Towards a realistic and integrated strain design in batch bioreactor

Guillaume Jeanne, Anne Goelzer, Sihem Tebbani, Didier Dumur, Vincent Fromion

Towards a realistic and integrated strain design in batch bioreactor

Guillaume Jeanne[1,2] Anne Goelzer[1] Sihem Tebbani[2] Didier Dumur[2] Vincent Fromion[1]

Abstract— **Recent advances in modeling bacterial cell functioning have deeply renewed questions about bioprocess design and opens the way towards the development of computer aided design (CAD) for strains. This article aims at explaining and exploring the consequences opened by the new cell models, investigating the questions related to the biological implementation of optimal strategies and in conclusion the possible role of the complex regulatory network in the retuning of the strain.**

I. INTRODUCTION

A bioprocess can be seen as a complex interaction between a specific medium and a micro-organism population. When successfully managed, the combination of the two can lead to an efficient production of compounds of interest. For decades, bioprocesses have been optimized through an iterative procedure combining two separate stages: a first stage where the micro-organism is selected or directly genetically modified through genome engineering tools, e.g. gene cloning; a second stage where the culture conditions and the composition of the medium are optimized in order to maximize the production of compounds of interest by the modified strain.

The optimization of culture conditions has been already investigated in the automatic control field, see e.g. [1], [2], in contrast to the problem of the strain design. The strain design is usually achieved without the help of mathematical models, by an intensive experimental trial-and-error cycle, i.e. a cycle of strain modifications followed by the experimental characterization the modified strain. However, the continuous progress in the understanding of the cell functioning along these last decades opened the way for approaches where mathematical models play a more important role in the strain design. For instance, the development of genome-scale metabolic models of micro-organisms supported the developments of metabolic engineering tools along the 90's and 2000's [3], and even for the strain design [4].

However, a few key ingredients are still missing in metabolic models, which prevent their use for computer-aided design (CAD) of strains. Actually, metabolic models integrate a mass balance constraint within the metabolic network, but do not capture the cellular constraints limiting growth rate structurally and governing the resource allocation between cellular processes [5]. Indeed, a cell is composed of cellular subsystems (the cellular processes) that share a common set of resources and that all together contribute to the growth rate. This results in the inability of metabolic models to anticipate the consequences of strain modification at the cell scale for rerouting the resources devoted to biomass production towards the production of compounds of interest. In conclusion even if useful methods for strain design already exist [4], they do not incorporate some constraints that are essential for predicting the cell behavior. Without integrating those constraints, it seems illusory to develop the next generation of computer-aided strain design.

Recently, several mathematical approaches were proposed in [5], [6], [7], [8] to integrate within a mathematical framework the constraints governing the resource allocation between cellular processes. Among these approaches, only the RBA (Resource Balance Analysis) approach introduced in [6], [9] led to formulate the problem in a way compatible to the development of CAD strain design, since it is at the genome scale and is formulated as a quasi-convex optimization problem. The RBA approach was biologically validated in 2015 in [10] for bacteria and predicts for a given medium in steady-state the maximal growth rate, the abundances of proteins and of molecular machineries, and the flux distribution at the cell scale. Moreover, it has been recently proved that the general principle within RBA, i.e. the parsimonious use of resources by the cell, may explain the emergence of the cell's regulatory network [11]. Clearly, the RBA approach forms the premises and possible basis of a strain design method.

However, we have also to consider additional constraints related to the implementation of the designed laws. For bacteria, the general principles governing gene expression emerged recently and allow to propose a first way to retune the level of existing genes or to plug new genes in modified strain. Actually, the new experimental tools for genome engineering, pave the way for massive bacterial genome modifications in a near future through for example the reprogramming of an existing bacteria chromosome (see e.g. [12], [13] to cite a very few). Today, a strong imbalance exists between the possibilities opened up by new genetic engineering tools and our ability to design a strain rationally. In a near future, the question will be no longer whether the changes can be made but what they are.

This paper proposes a first step in this direction. We revisited the standard design approach where only the culture conditions are optimized while integrating the constraints managing resource allocation and implementation of designed laws. In view of the complexity of cell functioning, our idea here is to present the bases of a new approach for the strain design. We emphasize in the rest of this article how

[1] MaIAGE, INRA, UR1404, Université Paris-Saclay, 78350 Jouy-en-Josas, France firstname.name@inra.fr

[2] Laboratoire des Signaux et Systèmes, CentraleSupélec - CNRS - Univ. Paris-Sud, Université Paris-Saclay, Control Department, Plateau du Moulon, 91190 Gif-sur-Yvette, France firstname.name@centralesupelec.fr

the degrees of freedom that this approach offers can be taken into account. The paper is organized as follows. In Section II, a general model including a cellular and an external medium description is presented. Then, the problem of strain design is raised on Section III. Realism in design appears in Section IV where the gene expressions are optimized versus growth rate. Finally, conclusions are drawn in Section V.

II. Modeling framework

A. Context

The model developed in this paper combines a batch reactor model and a bacterial cell model. The bacterial cell model is new and corresponds to the dynamical extension of the model developed in steady-state within the RBA framework [6], [9]. We then consider that a single bacterial species is immersed in a bioreactor operating in batch mode, i.e. with a given and constant volume of medium. The medium is supposed perfectly stirred and homogeneous.

B. Bacterial cell model

The cell functioning is described by its cell processes. These processes are composed of chemical reactions, transforming reactants into products. These reactions can be depicted by their flux: the transformation rate of the reaction. The flux of process $\mathscr{P}_*$ is denoted v_* and its value is $\bar{v}_*$. By convention, $\bar{v}_*$ are non negative. In this paper, we only consider the main cell processes, $\mathscr{P}_M$ for which compounds are transformed under the action of catalytic species (including transport). We also include the production processes of catalytic species, $\mathscr{P}_E$, that is to say gene expression processes, from gene to proteins. These processes are known to be the biggest energy and resource expenses in cell. Moreover, as presented in [5], the balance between these processes is known to limit the cell growth and force the cell to make choices and spare resource. All these processes are catalyzed by enzymes or macromolecules complexes, as e.g. ribosomes. These compounds catalyzing reactions are gathered in the set $\mathscr{E}$. The set of other compounds, reactants and products of processes $\mathscr{P}_M$, is denoted $\mathscr{M}$. By convention, we assume that every process (element of $\mathscr{P}_M$ of $\mathscr{P}_E$) is catalyzed by one and only one element of $\mathscr{E}$. Nonetheless, an element of $\mathscr{E}$ can catalyze multiple processes. The concentrations of elements of $\mathscr{E}$ (resp. $\mathscr{M}$) with respect to cell volume are gathered in vector E (resp. M).

The differential equations associated to M and E are given by:

$$\left\{\begin{array}{ll} \dot{M}(t) & = \Omega\bar{v}_M(t) + \Omega^E\bar{v}_E(t) - \mu(t)M(t) \\ \dot{E}(t) & = \bar{v}_E(t) - \mu(t)E(t) \end{array}\right. \tag{1}$$

where element $\Omega_{i,j}$ (resp. $\Omega^E_{i,j}$) of matrix Ω (resp. Ω^E) is the algebraic number of metabolite $\mathscr{M}_i$ produced, if positive, or consumed, if negative, by flux $v_{M,j}$ (resp. $v_{E,j}$). Factors $-\mu M$ and $-\mu E$ take into account the effects due to cell volume increase, where μ is the bacterial specific growth rate (see below).

As in RBA [9], enzymes and elements of $\mathscr{E}$ are supposed to have limited efficiency. Then, for each element $\mathscr{E}_i$ of $\mathscr{E}$:

$$\sum_{\mathscr{P}_j \in \mathscr{P}} \delta_{i,j}\bar{v}_j(t) \leq k_i(t)E_i(t) \tag{2}$$

where $k_i(t)$ is the efficiency of element $\mathscr{E}_i$ and where $\delta_{i,j} = 1$ if process $\mathscr{P}_j$ is catalyzed by $\mathscr{E}_i$, otherwise, $\delta_{i,j} = 0$. By biochemical knowledge, the efficiency of an enzyme depends on the concentration of the reactants and products of the reaction, leading to: $k_i(t) = k_i(M(t))$. Constraint (2) is rewritten in a more compact way as

$$\Delta\begin{bmatrix} \bar{v}_M(t) \\ \bar{v}_E(t) \end{bmatrix} \leq diag[k_M(t)]E(t) \tag{3}$$

where $\Delta = (\delta_{i,j})_{i,j}$ is a suitable matrix.

A subset of $\mathscr{M}$ are macrocomponents whose concentration is known to be constant in the cell, e.g. cell wall and membrane components. Denoting $\mathscr{M}_c$ this set of elements and M_c their concentrations, it comes:

$$\dot{M}_c(t) = \Omega_c\bar{v}_M(t) + \Omega^E_c\bar{v}_E(t) - \mu(t)M_c = 0 \tag{4}$$

where Ω_c (resp. Ω^E_c) is the submatrix of Ω (resp. Ω^E) corresponding to elements of $\mathscr{M}_c$. The relative complement of $\mathscr{M}_c$ in $\mathscr{M}$, $\mathscr{M}_i = \mathscr{M}\backslash\mathscr{M}_c$, gathers the elements with non zero dynamics. Their concentrations are gathered in M_i.

It remains to introduce an essential constraint of RBA, the so-called density constraint. Indeed, Kubitshek in [14] revealed that cell density D_0 is constant through different growth conditions and also along the cell cycle of bacteria as *Escherichia coli* or *Bacillus subtilis*. That leads, as in [15], to define a density D_0, related to the protein components from $\mathscr{E}$, by such a relation

$$D_0 = \sum_{\mathscr{E}} \ell_{E_i}E_i(t) \tag{5}$$

where ℓ_{E_i} corresponds to the number of amino-acids (or an equivalent) in the protein $\mathscr{E}_i$. Finally, in order to keep the density constant, and with E_i dynamics given by (1) the bacterial cell volume is increasing when new proteins are produced and thus is given by

$$\mu(t) = \frac{1}{D_0}\ell^T\bar{v}_E(t). \tag{6}$$

where ℓ is a vector containing the number of amino-acids of each element of $\mathscr{E}$.

Finally, the constraint of the density implies that the time evolution of the bacterial population concentration in mass in the bioreactor, denoted X, is given by:

$$\dot{X}(t) = \mu(t)X(t) \tag{7}$$

C. External reactor dynamics

We conclude by writing the dynamics for bioreactor components, denoted $\mathscr{M}_e$ (both produced and consumed by the cells in extracellular medium) whose concentrations, M_e, are expressed versus bioreactor volume. They are given by multiplying flux per unit of cell by biomass concentration in bioreactor:

$$\dot{M}_e(t) = \Omega_e\bar{v}_M(t)X(t) \tag{8}$$

where $\Omega_{ei,j}$ is the number of $\mathscr{M}_{ei}$ produced (consumed if negative) by reaction v_{Mj}.

D. The full model

The ODE and constraints describing the cell and the bioreactor are given by:

$$\left\{\begin{array}{rl} \dot{M}_i(t) & = \Omega_i \bar{v}_M(t) + \Omega_i^E \bar{v}_E(t) - M_i(t)\mu(t) \\ \dot{E}(t) & = \bar{v}_E(t) - E(t)\mu(t) \\ \dot{M}_e(t) & = \Omega_e \bar{v}_M(t) X(t) \\ \dot{X}(t) & = \mu(t) X(t) \end{array}\right. \tag{9}$$

$$w.r.t. \left\{\begin{array}{l} \Omega_c \bar{v}_M(t) + \Omega_c^E \bar{v}_E(t) - M_c \mu(t) = 0 \\ \mu(t) = \frac{1}{D_0} \ell^T \bar{v}_E(t) \\ \Delta \left[\begin{array}{c} \bar{v}_M(t) \\ \bar{v}_E(t) \end{array}\right] \leq diag[k_M(t)]E(t) \\ \bar{v}_E, \bar{v}_M, M_i, E, M_e, X \geq 0 \end{array}\right. \tag{10}$$

Substituting μ by its linear expression (6) in terms of $\bar{v}_E$ in (9) and (10) and with $\bar{v} = [\bar{v}_E^T, \bar{v}_M^T]^T$, $dim(\bar{v}) = m \times 1$, and $x = [M_i^T, E^T, M_e^T, X]^T$, $dim(x) = n \times 1$, dynamics and constraints can be expressed as:

$$\begin{array}{rl} & \dot{x}(t) = \Phi(x(t))\bar{v}(t) \\ w.r.t. & \left| \begin{array}{l} L\bar{v}(t) = 0 \\ \Delta \bar{v}(t) \leq \Psi(x(t)) \\ \bar{v}(t), x(t) \geq 0 \end{array}\right. \end{array} \tag{11}$$

with $dim(\Phi(x)) = n \times m$, $dim(L) = |\mathscr{M}_c| \times m$, $dim(\Delta) = |\mathscr{E}| \times m$, $dim(\Psi(x)) = |\mathscr{E}| \times 1$.

With an equivalent of the genome-scale model developed in [10], $|\mathscr{E}|$ would equal several hundreds, as for m and n (maybe several thousands). Clearly, the point is not to solve that problem at this scale, for the moment.

E. Simplified seven-process model

In the sequel, we derived a simplified model from the above general framework that contains the minimal key ingredients for the rational design of a strain dedicated to the production of a compound of interest. Indeed, as expressed in introduction, the objective of this paper is not to propose a design on hundreds of genes, but to give the essence of a strain design with realistic biological constraints. In this simplified model, we consider only one substrate, denoted $\mathscr{G}$ (like Glucose), within the bioreactor. The substrate $\mathscr{G}$ is imported and transformed by the process $\mathscr{P}_T$ into an intracellular substrate $\mathscr{S}$. The substrate $\mathscr{S}$ is used by process $\mathscr{P}_B$ to produce a macro-component $\mathscr{B}$ whose concentration shall remain equal to B_0. $\mathscr{S}$ is also used by the process $\mathscr{P}_P$ to produce and secrete a product of interest $\mathscr{P}$ into the culture medium. Finally, $\mathscr{S}$ is also the elementary brick that is used by the non-metabolic processes $\mathscr{P}_E$, divided in four gene expression processes $\mathscr{P}_{E_T}$, $\mathscr{P}_{E_B}$, $\mathscr{P}_{E_P}$, $\mathscr{P}_{E_R}$, respectively building up $\mathscr{E}_T$, $\mathscr{E}_B$, $\mathscr{E}_P$ and $\mathscr{E}_R$, which are the catalyzing compounds for processes $\mathscr{P}_T$, $\mathscr{P}_B$, $\mathscr{P}_P$, and the whole $\mathscr{P}_E$, respectively.

Following the formalism of the previous section, we have: $M_i = [S]$, $M_c = [B]$, $M_e = [G,P]^T$, $E = [E_T, E_B, E_P, E_R]^T$, $\bar{v}_M = [\bar{v}_T, \bar{v}_B, \bar{v}_P,]^T$, $\bar{v}_E = [\bar{v}_{E_T}, \bar{v}_{E_B}, \bar{v}_{E_P}, \bar{v}_{E_R}]^T$. We thus deduced the matrices $\Omega_i = [1,-1,-1]$, $\Omega_i^E = [1,1,1,1]$, $\Omega_c = [0,1,0]$, $\Omega_c^E = [0,0,0,0]$, $\Omega_e = \left[\begin{array}{ccc} -1 & 0 & 0 \\ 0 & 0 & 1 \end{array}\right]$, $\ell = [\ell_{E_T}, \ell_{E_B}, \ell_{E_P}, \ell_{E_R}]^T$ where ℓ_{E_T} (resp. ℓ_{E_B}, ℓ_{E_P}, ℓ_{E_R}) is the length in amino acids of compound $\mathscr{E}_T$ (resp. $\mathscr{E}_B$, $\mathscr{E}_P$, $\mathscr{E}_R$). Finally, we have

$$\Delta = \begin{array}{c} (\mathscr{P}_T :) \\ (\mathscr{P}_B :) \\ (\mathscr{P}_P :) \\ (\mathscr{P}_E :) \end{array} \left[\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]$$

and

$$k_M(t) = \left[\begin{array}{cccc} \frac{v_{m,T} G(t)}{G(t)+K_T+S(t)/K_S}, & \frac{v_{m,B} S(t)}{S(t)+K_B}, & \frac{v_{m,P} S(t)}{S(t)+K_P}, & \frac{v_{m,R} S(t)}{S(t)+K_R} \end{array}\right].$$

$\mathscr{P}_B$ and $\mathscr{P}_P$ are metabolic processes, and k_B, k_P are assumed to be two classical Michaelis-Menten-like efficiencies depending on substrate $\mathscr{S}$. The efficiency of $\mathscr{P}_E$, k_R, is also assumed to be a Michaelis-Menten-like efficiency increasing and saturated by $\mathscr{S}$ concentration, in accordance with [15]. For process $\mathscr{P}_T$, we further integrated the cost of transport of $\mathscr{G}$ as a function of the internal concentration $\mathscr{S}$. The efficiency k_T is assumed to follow a Briggs-Haldane like efficiency law, increasing in $\mathscr{G}$, decreasing in $\mathscr{S}$ and saturated for large values of $\mathscr{G}$. Finally, in each function of efficiencies, $v_{m,*}$ corresponds to the maximal rate of $\mathscr{P}_*$, K_* corresponds to the Michaelis-Menten constants or inhibition constants depending on the context.

Representing the whole-cell through only five major cell processes ($\mathscr{P}_T$, $\mathscr{P}_B$, $\mathscr{P}_{E_T}$, $\mathscr{P}_{E_B}$, $\mathscr{P}_{E_R}$) plus two ($\mathscr{P}_P$, $\mathscr{P}_{E_P}$) dedicated to the production of $\mathscr{P}$ is obviously drastic compared to the 73 cellular processes of the complete model [10]. However, we can distribute easily all cellular processes into $\mathscr{P}_T$, $\mathscr{P}_B$, $\mathscr{P}_E$ due to the systemic nature of the bacterial cell. Only the introduction of $\mathscr{S}$ has to be discussed with respect to the RBA framework. Actually, the concentrations of internal metabolites are known to be globally higher in rich medium that in poor medium [16]. This leads to increase globally the efficiency of almost all cell processes in rich medium compared to poor medium. $\mathscr{S}$ was thus introduced to capture the impact of the medium composition (rich vs. poor) on process efficiencies.

III. Open-loop optimal control

In this section we firstly define the criteria that has to be maximized, then recall some results guaranteeing the existence of a maximum and its associated open-loop optimal control and thus a way to compute it. The main interest of this first design is to provide an upper limit of the achievable criteria.

A. Criterion

The maximization of production can be formulated in various ways. The final quantity or concentration of product of interest is the most natural criterion, but the time necessary to obtain this maximal quantity is also an essential issue. A compromise between product quantity and time of the process has to be found. As suggested in [17], a solution could be to find the Pareto front of maximum final quantity of product versus culture duration. In view of our goal, we consider another compromise as in [18], defining the criterion as the ratio between the concentration of the product at the final time over the culture duration. If x_j is the concentration of the product of interest in vector x, criterion to maximize would be $J = \frac{x_j(t_f)}{t_f}$, where by definition $t_f > 0$ is the final time.

B. Optimization problem formulation

The problem of strain design for maximizing a product of interest can be formulated in a compact way as a Mayer optimization problem:

$$\max_{\bar{\nu},t_f} J(x(t_f),t_f) = \frac{1}{t_f} c^T x(t_f)$$

$$\textit{with respect to} \left\{ \begin{array}{l} \textit{Dynamics } (9) : \dot{x}(t) = \Phi(x(t))\bar{\nu}(t) \\ \textit{Constraints } (10) : \left| \begin{array}{l} L\bar{\nu}(t) = 0 \\ \Delta\bar{\nu}(t) \leq \Psi(x(t)) \\ \bar{\nu}(t), x(t) \geq 0 \end{array} \right. \\ (x(t_0),t_0) \in Z_0,\ (x(t_f),t_f) \in Z_f \end{array} \right. \tag{12}$$

where c is the vector of cost for each compound (typically, all zero except 1 for the index of the compound of interest), Z_0 and Z_f are defined as initial and final suitable state sets, in particular $t_f \geq \varepsilon > 0$ is part of Z_f definition.

Notice that the formulation is exactly the same, no matter is the type of the compound of interest: it can be indifferently biomass, protein, intra- or extracellular metabolite.

C. Existence of the open-loop optimal control & necessary conditions

Let us define $\Pi(x) = \{\bar{\nu} \in \mathbb{R}^m_+ | L\bar{\nu} = 0, \Delta\bar{\nu} \leq \Psi(x)\}$ and $N(x) = \{\Phi(x)\bar{\nu} | \bar{\nu} \in \Pi(x)\}$. By construction, it is possible to prove that $\Psi(x(t))$ is necessarily bounded and the constraints defining $\Pi(x)$ for a given x are then linear constraints. Consequently the set $\Pi(x)$ is necessarily a convex set. The convexity of $\Pi(x)$ implies the one of $N(x)$. In our case, it is a simple exercise to show that the set of feasible trajectories is not empty, that x, $\bar{\nu}$ and t_f are necessarily all bounded. Following these preliminary elements, the classic Filippov's existence theorem can be invoked in order to prove that the optimal control problem has a solution (see e.g. [19][Chapter 4.3 in particular] or [20] and [21]. We then know that there exists necessarily an optimal triple $\{t_f^*, x^*(t), \bar{\nu}^*(t)\}$ for problem (12) with $\bar{\nu}^*$ measurable.

In order to compute an open-loop optimal control input, it is then classical to invoke the Pontryagin's maximum principle. However, our optimal control problem defined by (12) corresponds to a so-called mixed state-control constraint. Indeed, for a given x, the set of input values has to belong to $\Pi(x)$. In this context, the maximum principle is replaced by more advanced conditions such as the ones associated to a Mayer problem where the dynamics of the system is defined by a nonlinear differential inclusion (see e.g. [22], [23]). If we further assume that the optimal input has some suitable continuity property, more classic conditions can be obtained such as the ones derived for instance in [21]. In view of our goal, we do not detailed further this advanced aspect on optimization. We only emphasize here the necessity to choose a numerical scheme that can handle specificities of mixed-control-state constraints (we refer readers to [24] for a complete and clear presentation of this general issue and possible remedies).

D. Numerical resolution for the simplified seven-process model

We thus use `Bocop` [25] to numerically solve the optimal problem defined by (12). This open-source toolbox converts the infinite dimensional optimal control problem into a finite dimensional non-linear optimization problem by time-discretization on state and control variables, i.e. the so-called direct transcription approach, see [24] for details. Following the discretization of the initial problem, the following non-linear optimal problem is defined:

$$\begin{array}{l} \max\limits_{\mathbf{X} \in \mathbb{R}^{(n+m)\times N+1}} F(\mathbf{X}), \\ w.r.t.\ C(\mathbf{X}) \geq 0 \end{array} \tag{13}$$

with $\mathbf{X} = \{\{x_i, \bar{\nu}_i\}_{i\in\{1,\ldots,N\}}, t_f\} \in \mathbb{R}^{(n+m)\times N+1}$, where N is the number of discretization points, n the dimension of x, m the dimension of $\bar{\nu}$. Finally, $C(\mathbf{X}) \geq 0$ sums up the set of all the constraints, including equality constraints. It thus takes into account all constraints associated to the discretization of system dynamics, i.e. $x_{i+1} = x_i + \frac{t_f - t_i}{N}\Phi(x_i)\bar{\nu}_i$; the inequality constrains on inputs at each time points, i.e. $L\bar{\nu}_i = 0$; $\Delta\bar{\nu}_i \leq \Psi(x_i)$; $\bar{\nu}_i \geq 0$; $x_i \geq 0$; and finally integrates the boundary constraints, i.e. $(x_0,t_0) \in Z_0$ and $(x_N,t_f) \in Z_f$. This nonlinear optimal problem is then solved by `Bocop`, see [24], [25].

E. Optimal open-loop control as a reference

Parameters are chosen in accordance with RBA approach [10] with initial and final conditions, corresponding to sets Z_0 and Z_f in (12): $G(t_i) = 10 mmol.L^{-1}$, $P(t_i) = 0$, $X(t_i) = 4.5 mg.L^{-1}$, $\ell^T E(t_i) = D_0 = 2.51 mmol.g_{CDW}^{-1}$, $S(t_i) = 1 \mu mol.g_{CDW}^{-1}$, $B(t_i) = 2 \mu mol.g_{CDW}^{-1}$, $G(t_f) = 0$, $t_f \geq \varepsilon$, where g_{CDW} stands for gram of cell dry weight.

The time evolution of the main variables, obtained by the resolution of (13) with above initial conditions, are given in Fig 1. The numerical optimal trajectory presents three main phases:

(i) Proteins are first allocated towards biomass synthesis only: cells grow at constant growth rate, without production of $\mathscr{P}$. $\bar{\nu}_{E_P}$ is null. A balanced exponential regime is recovered.

(ii) Cell mass increases linearly while the protein repartition is switching towards the synthesis pathway of product of interest. The growth is decreasing, reflecting the arrest of the synthesis of proteins within processes linked to growth: $\bar{\nu}_{E_B}$ and $\bar{\nu}_{E_R}$ are null.

(iii) The production of all proteins stops leading to a last phase in which the production is at its maximum, with few $\mathscr{E}_B$ and $\mathscr{E}_R$. Compared to phase (i), the repartition of proteins shows that $\mathscr{E}_P$ has replaced $\mathscr{E}_B$.

We obtained an optimal production strategy using our seven-process model that is close to a standard bang-bang politics [2]. We already obtain an optimal control profile that is realistic from the biological point of view. Indeed, the optimal profile integrates explicitly the time necessary (i.e. a transition phase) to build all cellular components required for the production of $\mathscr{P}$. This is due to the operating constraints that we integrated into our optimization problem. They prevent to switch instantaneously from the configuration of biomass production to the one of $\mathscr{P}$ production.

IV. Constraints implied by biological implementation

A. Motivation for a growth rate dependent gene expression

Here we introduce in the design optimization problem the constraints related to the biological ‘implementation’ of the designed laws. At a laboratory scale, gene expression can be controlled through external signals such as a given and defined metabolite. However, the number of such signals is very low. Moreover, using external signals imposes strong constraints on the bioreactors operating at the industrial scale (i.e. a volume of a few hundred liters typically, even more), and can be very costly. Using the internal genetic regulatory mechanisms of the cell should thus be more suitable. This seems achievable today.

Indeed, the major principles governing the bacterial gene expressions are now better understood and biologically characterized. The genetic regulatory network (e.g. transcription factors, etc.) is now widely identified for bacteria such as *E. coli* or *B. subtilis*. Moreover, in addition to these regulatory mechanisms, bacteria have a class of specific genes, called constitutive, without any known genetic regulation. The evolution of the expression of each constitutive gene is specific and growth rate dependent. The growth rate dependent regulation is mainly achieved at the level of the initiation rate of gene transcription, see e.g. [26], [27] and of messenger translation, see e.g. [28]. Consequently, we have theoretically access to a large well-characterized set of gene expression profiles (as functions of cell growth rate).

B. Problem formulation

The objective is to have only a few number of processes that are controlled from an external signal and the majority of the processes that are controlled using internal regulatory mechanisms of the cell, in particular through a growth controlled gene expression. Let $\mathscr{P}_\mu$ the set of processes that shall be controlled by growth, and ν_μ the associated fluxes. Then, the problem is the same as in Section III with additional constraints on $\bar{\nu}_\mu$ for growth dependency: $\bar{\nu}_E(t) = \bar{\nu}_E(\mu(t))$.

From the regular shape of the curves presented in [27], [28], it seems reasonnable to approximate $\bar{\nu}_E(\mu)$ functions by polynomial expressions: $\bar{\nu}_E(\mu) = \sum_1^d u_k \mu^k$, where u_k are polynomial coefficients and d the degree of polynomials. Note that from (6) and as $\bar{\nu}_E \geq 0$, $u_0 = 0$. This leads to the implementation constraint :

$$\bar{\nu}_{E_i}(t) = \sum_{k=1}^{d} u_{i,k} \mu(t)^k, \forall \nu_{E_i} \in \nu_\mu \tag{14}$$

By adding the u coefficients in the optimization variables, the new mixed-control-state constraints are fully handled by `Bocop`. By abuse of notation, this kind of optimization problem are called closed-loop optimization problem in the sequel.

TABLE I

Comparison of optimality indexes for open-loop optimization (OL), and closed-loop optimization (CL) cases A and B

	OL	CL Case A	CL Case B
$J = P(t_f)/t_f$ (in $mmol.L^{-1}.h^{-1}$)	0.38	0.35	0.31
$P(t_f)$ (in $mmol.L^{-1}$)	2.3	2.2	2.3
t_f (in h)	6.1	6.3	7.4

C. Solution for the seven-process model

1) Case A (Full design): In the simplified seven-process model, the three non-metabolic processes are supposed to be growth controlled, $\mathscr{P}_\mu = \{\mathscr{P}_{E_P}, \mathscr{P}_{E_B}, \mathscr{P}_{E_T}\}$. We choose $d = 3$ for simplicity. Solution of this close-loop problem A is depicted in blue in Fig. 1. Macroscopic quantities as biomass concentration X, growth rate μ and product concentration P are much more regular but remain very close to the open-loop optimal solution. Moreover, the final concentrations of $\mathscr{P}$ are also very close, see Tab. I. This suggests that the loss of optimality is reasonable compared to the fact that no process is controlled by an external signal.

2) Case B (Rational design): In Case A, we assumed that all genes in the bacterium are somehow re-adjusted/retuned. Here, we can realistically assume that the evolution of protein concentrations dedicated to the production of the biomass, i.e. E_B, follows a linear law with respect to μ, i.e. $E_B = \alpha_B \mu$. With quasi-steady state assumption on E_B dynamics (i.e. $\dot{E}_B(t) = 0$), it comes: $\bar{\nu}_{E_B}(t) = \alpha_B \mu(t)^2$, where α_B is assumed to be a known parameter.

We furthermore assume that the evolution of protein concentrations dedicated to the production pathway E_P are also linear in μ. Their gene expressions thus follow a quadratic law, i.e. $\bar{\nu}_{E_P}(t) = u_P \mu(t)^2$ where the coefficient u_P is then a parameter to be optimized. In addition to this quadratic formulation, we introduce an external time control that activates the expression of genes associated to $\mathscr{P}_{E_P}$, thus mimicking the activation of $\mathscr{P}_{E_P}$ gene expression by a transcription factor sensitive to an extracellular signal. This control signal, $\varepsilon(t)$, is binary and can switch once and only once at time t_s: $\varepsilon(t) = 0$ for $t \leq t_s$, $\varepsilon(t) = 1$ for $t > t_s$. Gene expression of $\mathscr{E}_P$ is then given by:

$$\bar{\nu}_{E_P}(t) = u_P \mu(t)^2 \varepsilon(t) \tag{15}$$

with coefficient u_P and switching time t_s to be determined.

We finally assume that processes $\mathscr{P}_T$ and $\mathscr{P}_R$ are adapted in agreement to the demand and thus $\bar{\nu}_{E_T}$ and $\bar{\nu}_{E_R}$ are assumed to be free variables with respect to the growth rate.

Numerical solution is obtained by Bocop and is presented on Fig. 1 We obtain the same three phases (growth, transition, production) as in the open-loop case. The sharpness of $\mu(t)$ is due to the switch occurring at $t_s = 1.85h$.

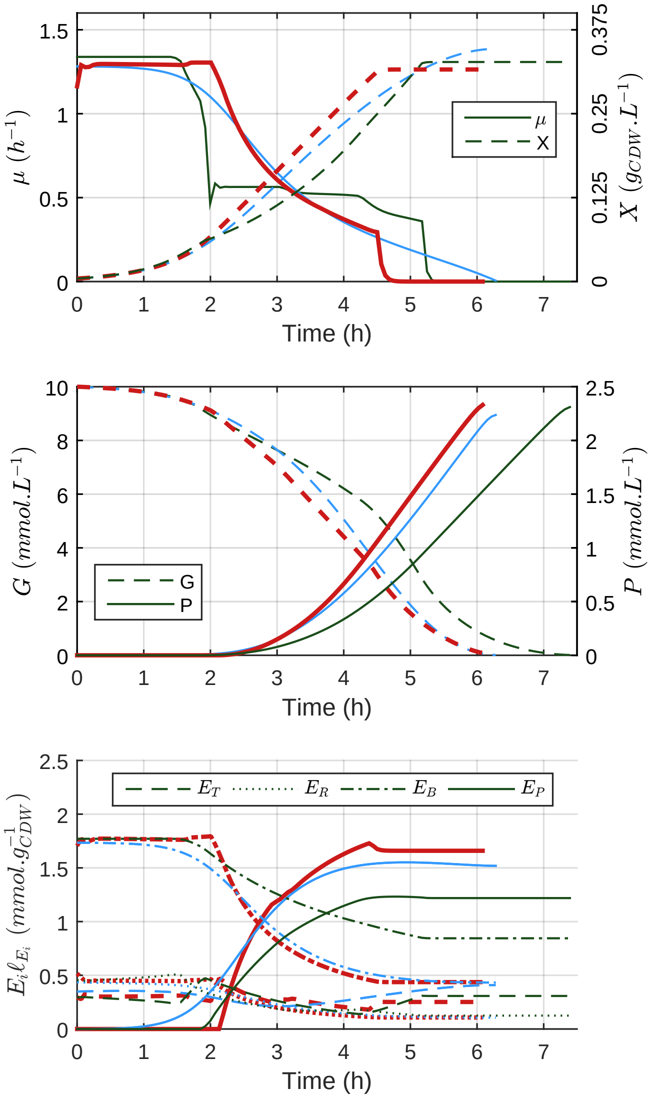

Fig. 1. Evolution of variables for the open-loop optimization problem, in red. Results for close-loop optimization problem case A, in blue. Results for close-loop optimization problem case B, in green.

V. Conclusions

To conclude, we introduced in this paper a first step towards an integrated strain design in batch bioprocess. We investigated the question of the biological implementation of optimal control. We showed that the closed-loop implementation of the optimal control using the growth rate dependency of gene expression slightly degraded the performance obtained through an open-loop optimal control. Finally, we emphasize through a simple example that the procedure of strain design can integrate the existing regulatory network of the cell in order to minimize as much as possible the number of genome modifications.

References

[1] D. Dochain, *Automatic control of bioprocesses*. John Wiley & Sons, 2013.

[2] J. E. Cuthrell and L. T. Biegler, “Simultaneous optimization and solution methods for batch reactor control profiles,” *Computers & Chemical Engineering*, vol. 13, no. 1-2, pp. 49–62, 1989.

[3] N. E. Lewis *et al.*, “Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods,” *Nature Reviews Microbiology*, vol. 10, no. 4, p. 291, 2012.

[4] A. Chowdhury *et al.*, “Bilevel optimization techniques in computational strain design,” *Computers & Chemical Engineering*, vol. 72, pp. 363–372, 2015.

[5] A. Goelzer and V. Fromion, “Bacterial growth rate reflects a bottleneck in resource allocation,” *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1810, no. 10, pp. 978–988, 2011.

[6] A. Goelzer *et al.*, “Cell design in bacteria as a convex optimization problem,” in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, December 2009, pp. 4517–4522.

[7] D. Molenaar *et al.*, “Shifts in growth strategies reflect tradeoffs in cellular economics,” *Molecular Systems Biology*, vol. 5, nov 2009.

[8] M. Scott *et al.*, “Interdependence of cell growth and gene expression: Origins and consequences,” *Science*, vol. 330, no. 6007, pp. 1099–1102, nov 2010.

[9] A. Goelzer *et al.*, “Cell design in bacteria as a convex optimization problem,” *Automatica*, vol. 47, no. 6, pp. 1210–1218, 2011.

[10] ——, “Quantitative prediction of genome-wide resource allocation in bacteria,” *Metabolic engineering*, vol. 32, pp. 232–243, 2015.

[11] L. Tournier *et al.*, “Optimal resource allocation enables mathematical exploration of microbial metabolic configurations,” *Journal of Mathematical Biology*, vol. 75, no. 6-7, pp. 1349–1380, mar 2017.

[12] D. G. Gibson *et al.*, “Creation of a bacterial cell controlled by a chemically synthesized genome,” *Science*, vol. 329, no. 5987, pp. 52–56, may 2010.

[13] H. Wang *et al.*, “CRISPR/cas9 in genome editing and beyond,” *Annual Review of Biochemistry*, vol. 85, no. 1, pp. 227–264, jun 2016.

[14] H. E. Kubitschek *et al.*, “Independence of buoyant cell density and growth rate in escherichia coli.” *Journal of bacteriology*, vol. 158, no. 1, pp. 296–299, 1984.

[15] A. G. Marr, “Growth rate of escherichia coli.” *Microbiological reviews*, vol. 55, no. 2, pp. 316–333, 1991.

[16] V. M. Boer *et al.*, “Growth-limiting intracellular metabolites in yeast growing under diverse nutrient limitations,” *Molecular biology of the cell*, vol. 21, no. 1, pp. 198–211, 2010.

[17] K. G. Gadkar *et al.*, “Estimating optimal profiles of genetic alterations using constraint-based models,” *Biotechnology and bioengineering*, vol. 89, no. 2, pp. 243–251, 2005.

[18] B. Jabarivelisdeh and S. Waldherr, “Improving bioprocess productivity using constraint-based models in a dynamic optimization scheme,” *IFAC-PapersOnLine*, vol. 49, no. 26, pp. 245–251, 2016.

[19] L. Cesari, *Optimization-theory and applications: problems with ordinary differential equations*, ser. Application Mathematics. Springer-Verlarg, 1984, vol. 17.

[20] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*. John Wiley and Sons, 1967.

[21] R. F. Hartl *et al.*, “A survey of the maximum principles for optimal control problems with state constraints,” *SIAM review*, vol. 37, no. 2, pp. 181–218, 1995.

[22] H. Frankowska, “The maximum principle for a differential inclusion problem,” in *Analysis and Optimization of Systems*. Springer, 1984, pp. 517–531.

[23] R. Vinter, *Optimal control*. Springer Science, 2010.

[24] L. T. Biegler, *Nonlinear programming: concepts, algorithms, and applications to chemical processes*. Siam, 2010, vol. 10.

[25] J. Bonnans, Frederic *et al.*, “Bocop A collection of examples,” INRIA, Tech. Rep., 2017. [Online]. Available: http://www.bocop.org

[26] S. Klumpp and T. Hwa, “Growth-rate-dependent partitioning of RNA polymerases in bacteria,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 51, pp. 20 245–20 250, 2008.

[27] L. Gerosa *et al.*, “Dissecting specific and global transcriptional regulation of bacterial gene expression,” *Molecular systems biology*, vol. 9, no. 1, p. 658, 2013.

[28] O. Borkowski *et al.*, “Translation elicits a growth rate-dependent, genome-wide, differential protein production in bacillus subtilis,” *Molecular systems biology*, vol. 12, no. 5, p. 870, 2016.