geq 0$ such that $P_f(\mu^*)$ is feasible and for all $\mu > \mu^*$, $P_f(\mu^*)$ is infeasible.

From a biological point of view, we show here that there exists a resource distribution between the metabolic network and the ribosomes for the growth rate value μ and for lower values (Item 2 of Proposition 3.2). Hence we predict that the bacterium can grow with the growth rate μ , and of course for lower growth rate values. Moreover, there exists a maximal value for the growth rate (μ^*) with respect to a specific medium (Item 3 of Proposition 3.2).

B. Parameter variation

Proposition 3.3: If for fixed $P_G > 0$ and $\mu > 0$, $\mathcal{C}_{\mu, P_G} \neq \emptyset$ then for all $\delta P_G > 0$ such that $\delta P_G \leq P_G$ there exists $\delta \mu > 0$ such that $\mathcal{C}_{\mu + \delta \mu, P_G - \delta P_G} \neq \emptyset$.

Proposition 3.3 indicates that the growth rate is increasing when the set of proteins \mathcal{P}_G is decreasing. Proteins involved in \mathcal{P}_G depend on the physiological state of the bacterium and is directly linked to its adaptation to the ecological niche. Biological experiments available in the literature confirm the impact of \mathcal{P}_G proteins on the growth rate. For *B. subtilis*, the synthesis of proteins involved in the mobility (present in our \mathcal{P}_G set) is active in exponential phase. If the inductor of the mobility is deleted, the mutant strain grows faster than the wild type [9]. Moreover, for low growth rates, the weight of \mathcal{P}_G is increasing compared to fast growth rates. Indeed, the strategies developed by the bacteria could be complex during low growth rates or during the transition between exponential to stationary phase leading to change the \mathcal{P}_G set. High level decisions such as the general stress response, competence, initiation of sporulation could be induced which could impact the P_G value.

C. A Linear Programming feasibility problem

$P_f(\mu)$ then corresponds to a nondifferentiable convex feasibility problem, for which no efficient algorithms currently exist for their resolution [10]. However, we show in the sequel that $P_f(\mu)$ is equivalent to a LP feasibility problem for which many polynomial-time algorithms based on the interior point method are available [11], [10], [12].

Let us introduce the following LP feasibility problem $P_f^{lp}(\mu)$:

$$\begin{aligned}
& \text{find } R_a \geq 0, v \in \mathcal{R}^m, v^{\max} \in \mathcal{R}_+^m \\
& \text{subject to} \\
& (C_{1a}^{lp}) \quad \text{for all } i \in I_p, \\
& \quad -\sum_{j=1}^m S_{p_{ij}} v_j + \dots \\
& \quad \mu \left(\sum_{j=1}^m C_{M_{ij}}^{M_p} v_j^{\max} + C_{R_i}^{M_p} R_a + C_{G_i}^{M_p} P_G + \bar{X}_c \right) + v_Y \leq 0 \\
& (C_{1b}^{lp}) \quad \text{for all } i \in I_c, \\
& \quad -\sum_{j=1}^m S_{c_{ij}} v_j + \mu \bar{X}_c \leq 0 \\
& (C_{1c}^{lp}) \quad \text{for all } i \in I_r, \\
& \quad \sum_{j=1}^m S_{r_{ij}} v_j + \mu \left(\sum_{j=1}^m C_{M_{ij}}^{M_r} v_j^{\max} + C_{R_i}^{M_r} R_a + C_{G_i}^{M_r} P_G \right) \leq 0 \\
& (C_{1d}^{lp}) \quad \text{for all } i \in I_l, \\
& \quad \sum_{j=1}^m S_{l_{ij}} v_j = 0 \\
& (C_2^{lp}) \quad \mu \left(\sum_{j=1}^m C_{M_j}^R v_j^{\max} + C_R^R R_a + C_G^R P_G \right) - k_T R_a \leq 0 \\
& (C_3^{lp}) \quad \sum_{j=1}^m C_{M_j}^D v_j^{\max} + C_R^D R_a + C_G^D P_G - \bar{D} \leq 0 \\
& (C_4^{lp}) \quad \text{for all } j \in I_m, \\
& \quad v_j - v_j^{\max} \leq 0 \text{ and } (v_j + v_j^{\max}) \leq 0
\end{aligned}$$

and the associated set of these inequalities and equalities:

$$\mathcal{C}_{\mu, P_G}^{lp} = \{(R_a, v) \in \mathcal{R}^+ \times \mathcal{R}^m \mid (C_{1a}^{lp}), (C_{1b}^{lp}), (C_{1c}^{lp}), \dots, (C_{1d}^{lp}), (C_2^{lp}), (C_3^{lp})\}.$$

Proposition 3.4: For fixed $P_G \geq 0$, $\mu \geq 0$, $\mathcal{C}_{\mu, P_G} = \mathcal{C}_{\mu, P_G}^{lp}$.

Proof: (if) Let us first prove that $\mathcal{C}_{\mu, P_G}^{lp} \subseteq \mathcal{C}_{\mu, P_G}$. Let us assume that for fixed $\mu \geq 0$, $P_G \geq 0$, $\mathcal{C}_{\mu, P_G}^{lp} \neq \emptyset$. Let $(\bar{R}_a, \bar{v}) \in \mathcal{C}_{\mu, P_G}^{lp}$. Then there exists a \bar{v}^{\max} such that (C_{1a}^{lp}) , (C_{1b}^{lp}) , (C_{1c}^{lp}) , (C_{1d}^{lp}) , (C_2^{lp}) , (C_3^{lp}) and (C_4^{lp}) are satisfied. (C_4^{lp}) imply that for all $j \in I_m$, $|\bar{v}_j| \leq \bar{v}_j^{\max}$. Since for (C_{1a}^{lp}) , (C_{1c}^{lp}) , (C_2^{lp}) and (C_3^{lp}) , (i) the coefficients multiplying \bar{v}_j^{\max} are nonnegative, and (ii) \bar{R}_a , \bar{v} , \bar{v}^{\max} satisfy them, \bar{R}_a , \bar{v} satisfy (C_{1a}) , (C_{1c}) , (C_2) and (C_3) . (C_{1b}) and (C_{1d}) are obviously satisfied for \bar{v} . So $(\bar{R}_a, \bar{v}) \in \mathcal{C}_{\mu, P_G}$ and $\mathcal{C}_{\mu, P_G}^{lp} \subseteq \mathcal{C}_{\mu, P_G}$.

(only if) Let us now prove that $\mathcal{C}_{\mu, P_G} \subseteq \mathcal{C}_{\mu, P_G}^{lp}$. Let us assume that for fixed $\mu \geq 0$, $P_G \geq 0$, $\mathcal{C}_{\mu, P_G} \neq \emptyset$. Let $(\bar{R}_a, \bar{v}) \in \mathcal{C}_{\mu, P_G}$ and let us introduce for each $j \in I_m$, $v_j^{\max} \geq 0$ such that $v_j^{\max} = |v_j|$. (C_4^{lp}) are obviously satisfied. The other constraints defining $\mathcal{C}_{\mu, P_G}^{lp}$ are obtained by the direct substitution of $|v_j|$ by v_j^{\max} . So $(\bar{R}_a, \bar{v}) \in \mathcal{C}_{\mu, P_G}^{lp}$ which concludes the proof. ■

Proposition 3.4 emphasizes that $P_f(\mu)$ is equivalent to the LP feasibility problem $P_f^{lp}(\mu)$ [11], [10], [12]. Since Proposition 3.2 obtained for $P_f(\mu)$ can be extended to $P_f^{lp}(\mu)$, we deduce that for a set of external resources, μ^* can be computed by dichotomy through an iterative resolution of $P_f^{lp}(\mu)$ for each μ value. We also obtain through the resolution of $P_f^{lp}(\mu)$ the corresponding flux distribution, the concentration of proteins involved in the metabolic network and of ribosomes for μ^* .

IV. PREDICTION OF THE MODULAR STRUCTURE OF THE METABOLIC NETWORK

The feasibility problem $P_f(\mu)$ allows to manage the priority of external resource uptakes according to the cost of their assimilation pathway or their *de novo* synthesis pathway respectively. A metabolic pathway is indeed composed of several enzymes, each one having distinct characteristics (amino acid composition, length, etc). Several metabolic pathways can lead to the production of the same metabolite. Hence, choosing between two metabolic pathways can be crucial for the cell if the growth rate is impacted by this choice. We proved in [6] that the constraints integrated in $P_f(\mu)$ lead to turn off the more expensive metabolic pathway.

Let us consider two alternative metabolic pathways mp_1 and mp_2 to produce the k -th metabolic precursor X_{pk} , for $k \in I_p$. Each pathway is respectively composed of N_{mp_1} and N_{mp_2} distinct enzymes and such that no other co-metabolites are solely produced or consumed with the exception of X_{pk} .

Let $I_{mp_1} \subseteq I_m$ and $I_{mp_2} \subseteq I_m$ be the index of fluxes associated to the enzymes belonging to these pathways.

Assumption 4.1: For the two metabolic pathways mp_1 and mp_2 previously introduced:

- (i) The number of the each metabolic precursor required for the synthesis of the pathway mp_2 is lower than the one required for the synthesis of mp_1 , which corresponds to for all $i \in I_p$, $\sum_{j \in I_{mp_2}} C_{M_{ij}}^{M_p} < \sum_{j \in I_{mp_1}} C_{M_{ij}}^{M_p}$,
- (ii) The number of the each recycled metabolite produced during the synthesis of the pathway mp_2 is lower than the one produced during the synthesis of mp_1 , which corresponds to for all $i \in I_r$, $\sum_{j \in I_{mp_2}} C_{M_{ij}}^{M_r} < \sum_{j \in I_{mp_1}} C_{M_{ij}}^{M_r}$,
- (iii) The number of ribosomes required for the synthesis of all proteins belonging to mp_2 is lower than the one required for the synthesis of all proteins belonging to mp_1 , which corresponds to $\sum_{j \in I_{mp_2}} C_{M_j}^R < \sum_{j \in I_{mp_1}} C_{M_j}^R$,
- (iv) The intracellular space occupied by all proteins belonging to mp_2 is lower than the one occupied by all proteins belonging to mp_1 , which corresponds to $\sum_{j \in I_{mp_2}} C_{M_j}^D < \sum_{j \in I_{mp_1}} C_{M_j}^D$.

Proposition 4.2: Let Assumption 4.1 be hold. For all $P_G \geq 0$ and all $\mu \geq 0$ if $P_f(\mu)$ is feasible and (\bar{R}_a, \bar{v}) is such that $\bar{v}_j \neq 0$ for $j \in I_{mp_1} \cup I_{mp_2}$ then there exists $\delta\mu > 0$ such that $P_f(\mu + \delta\mu)$ is feasible too.

Proposition 4.2 indicates that choosing a “cheap” synthesis pathway in terms of metabolic precursors instead of an “expensive” one allows to increase the growth rate. For example, solving the optimization problem leads to activate an amino-acid transporter instead of inducing the entire *de novo* pathway when this amino acid is present in the medium. However, the choice between two metabolic pathways is usually much more difficult to evaluate analytically since most of metabolic pathways also include cofactors or co-metabolites (contrary to the chosen example). In particular, some resource distribution solution of $P_f(\mu)$ could induce both the “cheap” and the “expensive” pathway if the production cost of one co-metabolite produced by the expensive pathway is cheaper than the produced by an alternative pathway. The global flux distribution of the metabolic network obtained during the resolution of $P_f(\mu)$ is thus strongly dependent both on the stoichiometry of the metabolic network, and on the cost in metabolic precursors of the induction of of the whole metabolic network. To conclude, we predict that the bacterium can develop strategies such as genetic regulations to modulate the flux of the expensive metabolic pathway.

V. VALIDATION FOR B. SUBTILIS

A complete quantitative validation would require the identification of all k_p parameters for the enzymes, which seems unreasonable due to the lack of available data. However, the choice of the included design constraints can be validated by comparison with the existing knowledge in the literature.

In particular, (i) the predictions of the regulation structure of the metabolic network can be compared with the known regulatory network for one specific organism; (ii) the predicted resource repartition between enzymes and ribosomes can be compared with the known distributions for model-organisms such as *E. coli* [13] or *S. cerevisiae* [14].

We considered the main metabolic pathways of *Bacillus subtilis* to build S : the central carbon pathway with the glucose assimilation, aerobic respiration, amino-acids metabolism, the synthesis of nucleotides, fatty-acids, phospholipids, peptidoglycan and teichoic acids [15], [1]. X_c corresponds to the concentrations for phospholipids, peptidoglycan and teichoic acids during exponential growth. These metabolic pathways include 301 genes, coding for 250 enzymes, 31 transporters, which represents 325 reactions. We considered that a ribosome is composed of one rRNA of 4593 nucleotides and 52 ribosomal proteins, with $k_T = 15aa/s$. For each protein, all coefficients C_i^j of $P_f(\mu)$ are deduced from the exact amino acid length and the mean composition in amino acids proposed in [16]. The exact composition in rRNA nucleotides is used to compute the coefficient C_R^M . We also used (i) for \bar{X}_c , the mean concentrations available in [15]; (ii) the same turnover $k_P = 50s^{-1}$ for all enzymes; (iii) $\bar{D} = 1.117g/ml$ [17]; (iv) P_G is used as a scaling parameter, and is set to around 45% of all cell proteins.

A. Recovery of the known functional modules

We solved $P_f^{lp}(\mu)$ for a set of various media. We obtained the groups of enzymes (modules) that are switched on/off according to the media composition, and thus that could be controlled by a common regulator. We compared our predictions with the results in [1]. All the known modules have been recovered except for one, for which the known genetic regulation is quite unclear. Moreover, we also predicted the existence of 11 additional modules in the metabolism of amino acids. Among them, 6 can be found for the gram-negative model bacterium *E. coli*.

B. Predictions of the resource repartition

We displayed on Figure 2 the predictions of the number of amino acid residues used for ribosomes and for the metabolic network obtained for various media (and so different growth rate). The ribosome concentration is increasing with the growth rate, while the protein concentration involved in the metabolic network is decreasing. We obtained the same qualitative behavior as the resource repartition in *E. coli* [13] and in *S. cerevisiae* [14], for which three set of genes can be distinguished. The expression of two is growth-rate dependent (induction or repression) while the third one is growth-rate independent and corresponds to our \mathcal{P}_G set [14]. Following our results and [13], [14], for a given set of environmental conditions, every protein saved through the repression of a metabolic pathway for example, reduces the amount of metabolic precursors allocated to the metabolic network. This saved set of metabolic precursors can be shared between the three sets of proteins (\mathcal{P}_G , metabolic and ribosomal) in order to increase the concentration of

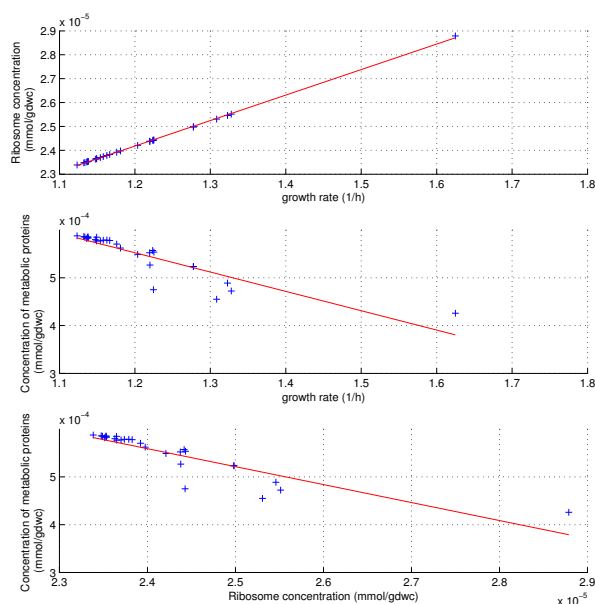


Fig. 2. Resource repartition between ribosomes and metabolic enzymes

ribosomes, and thus to increase the growth rate. Reciprocally, every metabolic precursor unused to increase the ribosome concentration leads to decrease the growth rate.

Hence, the cell can develop strategies such as the genetic regulations to turn on/off the synthesis of proteins and entire metabolic pathways when they are dispensable and thus to save the corresponding metabolic precursors. We proved in this paper through the feasibility problem $P_f(\mu)$ the biological fact usually observed: the genetic regulations appear to save proteins.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we demonstrated that the problem of resource management in bacteria for a fixed growth rate can be formalised into a nondifferentiable convex constraint-based feasibility problem $P_f(\mu)$ through the integration of three structural constraints. This feasibility problem can be easily transformed into an equivalent LP feasibility problem $P_f^{LP}(\mu)$, for which many classical polynomial-time solvers are available [10], [18]. The resolution of the LP feasibility problem leads to predict not only the flux distribution and the maximal growth rate, but also the concentrations of ribosomes, and of the proteins involved in the metabolic network and thus the composition of the cell for different growth rates. Moreover, the modular structure of the metabolic network can also be predicted with respect to the medium composition.

Another major conclusion of this paper is the successful use of tools and methods based on convex optimisation in biology. The formalisation of the cell behavior is suitable for convex optimisation and strong structural properties have been obtained allowing to explain the emergence of

functional modules in the metabolic network. The links between these two fields (biology and optimization) have to be strengthened in order to investigate fundamental questions such as the evolution of regulatory networks of organisms with respect to the ecological niche.

VII. ACKNOWLEDGMENTS

This work was supported in part by ANR Dynamocell (NT05-2_44860) and in part by the BaSysBio project (LSHG-CT-2006-037469).

REFERENCES

- [1] A. Goelzer, F. Bekkal Brikci, I. Martin-Verstraete, P. Noirot, P. Bessières, S. Aymerich, and V. Fromion, "Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*," *BMC Syst Biol*, vol. 2, p. 20, February 2008.
- [2] A. Varma and B. Palsson, "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* w3110," *Appl Environ Microbiol*, vol. 60, no. 10, pp. 3724–3731, October 1994.
- [3] J. Edwards and B. Palsson, "The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities," *Proc Natl Acad Sci USA*, vol. 97, no. 10, pp. 5528–5533, May 2000.
- [4] B. Papp, C. Pal, and L. Hurst, "Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast," *Nature*, vol. 429, no. 6945, pp. 661–664, June 2004.
- [5] H. Bremer and P. Dennis, "Modulation of chemical composition and other parameters of the cell by growth rate," in *Escherichia coli and salmonella: cellular and molecular biology*, 2nd ed., F. Neidhart, Ed. Washington D.C., USA: American Society of Microbiology Press, 1996, pp. 1553–1569.
- [6] A. Goelzer, V. Fromion, and G. Scorletti, *Cell design in bacteria as a convex optimization problem*, 2009.
- [7] Q. Beg, A. Vazquez, J. Ernst, M. Demenezes, Z. Bar-Joseph, A. Barabasi, and Z. Oltvai, "Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity," *Proc Natl Acad Sci USA*, vol. 104, no. 31, pp. 12 663–12 668, July 2007.
- [8] A. Vazquez, Q. Beg, M. Demenezes, J. Ernst, Z. Bar-Joseph, A. Barabasi, L. Boros, and Z. Oltvai, "Impact of the solvent capacity constraint on *E. coli* metabolism," *BMC Syst Biol*, vol. 2, p. 7, January 2008.
- [9] E. Fischer and U. Sauer, "Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism," *Nat Genet*, vol. 37, no. 6, pp. 636–640, 2005.
- [10] Y. Nesterov, *Introductory lectures on convex optimization: a basic course*. Kluwer Academic Publishers, 2004.
- [11] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. MPS/SIAM Series on Optimization, 2001.
- [12] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [13] A. Marr, "Growth rate of *Escherichia coli*," *Microbiol Rev*, vol. 55, no. 2, pp. 316–333, June 1991.
- [14] M. Brauer, C. Huttenhower, E. Airoldi, R. Rosenstein, J. Matese, D. Gresham, V. Boer, O. Troyanskaya, and D. Botstein, "Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast," *Mol Biol Cell*, vol. 19, no. 1, pp. 352–367, January 2008.
- [15] Y. Oh, B. Palsson, S. Park, C. Schilling, and R. Mahadevan, "Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data," *J Biol Chem*, vol. 282, no. 39, pp. 28 791–28 799, September 2007.
- [16] R. Alves and M. Savageau, "Evidence of selection for low cognate amino acid bias in amino acid biosynthetic enzymes," *Mol Microbiol*, vol. 56, no. 4, pp. 1017–1034, May 2005.
- [17] A. Hart and C. Edwards, "Buoyant density fluctuations during the cell cycle of *Bacillus subtilis*," *Arch Microbiol*, vol. 147, no. 1, pp. 68–72, February 1987.
- [18] J.-C. Gilbert, C. Lemaréchal, and C. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects*, 2nd ed. Springer-Verlag, 2006.