Automatica 47 (2011) 1210–1218

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Cell design in bacteria as a convex optimization problem☆

Anne Goelzer [a,b], Vincent Fromion [a,*], Gérard Scorletti [b]

[a] Institut National de la Recherche en Agronomie, Unité de Mathématique, Informatique et Génome UR1077, F-78350 Jouy-en-Josas, France
[b] Laboratoire Ampère, École Centrale de Lyon - BP163 - 69131 Ecully, France

ARTICLE INFO

Article history:
Received 1 February 2010
Received in revised form
27 January 2011
Accepted 31 January 2011
Available online 21 March 2011

Keywords:
FBA
Cell design
Systems biology
Convex optimization

ABSTRACT

In this paper, we investigate the prediction of the cell composition of bacteria with respect to their medium. By modeling the bacterium as an interconnection of subsystems, the problem is written as a non-smooth convex optimization problem equivalent to a Linear Programming feasibility problem. We then obtain a new method, called Resource Balance Analysis (RBA), predicting the distribution of the available resources in the medium among the various cellular subsystems. Beyond its predictive capability, the proposed approach grasps some fundamental aspects of the bacterium physiology by including a refined model. This method reveals the existence of an intrinsic bottleneck in the system resource distribution of the bacterium, leading to the existence of a structural limitation of its growth rate which can be predicted. RBA is also able to predict the configuration of the metabolic network for a given medium at steady-state regimen which nicely fits the available experimental results for the gram-positive model bacterium *Bacillus subtilis*.

1. Introduction

Superior organisms such as mammals are composed of organs performing well-defined tasks: they can be considered as highly organized systems composed of specialized subsystems. During the second part of the 20th century, theoretical and technological advances, especially in the field of molecular biology, revealed that the modularity observed at the scale of an organism also exists at the scale of the cell. Like superior organisms, cells are organized systems, composed of a large number of elementary and specialized processes interacting together. This systemic view of the cell has recently renewed the interest of System Theory to address the cell complexity. This point of view motivated the emergence of the so-called *Systems Biology* research field (see Kitano, 2001 and references therein).

In a previous report (Goelzer et al., 2008), our group proposed a model of the metabolic pathway regulation network of the gram-positive bacterium *Bacillus subtilis* and described the properties of this network using a mathematical analysis. The metabolic network of *B. subtilis* was broken down into elementary functional modules locally controlled by a genetic regulation sensing the internal state of each module. Our analysis revealed that these modules are further coordinated by the so-called global genetic regulations in response to physiological changes, suggesting a strong modular and hierarchical organization of the metabolic network regulation. The main objective of the work reported here is to investigate the existence of possible design rules/constraints explaining this modular organization.

Some characteristics of the biological evolution processes can be interpreted from the perspective of Automatic Control. Biological systems, such as bacteria, animals or plants are the result of a long evolution where two main processes played complementary roles: the genetic variations (*e.g.* due to mutations or recombinations) and the natural selection. From the perspective of control, the evolution process is similar to an optimization process where natural selection defines the "criterion" and genetic variations allow to explore a large set of possible solutions. Additionally, natural selection and genetic variations still continue to constrain the design of biological systems. As a result, the optimality principle and optimization tools have been widely applied to identify some design principles in Systems Biology (Dekel & Alon, 2005; Ehrenberg & Kurland, 1984; Zaslaver et al., 2004). An emerging optimal principle proposed for the design of metabolic networks of microorganisms is the maximization of the growth with respect to a given extracellular medium. This aspect is indeed crucial in the context of the competition between bacteria. This principle is the starting point of the flux balance analysis (FBA) method based on the following Linear Programming (LP) problem (Varma & Palsson, 1994):

$$\begin{array}{ll} \text{maximize} & c^T v \\ v \in \mathbb{R}^m & \\ \text{subject to} & S.v = 0 \\ & \alpha_i \leq v_i \leq \beta_i, \quad i \in \{1, \ldots, m\} \end{array}$$

☆ The material in this paper was partially presented at the 48th IEEE Conference on Decision and Control, December 16–18, 2009, Shanghai, China. This paper was recommended for publication in revised form by Associate Editor Murat Arcak under the direction of Guest Editor Francis J. Doyle III.

* Corresponding author. Tel.: +33 1 34 65 28 81; fax: +33 1 34 65 29 01.
E-mail addresses: anne.goelzer@jouy.inra.fr (A. Goelzer), vincent.fromion@jouy.inra.fr (V. Fromion), gerard.scorletti@ec-lyon.fr (G. Scorletti).

0005-1098/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2011.02.038

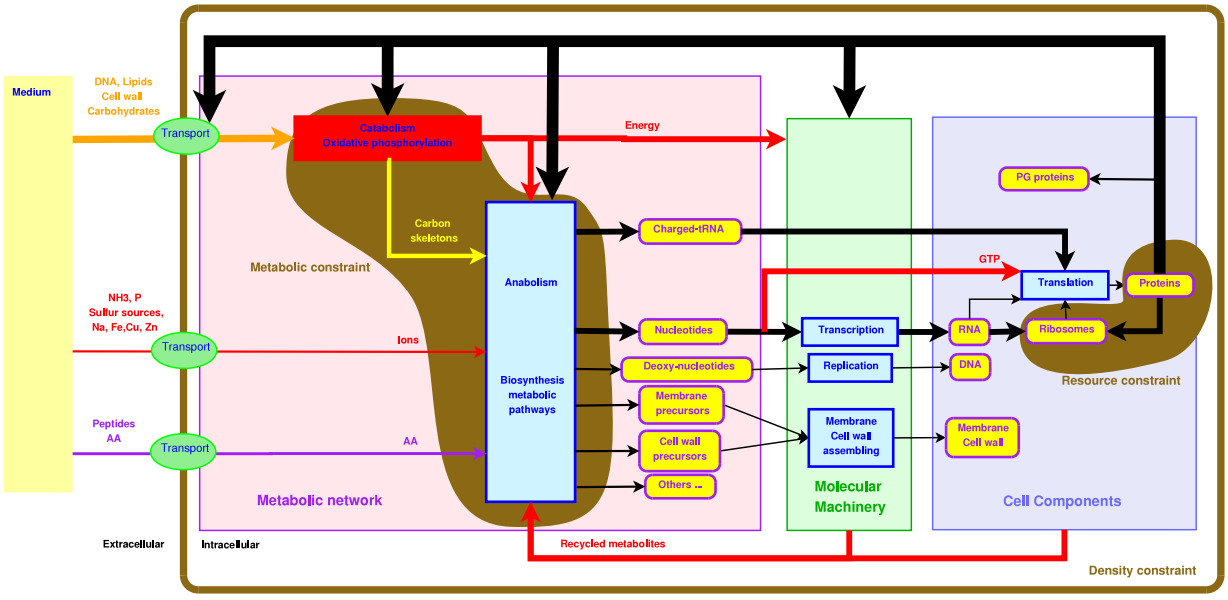

Fig. 1. A systemic view of the cell.

where the metabolic network is defined by the stoichiometry matrix S. At steady-state, the metabolic fluxes ν are solutions of the equation $S.\nu = 0$. The cost function $c^T\nu$ is the rate of biomass production, also referred to as the growth rate. The vector c is usually obtained as the mean composition of the cell at a given growth rate (Varma & Palsson, 1994), which is a strong assumption since an important phenomenon is the dependence of the cell composition with the growth rate (Bremer & Dennis, 1996). Thus, although the FBA method was experimentally validated on several organisms (Edwards & Palsson, 2000; Papp et al., 2004), there exist theoretical drawbacks due to the simplification of the cell design problem. Moreover, the FBA method requires the definition of suitable upper bounds β_i on some fluxes in order to obtain a bounded growth rate.

In this paper, we propose a new method whose range of prediction is much larger than the FBA one by considering a more refined model of the cell. We explicitly predict not only the metabolic fluxes but also the protein concentration and thus the variation of the cell composition with the growth rate. Another benefit is to propose a framework which addresses some fundamental open biological problems such as the structural growth rate limitation (Koch, 1997). To this purpose, in contrast with the FBA method, the cell is modeled as an interconnection of subsystems whose function is to perform some well-defined tasks to achieve the growth by consuming common resources. A given growth rate then strongly depends on the resource allocation among all these subsystems. Actually, due to the interconnected subsystems modeling, the proposed method predicts this resource allocation. This problem of resource allocation is captured by a non-smooth convex optimization problem, which is transformed into a Linear Programming optimization problem. A strong interest is that LP optimization problems can be efficiently solved even for large-scale problems.

Interesting properties of the solution are then investigated and discussed from a biological point of view. In particular, we can predict different configurations of the metabolic pathways for various media and then reveal which metabolic pathways with respect to the resource allocation of proteins should be turned on or turned off. It turns out that these predictions coincide with the configurations of the metabolic network resulting from the genetic regulation network of one specific organism, *B. subtilis*.

The paper is organized as follows. In Section 2, the cell design constraints are discussed and formalized as an optimization problem. Section 3 presents the properties that can be deduced from the analysis of the optimization problem. Section 4 details the LP problem. Section 5 investigates the different ways to increase the growth rate with respect to the analysis of our optimization problem. Finally, Section 6 reports an application of our approach for the gram-positive model bacterium *Bacillus subtilis*.

Notations: For a given integer N_x, I_x denotes the set $\{1, \ldots, N_x\}$. For a given vector $P_x \triangleq (P_{x_1}, \ldots, P_{x_{N_x}})$, $P_x \geq 0$ (respectively $P_x > 0$) denotes that $\forall i \in I_x$, $P_{x_i} \geq 0$ (resp. $P_{x_i} > 0$). The symbol $\triangleq$ means "equal by definition".

2. Cell design constraints

2.1. A simplified and systemic description of bacteria

A living bacterium is an integrated system, which can be decomposed into subsystems. Each subsystem performs an elementary task required for the survival and the replication of the bacterium. A simplified and systemic representation of the bacterium is presented in Fig. 1, where only the key subsystems involved in the growth phase are considered: the metabolic network, the translation apparatus, the DNA replication, the RNA transcription and the processes of the cell membrane and cell wall assembly. Two subsystems are important in the context of this paper:

- **the translation apparatus**: its main function is to produce proteins. It is based upon a complex molecular machinery made of proteins and RNA, called the ribosome. Ribosomes translate coding messengers (mRNA) into linear chains of amino acids, which are folded into proteins. The concentration of ribosomes is denoted by R.
- **the metabolic network**: its main function is to uptake, to break down (catabolism phase) and to transform extracellular nutrients to produce energy and metabolic precursors (anabolism phase). These metabolic precursors are further consumed by all cellular subsystems (such as the assembly processes in the Molecular Machinery box in Fig. 1) to produce all sorts of cell components (proteins, cell wall, DNA, etc.) and side-reaction products that are also recycled by the metabolic network. Beyond its role in the cell anabolism and catabolism, the metabolic