Research Article

Design of Stable Metabolic Networks

Jimena Di Maggio [1]

Aníbal M. Blanco[1]

J. Alberto Bandoni[1]

Juan Carlos Diaz Ricci[2]

M. Soledad Diaz[1]

[1] Planta Piloto de Ingeniería Química (PLAPIQUI), Universidad Nacional del Sur-CONICET, Camino La Carrindanga Km 7, Bahía Blanca 8000, Argentina

[2] Instituto Superior de Investigaciones Biológicas-INSIBIO (UNT-CONICET), Chacabuco 461, San Miguel de Tucumán 4000, Argentina

Received: 03 16, 2017; Revised: 06 02, 2017; Accepted: 06 06, 2017

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/elsc.201700065.

Correspondence: Prof. M. Soledad Diaz (sdiaz@plapiqui.edu.ar). Planta Piloto de Ingeniería Química (PLAPIQUI), Universidad Nacional del Sur-CONICET, Camino La Carrindanga Km 7, Bahía Blanca 8000, Argentina.

Keywords: Eigenvalue optimization, Metabolic networks, Non-linear dynamic models, Optimal design, Stability

Practical application

Modeling the metabolism of a microorganism requires a deep knowledge of the main metabolic and regulatory pathways participating, as well as the conditions under which the model can bring reliable results. In this sense, regardless how detailed the models may be, many times they fail, and the reason is that they further require the evaluation of the system stability so that culture conditions can remain stable under a certain range of conditions.

In this paper we explore the dynamic stability of the central metabolism of *E. coli* to investigate the main controlling steps. The aim of the paper is to provide a computational tool, based on nonlinear programming techniques and eigenvalue optimization theory, to evaluate whether the expression of a gene on a metabolic engineered metabolism is stable at particular culture conditions, or within which range of external variables we can expect a stable behavior.

Abstract

In this work, we propose eigenvalue optimization combined with Lyapunov theory concepts to ensure stability of the Embden-Meyerhof-Parnas pathway, the pentose-phosphate pathway, the phosphotransferase system and fermentation reactions of *Escherichia coli*. We address the design of a metabolic network for the maximization of different metabolite production rates. The first case study focuses on serine production, based on a model that consists of eighteen differential equations corresponding to dynamic mass balances for extracellular glucose and intracellular metabolites, and thirty kinetic rate expressions. A second case study addresses the design problem to maximize ethanol production, based on a dynamic model that involves mass balances for twenty-five metabolites and thirty eight kinetic rate equations. The nonlinear optimization problem including stability constraints has been solved with reduced space Successive Quadratic Programming techniques. Numerical results provide useful insights on the stability properties of the studied kinetic models.

1 Introduction

Metabolic network design can be formulated as an optimization problem to achieve a given objective, for example an increase in the production rate of a certain metabolite, subject to mass balance equations that represent the network. Due to nonlinear kinetics of the biochemical reactions and their coupling through common metabolites, biological systems may undergo drastic changes in their qualitative behavior when a variation on the enzyme level occurs. Kinetic models of metabolic networks allow the analysis of stability of the predicted states, which is of fundamental importance because biological systems may exhibit monotonic stable states, bi-stable switching threshold phenomena, oscillations and chaotic behavior. If no stability constraints are included in the design problem formulation, the optimal operating point might be unstable, making the metabolic network vulnerable to external disturbances. In other words, in spite of the presence of modest disturbances an unstable network may reach physiological constraints and collapse. Several authors have addressed the analysis of biological systems of small to moderate size [1, 2, 3, 4, 5]. Diaz Ricci [1] studies local stability of phosphofructokinase (PFK) enzyme through the analysis of the eigenvalues of a kinetic model that comprises mass balances for its substrate and product. Results show that the dynamic behavior of PFK could exhibit instabilities depending on substrate concentration. Vital-Lopez et al. [6] study the stability of the central carbon metabolism of *E. coli* at optimal enzyme levels determined by Nikolaev et al. [7] for the production of serine. The authors construct bifurcation diagrams that show that the central carbon metabolism can exhibit oscillatory behavior for certain parameter values. Ivanov et al. [4] study the properties of steady states in metabolic networks with monotonic kinetics, in relation to their stoichiometry and the number of metabolites that participate in every reaction. The study is based on the investigation of the Jacobian matrix properties..

The design-for-stability problem, an important sub problem of the general design-for-operability problem, has also motivated many contributions from the process systems engineering community. Different strategies have been proposed to include stability considerations within the design problem [8, 9]. In this sense, several authors have formulated different problems to determine optimal designs of metabolic networks. Visser et al.[10] formulate an optimal design problem to maximize serine production and the flux through the phosphotransferase system in *Escherichia coli*. Chang and Sahinidis [8] propose solution with a global optimization algorithm, for metabolic network design problem that includes stability constraints through the Routh-Hurwitz criterion and the Vieta theorem. The algorithm is applied to a small metabolic network model formulated to describe tryptophan biosynthesis and to another one that describes anaerobic fermentation in *Saccharomyces cerevisiae.*

Gerhard et al. [11] address the steady-state optimization problem that ensures the stability of a tryptophan biosynthesis network using the constructive nonlinear dynamics approach [12]. Furthermore, Nikolaev [13] formulates a mixed integer non-linear problem, based on a kinetic model for the central carbon metabolism of *Escherichia coli*, to compute which enzyme levels should be modulated and which enzyme regulatory structures should be altered to maximize serine production and the flux through the phosphotransferase system. The author includes constraints to ensure that the real part of the eigenvalues of the Jacobian matrix for the linearized system is negative. To solve the MINLP problem, the author implements a simulated annealing algorithm.

In this work, we propose an eigenvalue optimization approach to ensure stability of the glycolysis, the pentose-phosphate pathway, the phosphotransferase system and fermentation reactions of *Escherichia coli* K-12 W3110, on a detailed kinetic model [14]. The nonlinear optimization problem, corresponding to the right hand side of differential equations, rate equations and stability constraints, is solved with reduced space Successive Quadratic Programming techniques with program IPOPT [15]. Optimization results provide an improved metabolic network for the maximization of serine production in one case and the maximization of ethanol production, as a second case of study, in *Escherichia coli*.

2 Materials and Methods

2. 1. Mathematical model of *Escherichia coli* metabolism

Dynamic models of metabolic networks comprise a set of differential and algebraic equations that arises from mass balances for intra and extracellular metabolites and from kinetic expressions for enzymes catalytic activity. In this work, we present an extended mathematical model of an extension of the dynamic model for the central carbon metabolism of *Escherichia coli* K-12 W3110 ([16], [14]). The proposed model comprises the Embden-Meyerhof-Parnas pathway, the pentose phosphate pathway, the phosphotransferase system and fermentation reactions (Figure 1).

Mass balances for extracellular metabolites are represented by Eqn. (1), where C_i^{feed} (mM) and $C_i^{extracellular}$ (mM) represent each extracellular metabolite *i* concentration in the feed and the reactor, respectively. D (s^{-1}) is the dilution rate, r_j (mM s^{-1}) is the reaction rate associated to each enzyme *j*, v_{ij} is the stoichiometric coefficient for metabolite *i* in reaction *j*, C_x (g DW L^{-1} culture volume) and ρ_x (g DW L^{-1} cell volume) correspond to biomass concentration and specific weight, respectively.

$$\frac{dC_i^{extracellular}}{dt} = D\left(C_i^{feed} - C_i^{extracellular}\right) + v_{ij} r_j \left(\frac{C_x}{\rho_x}\right) \quad (1)$$

i = extracellular glucose, ethanol, acetate, formate, lactate, succinate
j *= PTS, LDH, ACK, ADH, PFL, SucSynthesis*

Fermentation products (ethanol, acetate, formate, lactate and succinate) are considered as extracellular metabolites since it is assumed that their transport from the cytoplasm to the culture medium is fast enough[17]. For these extracellular metabolites C_i^{feed} is equal to zero.

A general expression for intracellular metabolite mass balances is represented by Eqn. (2), where r_l (mM s^{-1}) is the reaction rate associated to each enzyme *l*, v_{kl} is the stoichiometric coefficient for metabolite *k* in reaction *l* and μ (s^{-1}) is the specific growth rate. The term μC_k in Eqn. (2) stands for the dilution rate of metabolite *k* due to biomass growth.

$$\frac{dC_k}{dt} = \sum_l v_{kl} r_l - \mu C_k \quad (2)$$

k = g6p, f6p, fdp, gap, dhap, pgp, 3pg, 2pg, pep, pyr, 6pg, ribu5p, xyl5p,

sed7p, rib5p, e4p, g1p, AcCoA, acetaldehyde, acetyl-P.

l = PTS, PGI, PFK, ALDO, TIS, GAPDH, PGK, PGluMu, ENO, PK, PDH, PEPCxylase, PGM, G1PAT, RPPK, G3PDH, SerSynth, MurSynth, DAHPS, TrpSynth, MetSynth, G6PDH, PGDH, Ru5p, R5PI, TKa, TKb, TA, Synth1, Synth2, ACALDH, PTA, LDH, ACK, ADH, PFL, SucSynthesis.

More details of the central carbon metabolism model of *Escherichia coli*, along with the PPP and fermentation reactions can be found in [14]. Mass balances and kinetic rate equations are shown in Tables A1 and A2 of Supporting information, respectively. Mass balances for pyruvate, intracellular metabolites involved in the fermentation reactions (acetyl-coenzyme A (AcCoA), acetyl-phosphate (Ac-P) and acetaldehyde (Acal)) and each fermentation product are described in Eqns. (3) to (11).

$$\frac{dC_{pyr}}{dt} = r_{PK} + r_{PTS} - r_{PDH} - r_{Synth2} + r_{MetSynthesis} + r_{TrpSynthesis} - r_{LDH} - r_{PFL} - \mu C_{pyr} \quad (3)$$

$$\frac{dC_{AcCoA}}{dt} = r_{PDH} + r_{PFL} - r_{ACALDH} - r_{PTA} - 2r_{MurSynth} - \mu C_{AcCoA} \quad (4)$$

$$\frac{dC_{Acal}}{dt} = r_{ACALDH} - r_{ADH} - \mu C_{Acal} \quad (5)$$

$$\frac{dC_{AcP}}{dt} = r_{PTA} - r_{ACK} - \mu C_{AcP} \quad (6)$$

$$\frac{dC_{Ethanol}}{dt} = r_{ADH}\left(\frac{C_x}{\rho_x}\right) - DC_{Ethanol} \quad (7)$$

$$\frac{dC_{Formate}}{dt} = r_{PFL}\left(\frac{C_x}{\rho_x}\right) - DC_{Formate} \quad (8)$$

$$\frac{dC_{Lactate}}{dt} = r_{LDH}\left(\frac{C_x}{\rho_x}\right) - DC_{Lactate} \quad (9)$$

$$\frac{dC_{Acetate}}{dt} = r_{ACK}\left(\frac{C_x}{\rho_x}\right) - DC_{Acetate} \quad (10)$$

$$\frac{dC_{Succinate}}{dt} = r_{SucSynthesis}\left(\frac{C_x}{\rho_x}\right) - DC_{Succinate} \quad (11)$$

Kinetic expressions of fermentation reactions rates are taken from the bibliography. In this sense, enzyme kinetic equations for acetate kinase (ACK), acetaldehyde dehydrogenase (ACALDH), alcohol dehydrogenase (ADH) and lactate dehydrogenase (LDH), Eqns. (12) to (15), are considered from [18], who propose a detailed kinetic model for fermentation pathways of lactic acid bacteria. On the other hand, kinetic expressions for phosphotransacetylase (PTA) (Eqn. (16)) and pyruvate formate lyase (PFL) (Eqn. (17)) are taken from [19] and [20], respectively. Synthesis of succinate (SucSynthesis) is described by a Michaelis-Menten equation (Eqn. (18)). Kinetic parameter descriptions are shown in Table A3 of Supporting information.

$$r_{PTA} = \frac{r_{PTA}^{max}\left(\frac{1}{K_{i,AcCoA}K_{mP}}\right)\left(C_{AcCoA}C_P - \frac{C_{AcP}C_{CoA}}{K_{eq}}\right)}{1 + \frac{C_{AcCoA}}{K_{i,AcCoA}} + \frac{C_P}{K_{i,P}} + \frac{C_{AcP}}{K_{i,AcP}} + \frac{C_{CoA}}{K_{i,CoA}} + \frac{C_{AcCoA}C_P}{K_{i,AcCoA}K_{mP}} + \frac{C_{AcP}C_{CoA}}{K_{i,CoA}K_{m,AcP}}} \quad (12)$$

$$r_{ACK} = \frac{r_{ACK}^{max}\left(K_{m,adp}K_{m,acp}\right)^{-1}\left(C_{acp}C_{adp} - \frac{C_{acetate}C_{atp}}{K_{eq}}\right)}{\left(1 + \frac{C_{acp}}{K_{m,AcP}} + \frac{C_{acetate}}{K_{m,acetate}}\right)\left(1 + \frac{C_{adp}}{K_{m,adp}} + \frac{C_{atp}}{K_{m,atp}}\right)} \quad (13)$$

$$r_{ACALDH} = \frac{r_{ACALDH}^{max}\left(\frac{1}{K_{m,AcCoA}K_{m,nadh}}\right)\left(C_{AcCoA}C_{nadh} - \frac{C_{CoA}C_{nad}C_{acal}}{K_{eq}}\right)}{\left(1 + \frac{C_{nad}}{K_{m,nad}}\frac{C_{nadh}}{K_{m,nadh}}\right)\left(1 + \frac{C_{AcCoA}}{K_{m,AcCoA}} + \frac{C_{CoA}}{K_{m,CoA}} + \frac{C_{acal}}{K_{m,acal}} + \frac{C_{CoA}C_{acal}}{K_{m,CoA}K_{m,acal}}\right)} \quad (14)$$

$$r_{ADH} = \frac{r_{ADH}^{max}\left(K_{m,acal}K_{m,nadh}\right)^{-1}\left(C_{acal}C_{nadh} - \frac{C_{ethanol}C_{nad}}{K_{eq}}\right)}{\left(1 + \frac{C_{nad}}{K_{m,nad}} + \frac{C_{nadh}}{K_{m,nadh}}\right)\left(1 + \frac{C_{acal}}{K_{m,acal}} + \frac{C_{ethanol}}{K_{m,ethanol}}\right)} \quad (15)$$

(16)

$$r_{PFL} = \frac{r^{max}_{PFL,for} C_{pyr} C_{CoA}}{\left(C_{pyr} C_{CoA} + C_{pyr} K_{m,CoA} + C_{CoA} K_{m,pyr}\right)} - \frac{r^{max}_{PFL,rev} C_{formate} C_{AcCoA}}{\left(C_{formate} C_{AcCoA} + C_{formate} K_{m,AcCoA} + C_{AcCoA} K_{m,formate}\right)}$$

$$r_{LDH} = \frac{r^{max}_{LDH} \left(K_{m,pyr} K_{m,nadh}\right)^{-1} \left(C_{pyr} C_{nadh} - \frac{C_{lactate} C_{nad}}{K_{eq}}\right)}{\left(1 + \frac{C_{pyr}}{K_{m,pyr}} + \frac{C_{lactate}}{K_{m,lactate}}\right)\left(1 + \frac{C_{nadh}}{K_{m,nadh}} + \frac{C_{nad}}{K_{m,nad}}\right)} \quad (17)$$

$$r_{SucSynthesis} = \frac{r^{max}_{SucSynthesis} C_{pep}}{K_{m,pep} + C_{pep}} \quad (18)$$

2.2. Optimization model description

Metabolic network design is formulated as an optimization problem based on a kinetic model to maximize the production of a valuable product. Maximum reaction rates (r_j^{max}) are considered as degrees of freedom, since they are associated to the corresponding enzyme concentration (Eqn. (19)) and can be tuned to maximize or minimize the production of a given metabolite.

$$r_j^{max} = k_{cat,j} e_j \quad (19)$$

where $k_{cat,j}$ is the catalytic constant and e_j is the corresponding enzyme concentration.

As Eqn. (19) indicates, changes in r_j^{max} imply that enzyme concentration or their catalytic activity will be modified, so the total enzyme activity will vary. Increase in proteins level could affect the diffusion of metabolites, as a consequence of the increase of cytoplasm viscosity; also inclusion bodies could appear due to protein precipitation [21]. On the other hand, increases in expression of enzymes imply larger amino acids synthesis rates and energy consumption, which could cause culture stress [10, 21, 22]. Kinetic models for metabolic networks can describe metabolic regulation of enzyme activity but they are not capable to describe the changes mentioned above, so it becomes

necessary to include additional constraints that ensure total enzyme activity level be maintained close to the initial steady state. To account for this condition Eqns. (20) and (21) are included [13],

$$\frac{1}{M}\sum_{j=1}^{M}\frac{r_j^{max}}{r_j^{max,0}} = 1 \tag{20}$$

where M is the number of enzymes involved in the metabolic network and $r_j^{max,0}$ is the maximum reaction rate at the reference steady state. Eqn. (20) imposes that an increase in certain enzyme level is compensated by a decrease in the activity of the remaining enzymes in order to keep the total enzyme activity [22, 23]. To account for changes in all enzymes when only L (L<M) are modulated; Eqn. (21) is included.

$$\frac{r_{j_1'}^{max}}{r_{j_1'}^{max,0}} = \cdots = \frac{r_{j_K'}^{max}}{r_{j_K'}^{max,0}} = \gamma \tag{21}$$

where $j_1' \ldots j_K'$ are the indices of non-modulated enzymes, K= M-L and L is the number of modulated enzymes (corresponding to maximum reaction rates being optimization variables). Equations (20) and (21) can be re-written as (22) and (23), respectively, to be included in the optimization problem formulation.

$$\frac{r_{j_1'}^{max}}{r_{j_1'}^{max,0}} + \cdots + \frac{r_{j_L'}^{max}}{r_{j_L'}^{max,0}} + K\gamma = M \tag{22}$$

$$r_{j_s}^{max} = \gamma r_{j_s}^{max,0} \qquad s = 1, \ldots, K \tag{23}$$

where $j_1' \ldots j_L'$ are the indices of modulated enzymes. Eqn. (23) is valid for non-modulated enzymes. On the other hand, as cells maintain homeostasis, if large changes in metabolite concentrations occur in a new steady state condition, it is expected that gene expression will be modified to return to the reference steady state. Also, an increase in metabolite concentration could inhibit enzyme activity [10, 7]. To account for this condition, Eqn. (24) is included in the optimization problem.

$$\frac{1}{N}\sum_{i=1}^{N}\frac{|C_i - C_i^0|}{C_i^0} \leq \delta \tag{24}$$

where δ corresponds to the allowed maximum fraction for changes in intracellular metabolites concentrations, C_i^0 is the concentration of metabolite i at the reference steady state and N is the number of intracellular metabolites.

In this work, we formulate an optimal design problem based on the extended metabolic network model for *Escherichia coli* described in Section 2.1, for the maximization of serine and ethanol, as two different case studies. Right hand side of dynamic mass balances, described by Eqns. (3) to (11) and Eqns. (A1) to (A17) of Supporting information, are written equal to zero in the design problem. Finally, the optimal design problem is formulated by Eqn. (25), as follows:

$$\max_{r_{j_1}^{max} \dots r_{j_L}^{max}} \quad r_m = r_m^{max} f(\mathbf{c}, \mathbf{q}) \quad m = SerSynth \text{ or } ADH$$

$$s.t.$$

$$D\left(C_i^{feed} - C_i^{extracellular}\right) + v_{ij} r_j \left(\frac{C_x}{\rho_x}\right) = 0 \qquad a\ i)$$

$$\sum_l v_{kl} r_l - \mu C_k = 0 \qquad a\ ii) \qquad (25)$$

$$r_j = r_j^{max} f(\mathbf{c}, \mathbf{q}) \qquad j = 1, \dots, M \qquad b)$$

$$\frac{r_{j_1'}^{max}}{r_{j_1'}^{max,0}} + \cdots + \frac{r_{j_L'}^{max}}{r_{j_L'}^{max,0}} + K\gamma = M \qquad c)$$

$$r_{j_s}^{max} = \gamma r_{j_s}^{max,0} \qquad s = 1, \dots, K \qquad d)$$

$$\frac{1}{N} \sum_{i=1}^{N} \frac{|C_i - C_i^0|}{C_i^0} \leq \delta \qquad e)$$

where **c** and **q** are concentrations and model parameters vectors, respectively; constraints 25 a i) and a ii) correspond to RHS of mass balances; constraints (25 b) correspond to kinetic rate equations; and constraints 25 c) to e) correspond to Eqns. (22) to (24).

2.3. Optimization under stability constraints

Eigenvalue analysis is usually carried out to assess asymptotic stability of dynamic systems. In this section, we give a brief description of the eigenvalue optimization approach used in this work to achieve the stable design of metabolic networks,. A detailed description of this methodology is presented in [9].

For asymptotically stability, the eigenvalues of the dynamic system Jacobian (**A**) must lie on the left half of the complex plane. By formulation of an optimization problem, the real parts of the eigenvalues of the Jacobian matrix are forced to be strictly negative, ensuring, in that way, asymptotic stability of the resulting equilibrium point. For matrices with dimensions larger than four, it is not possible to obtain mathematical expressions for their eigenvalues, which prevents their direct inclusion as constraints on large scale optimization problems. For a general treatment of the design problem under stability conditions, it is necessary to develop alternative formulations that are equivalent to the condition on the real part of the eigenvalues of matrix **A** be negative. Blanco and Bandoni [9] apply concepts from Lyapunov theory. When dealing with real asymmetric matrices, like matrix **A**, with complex eigenvalues, it is possible to translate the constraint on nonpositivity on real part of the eigenvalues of **A** into a positive definiteness condition on a real symmetric matrix P, defined through Lyapunov's matrix identity as follows:

$$\mathbf{A}(\mathbf{c})^T\mathbf{P} + \mathbf{P}\mathbf{A}(\mathbf{c}) + \mathbf{I} = 0 \quad (26)$$

where **P** is a symmetric and positive definite matrix and **I** is the Identity.
Positive definiteness condition for matrix P is derived from Sylvester criterion, which states that the necessary and sufficient conditions for a symmetric matrix to be positive definite, are that the determinants of its successive principal minors be positive. Sylvester criterion can be applied on the inverse of matrix **P** ($\mathbf{P^{-1}}$), to obtain a numerically better-posed constraint [9].
The optimal design problem for the metabolic network of *Escherichia coli* that ensures asymptotical stability of the steady state is, in this way, formulated including Eqn. (26) and the Sylvester criterion on $\mathbf{P^{-1}}$, as additional constraints to the formulation represented by Eqn. (25), as follows:

$$\max_{r_{j_1}^{max} \dots r_{j_L}^{max}} \quad r_m = r_m^{max} f(\mathbf{c}, \mathbf{q}) \quad m = SerSynth\ or\ ADH$$

$$s.t.$$

$$\begin{aligned}
& D\left(C_i^{feed} - C_i^{extracellular}\right) + v_{ij} r_j \left(\frac{C_x}{\rho_x}\right) = 0 && a\ i) \\
& \sum_l v_{kl} r_l - \mu C_k = 0 && a\ ii) \\
& r_j = r_j^{max} f(\mathbf{c}, \mathbf{q}) \qquad j = 1, \dots, M && b) \\
& \frac{r_{j_1'}^{max}}{r_{j_1'}^{max,0}} + \cdots + \frac{r_{j_L'}^{max}}{r_{j_L'}^{max,0}} + K\gamma = M && c) \\
& r_{j_s}^{max} = \gamma r_{j_s}^{max,0} \qquad s = 1, \dots, K && d) \\
& \frac{1}{N}\sum_{i=1}^{N} \frac{|C_i - C_i^0|}{C_i^0} \leq \delta && e) \\
& \mathbf{A}(\mathbf{c})^T\mathbf{P} + \mathbf{P}\mathbf{A}(\mathbf{c}) + \mathbf{I} = 0 && f) \\
& \det(\mathbf{P}_i^{-1}) > \xi && g)
\end{aligned} \quad (27)$$

where **c** and **q** are concentrations and model parameters vectors, respectively, **A** is the RHS of the dynamic system Jacobian matrix, constraint 27 f) corresponds to Lyapunov equation, $\det(\mathbf{P}_i^{-1})$ stands for determinants of the principal minors of $\mathbf{P^{-1}}$ and ξ is a user defined positive constant. Values of ξ allow to modify, in an indirect way, the spectrum of the Jacobian and, consequently, the dynamic evolution of the system towards the equilibrium. Constraints 27 a) to e) correspond to the formulation represented by Eqn. (25).
We formulate the optimal design problem under stability constraints (Eqn. (27)), based on eigenvalue optimization approach, for the extended metabolic network model for *Escherichia coli* described in Section 2.1.

3. Results and Discussion

We propose two case studies: the maximization of serine production, which is an aminoacid with great importance in different industries (Case 1); and the maximization of ethanol production (Case 2). In both cases we solve the design problem without stability constraints (Eqn. (25)) and under stability constraints to ensure a stable optimal network design (Eqn. (27)). Nonlinear optimization problems are solved with reduced space Successive Quadratic Programming techniques, with program IPOPT ([15]).

3.1. Case 1: Maximization of Serine production

Based on previous work on global sensitivity analysis on the dynamic metabolic network to main parameters ([24]), we have selected three maximum reaction rates as design variables (L=3), corresponding to Serine synthesis rate ($r^{max}_{SerSynth}$), glucose-6-phosphate isomerase rate (r^{max}_{PGI}), and Chorismate and Murein synthesis rate (r^{max}_{Synth1}), respectively. The design problem to maximize serine production rate ($r_{SerSynth}$), for the kinetic model of the Embden-Meyerhof-Parnas pathway, the pentose phosphate pathway and the phosphotransferase system comprises mass balances for eighteen metabolites (Eqns. (A1) to (A17) and Eqn. (3), thirty rate equations (Eqns. (A18) to (A47)) and twenty nine additional constraints standing for Eqns. (22), (23) and (24).
When formulating the design problem under stability constraints, Eqn. (27), additional equality constraints (eighteen) and inequalities (eighteen) are included in the optimization problem formulation, standing for Lyapunov's equation and nonnegativity on the determinants of the principal minors of $\mathbf{P^{-1}}$, to ensure its positive definiteness, respectively.

Numerical results for the metabolic network design to achieve maximization of Serine production are shown in Figures (2) and (3). For both formulations, without and with stability constraints Eqns. (25) and (27) δ=0.3 is considered as the upper bound for the homeostasis condition (Eqn. (24)). Table A4 of Supporting information shows values for metabolites concentration and enzyme reaction fluxes at the reference steady state. Values for the optimization variables and the objective function at the reference steady state and at the steady state achieved in the design problem are presented in Table 1.

Numerical results show that serine production can be increased from 0.014 mM/sec in a reference steady state (experimental) to 0.094 mM/sec when the maximum Serine synthesis reaction rate ($r^{max}_{SerSynth}$) is at its upper bound (=10* $r^{max,0}_{SerSynth}$) (Table 1). The eigenvalues of matrix **A**(**c**) show that the metabolic network is stable, being the real part of the largest eigenvalue -3.7E-4. While stable, the system is close to critical stability and small changes in some parameters could easily lead to an unstable equilibrium. Taking into account the difficulty to modulate enzyme levels with precision, this implies that a relative deviation on the enzyme levels from their desired value may produce unstable behavior in the system. These results are in agreement with [6] who determined that the metabolic network of *Escherichia coli* becomes instable when $r^{max}_{SerSynth}$> 9.8 $r^{max,0}_{SerSynth}$. Therefore, the inclusion of stability constraints in the design problem becomes necessary.

In order to modify the spectrum of matrix **A**(**c**), the optimization problem has been solved for different values of parameter ξ. For ξ= 1E-4, the largest eigenvalue real part is -1.5E-3. In this case,

maximum serine production can be increased 330% from its reference steady state value (0.014 mM/sec) to 0.059 mM/sec. It can be noted that some robustness of the network regarding dynamic stability was achieved at the expense of a decrease in the objective function. Furthermore, serine synthesis maximum reaction rate ($r^{max}_{SerSynth}$) is not at its upper bound (Table 1). In both cases (with and without stability constraints), optimal values for the maximization of serine production implies $r^{max}_{Synth1} = 0$. This corresponds to the deletion of Chorismate and Murein synthesis pathway.

Figure 2 shows the ratio of change with respect to the reference steady state for each metabolite in the metabolic network for the optimal design with and without stability constraints, respectively. The largest increments in metabolites concentration are obtained for g6p, f6p and g1p, being 49% for the optimal design and 23% for the stable optimal design (Figure 2).

Figure 3 shows the optimal flux distribution in the metabolic network of *Escherichia coli* at the reference steady state and at the steady state achieved for the stable optimal design. In the latter, it is observed that serine production rate ($r_{SerSynth}$) increases 6.8 and 4.3 times with respect to the reference steady state for the optimal design and the stable optimal design, respectively. From Figure 3 it can be observed that maximization of Serine production implies an increase in the glycolytic enzymes fluxes and in the pentose phosphate pathways (PPP). However for the last one it is observed that aromatic amino acids and nucleotides synthesis rates are not significantly modified with respect to their reference value. This implies that most of the carbon flux through PPP is derived to glycolysis due to transketolase b (TKb) enzyme activity (Figure 1), contributing to serine synthesis.

3.2. Case 2: Maximization of Ethanol production

The design problem to maximize ethanol production (r_{ADH}), for the steady state kinetic model of the Embden-Meyerhof-Parnas pathway, the pentose phosphate pathway, the phosphotransferase system and fermentation reactions comprises mass balances for twenty five metabolites (Eqns. (A1) to (A17) and Eqns. (3) to (11)), thirty eight kinetic rate equations (Eqns. (A18) to (A47) and Eqns. (12) to (18)) and thirty six additional constraints standing for Eqns. (22), (23) and (24). Maximum reaction rates of phosphofructokinase (r^{max}_{PFK}), pyruvate kinase (r^{max}_{PK}), alcohol dehydrogenase (r^{max}_{ADH}) and Chorismate and Mureine synthesis (r^{max}_{Synth1}), are considered as optimization variables for the design problem.

When formulating the design problem under stability constraints, Eqn. (27), additional equality constraints (twenty five) and inequalities (twenty five) are included in the optimization problem formulation, standing for Lyapunov's equation and nonnegativity on the determinants of the principal minors of $\mathbf{P^{-1}}$, to ensure its positive definiteness, respectively.

Numerical results for the metabolic network design when ethanol production is maximized are shown in Figures 4 and 5. For both formulations (Eqns. (25) and (27)) δ=0.3 is considered as the upper bound for the homeostasis condition (Eqn. (24)). Values for optimization variables and objective function for both optimal designs are presented in Table 2. Numerical results show that an increment of 38 % with respect to the reference steady state can be reached for ethanol production rate in the optimal design (Table 2). The eigenvalues of matrix **A(c)** show that the designed metabolic network is unstable, as

the real parts of some eigenvalues are positive. Thus, the addition of stability constraints on the design problem is required.

For the optimal design under stability constraints it is observed that, for ξ= 1E-5, the largest eigenvalue real part is -2.1E-3. In this case ethanol production rate is increased from 0.005 mM/sec to 0.006 mM/sec (Table 2). Therefore, stability of the new steady state is achieved with a 14% decrease in the objective function.

Figure 4 shows that concentrations of the glycolytic intermediates for the stable metabolic network are higher than for the reference steady state, being the largest increase in fructose-1,6-diphosphate (fdp) concentration.

Numerical results for the stable optimal design show that carbon fluxes through glycolysis and ethanol production are larger as compared to their reference steady state value. In particular, the reaction rate of pyruvate kinase presents the largest increment (Figure 5). This fact can be attributed to the combined effect of the increase in the maximum reaction rate of this enzyme (Table 2) and fdp concentration, which is an activator of pyruvate kinase (Figure 5).

Results also confirm that, even amplifying glycolytic enzyme fluxes and those involved in ethanol production (ACALDH and ADH) in *Escherichia coli*, there is no significant increase in the synthesis of ethanol, which justifies the insertion of heterologous pathways.

4. Concluding remarks

The proposed approach in the formulation of a design problem for a detailed kinetic model for a metabolic network including stability constraints, allows the determination of an improved network that maximizes a biotechnological objective (serine or ethanol production) by ensuring its stability at the design level.

Numerical results indicate that it is necessary to include stability constraints in the design problem since, otherwise, unstable steady states can be achieved and, as a consequence, the designed metabolic network can be vulnerable to external disturbances.

Finally, we can conclude that the approach of optimal design under stability constraints proposed in this work is an efficient and reliable tool for the design of large and stable metabolic networks.

Acknowledgment

The authors gratefully acknowledge financial support from the National Research Council, Universidad Nacional del Sur, Universidad Nacional de Tucumán and Ministry of Science, Technology and Productive Innovation, Argentina.

5. References

[1] Diaz Ricci, J. C.. Influence of phosphoenolpyruvate on the dynamic behavior of phosphofructokinase of *Escherichia coli*. Journal of Theoretical Biology, 1996, 178, 145–150.

[2] Hatzimanikatis, V., Bailey, J., Studies on glycolysis.1.Multiple steady states in bacterial glycolysis, Chemical Engineering Science, 1997, 52, 2579-2588.

[3] Haddad, W., Chellaboina, V., Stability and dissipativity theory for nonnegative dynamical systems: a unified analysis framework for biological and physiological systems, Nonlinear Analysis-Real World Applications, 2005, 6, 33-65.

[**4]** Ivanov, O., Schaft, A., Weissing, F., Steady states and stability in metabolic networks without regulation. Journal of Theoretical Biology, 2016, 401, 78–93.

[5] Theisen MK, Lafontaine Rivera JG, Liao JC., Stability of Ensemble Models Predicts Productivity of Enzymatic Systems. PLoS Comput Biol, 2016, 12(3): e1004800. doi:10.1371/journal.pcbi.1004800.

[6] Vital-Lopez, F., Maranas, C. y Armaou, A., Bifurcation analysis of the metabolism of E. coli at optimal enzyme levels. Proceedings of the 2006 American Control Conference, Minnneapolis, Minnesota, USA.

[7] Nikolaev, E., Pharkya, P., Maranas, C., Armaou, A., Optimal selection of enzyme levels using large-scale kinetic models, Proceedings of 16th I.F.A.C. World Congress, 2005, Prague, Czech Republic.

[8] Chang, Y., Sahinidis, N. Optimization of metabolic pathways under stability considerations. Computers & Chemical Engineering, 2005, 29, 467-479.

[9] Blanco, A.M., Bandoni, J.A., Eigenvalue Optimization Based Formulations for Dynamics and Control Problems, Chemical Engineering Process, 2007, 46, 1192-1199.

[10] Visser, D., Schmid, J., Mauch, K., Reuss, M., Heijnen, J., Optimal re-design of primary metabolism in *Escherichia coli* using linlog kinetics. Metabolic Engineering, 2004, 6, 378-390.

[11] Gerhard, J., Mönnigmann, M., Marquardt, W., Steady state optimization with guaranteed stability of a tryptophan biosynthesis model. Computers & Chemical Engineering, 2008, 32, 2914-2919.

[12] Mönnigmann, M., Marquardt, W., Steady state process optimization with guaranteed robust stability and robust feasibility. American Institute Of Chemical Engineers Journal, 2003, 49, 3110-3126.

[13] Nikolaev, E., The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems. Metabolic Engineering, 2010, 12 (1), 26-38.

[14] Di Maggio, J., Paulo, C., Estrada, V., Perotti, N., Diaz Ricci, J., Diaz, M.S., Parameter estimation in kinetic models for large scale biotechnological systems with advanced mathematical programming techniques, Biochemical Engineering Journal, 2014, 83, 104-115.

[15] Waechter, A., Biegler, L., On the Implementation of an Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming, Mathematical Programming, 2006, 106, 1, 25-57.

[16] Chassagnole, C., Noisommit-Rizzi, N., Schmid, J., Mauch, K., Reuss, M., Dynamic Modeling of the Central Carbon Metabolism of *Escherichia coli*, Biotechnology and Bioengineering, 2002, 79, 1, 53-73.

[17] Stephanopoulos, G., Aristidou, A. & Nielsen, J. (1998). Metabolic engineering: Principles and methodology. Elsevier.

[18] Hoefnagel, M., Starrenburg, M., Martens, D., Hugenholtz, J., Kleerebezem, M., Van Swam, I., Bongers, R., Westerhoff, H. & Snoep, J., Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. Microbiology, 2002, 148, 1003-1013.

[19] Henkin, J., Abeles, R., Evidence against an Acyl-Enzyme Intermediate in the Reaction Catalyzed by Clostridial Phosphotransacetylase. Biochemistry, 1976. 15 (16), 3472-3479.

[20] Knappe, J., Blaschkowski, H., Gröbner, P., Schmitt, T., Pyruvate Formate-Lyase of *Escherichia coli*: the Acetyl-Enzyme Intermediate. European Journal of Biochemistry, 1974, 50 (1), 253-263.

[21] Fredrickson, A. G. Formulation of structured growth models. Biotechnology and Bioengineering, 1976, 18, 1481-1486.

[22] Mauch, K., Buziol, S., Schmid, J., Reuss, M.. Computer aided design of metabolic networks. In: AIChE Symposium Series. Chemical Process Control-6 Conference, Tucson, Arizona, 2001.

[23] Bakker B.M., Hans V. Westerhoff and Paul A. M. Michels. Regulation and Control of Compartmentalized Glycolysis in Bloodstream Form Trypanosoma brucei. Journal of Bioenergetics and Biomembranes, Vol. 27, No. 5, 1995.

[24] Di Maggio, J., Diaz Ricci, J. C., Diaz, M. S., Global Sensitivity Analysis in dynamic metabolic networks. Computers & Chemical Engineering, 2010, 34 (5), 770-781.

Tables

Table 1. Optimization variables and objective function values for maximization of Serine production (Case 1)

Optimization variables (mM/sec)	Reference steady state	Optimal design	Optimal design under stability constraints
r_{PGI}^{max}	495.870	442.316	495.842
$r_{SerSynth}^{max}$	0.020	0.200	0.099
r_{Synyh1}^{max}	0.015	0.000	0.000
$r_{SerSynth}$	0.014	0.094	0.059

Table 2. Optimization variables and objective function values for maximization of ethanol production (Case 2)

Optimization variables (mM/sec)	Reference steady state	Optimal design	Optimal design under stability constraints
r_{PGI}^{max}	0.12	0.177	0.155
r_{PK}^{max}	0.047	0.278	0.117
r_{Synth1}^{max}	0.015	0.023	0.001
r_{ADH}^{max}	0.5	3.934	1.316
r_{ADH}	0.005	0.007	0.006

Figure legends

Figure 1. Central carbon metabolism and fermentation pathways of *Escherichia coli*

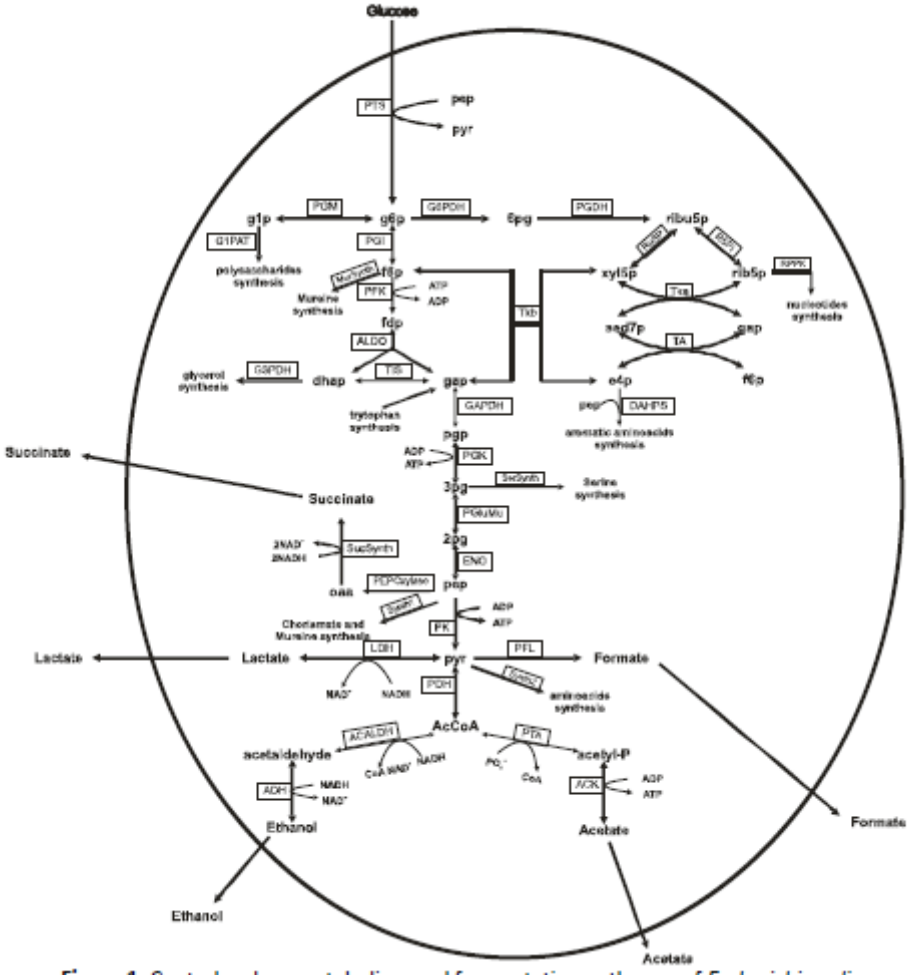

Figure 2. Optimal concentrations for metabolic network designs with and without stability constraints for maximization of Serine production.

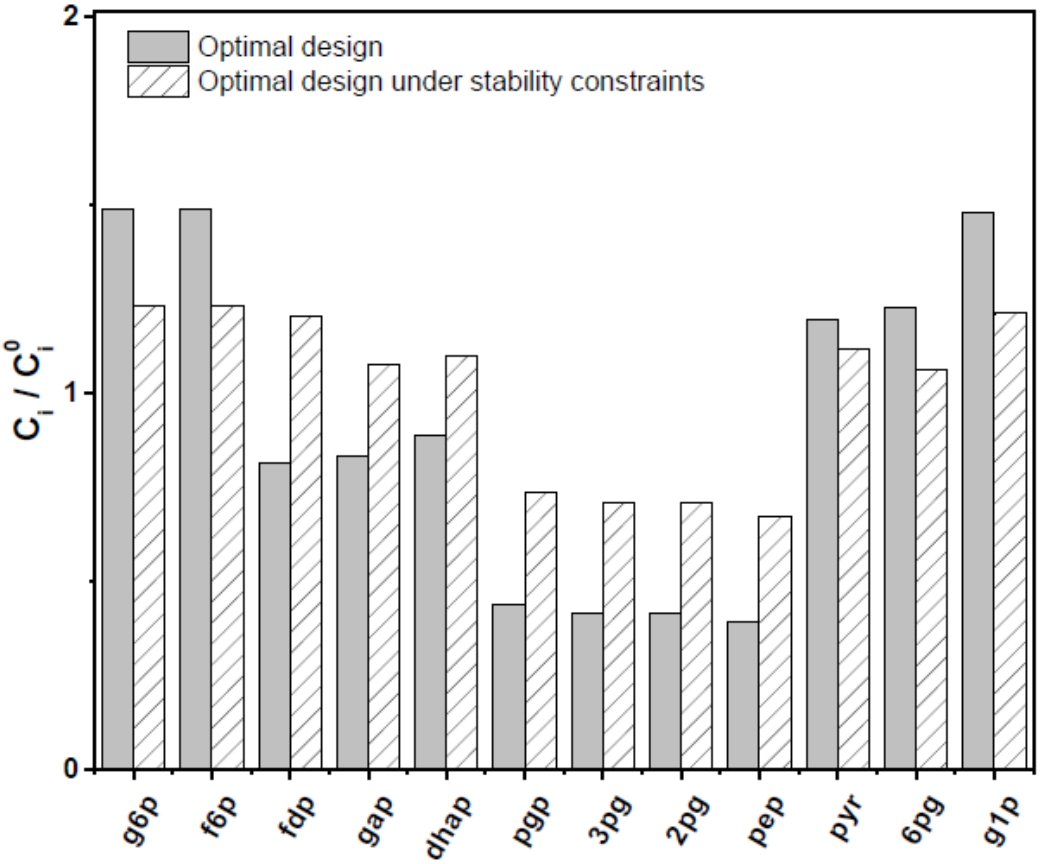

Figure 3. Optimal fluxes distribution for Serine production maximization at the reference steady state (⊡) and at the stable steady state.

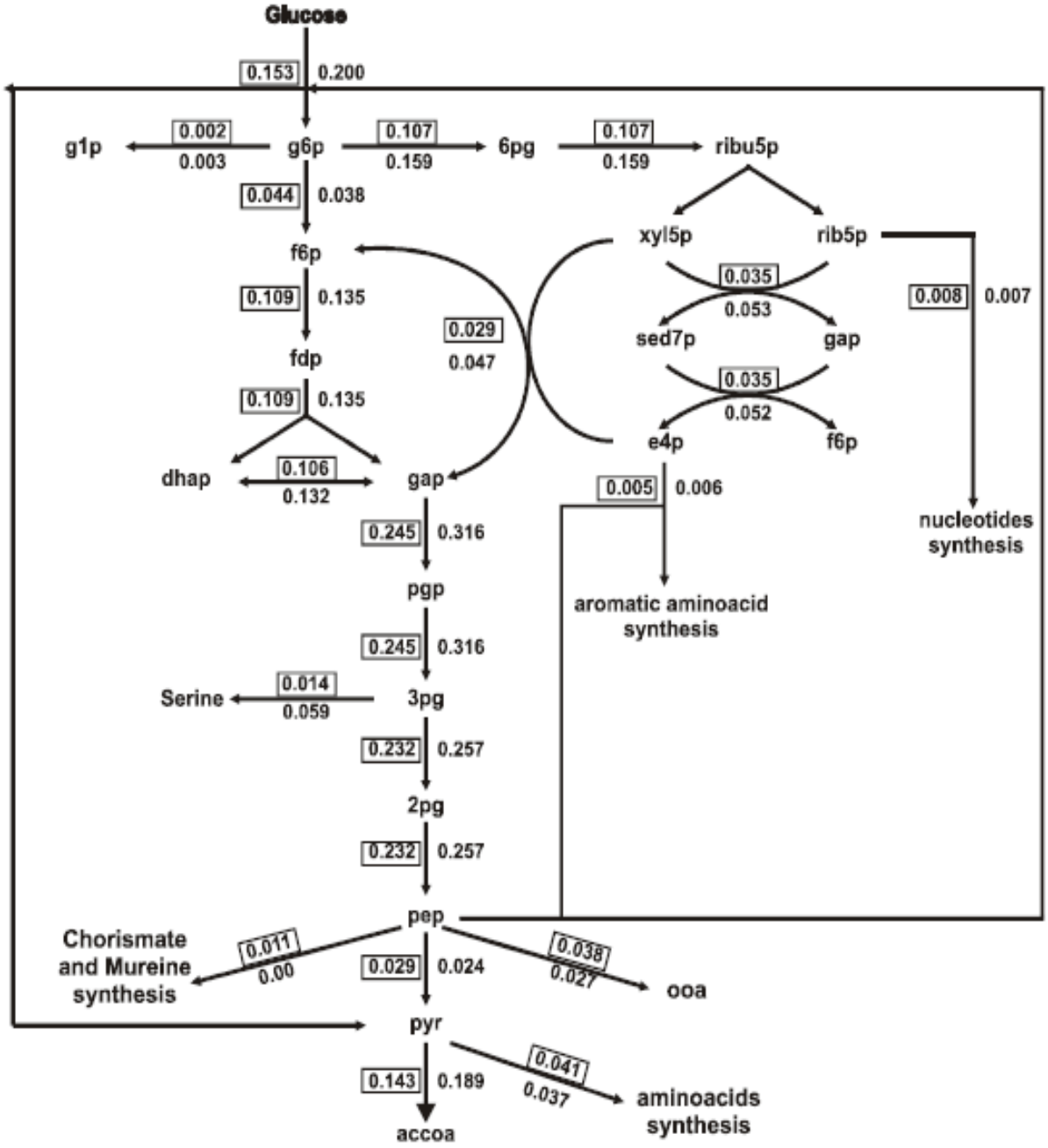

Figure 4. Optimal concentrations for metabolic network designs with and without stability constraints for maximization of ethanol production.

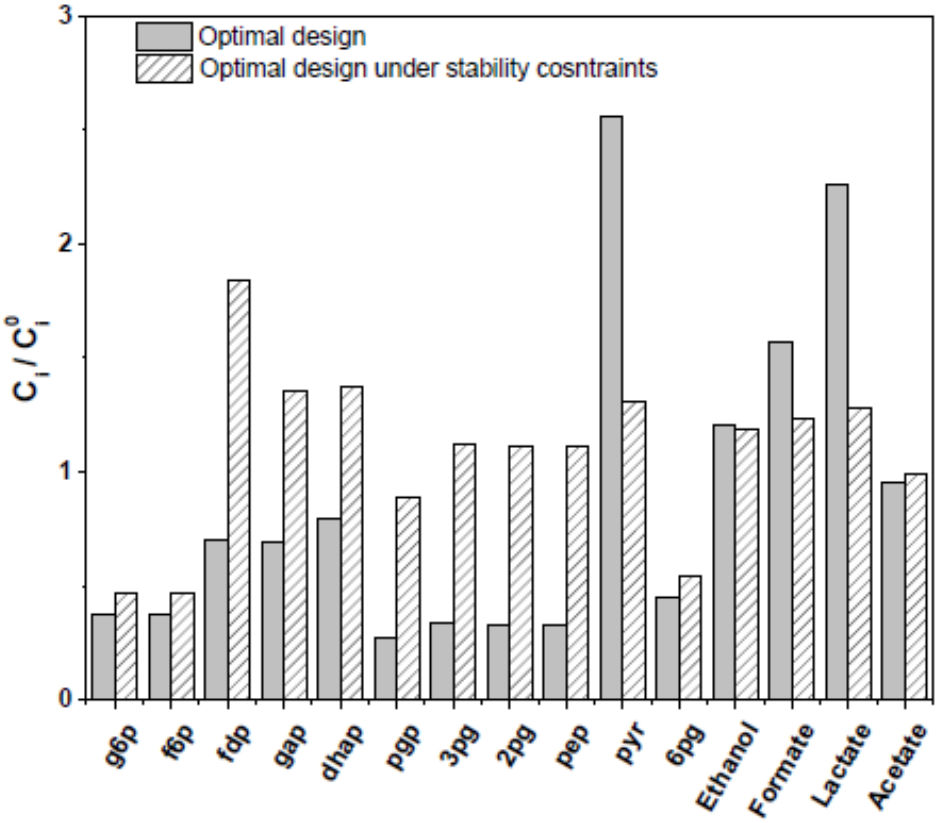

Figure 5. Optimal fluxes distribution for ethanol production maximization at the reference steady state (⊡) and at the stable steady state.