Provided for non-commercial research and education use. Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Journal of Biotechnology 149 (2010) 141-153

Contents lists available at ScienceDirect

ELSEVIER

Journal of Biotechnology



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

Optimization and evolution in metabolic pathways: Global optimization techniques in Generalized Mass Action models

Albert Sorribas^{a,*}, Carlos Pozo^b, Ester Vilaprinyo^c, Gonzalo Guillén-Gosálbez^b, Laureano Jiménez^b, Rui Alves^a

^a Departament de Ciències Mèdiques Bàsiques, Institut de Recerca Biomèdica de Lleida (IRBLLEIDA), University of Lleida, Montserrat Roig, 2, 25008 Lleida, Spain

^b Departament d'Enginyeria Química, Universitat Rovira i Virgili, Tarragona, Spain

^c Fundation Dr. Ferran, Hospital Verge de la Cinta, Tortosa, Spain

ARTICLE INFO

Article history: Received 15 September 2009 Received in revised form 23 January 2010 Accepted 29 January 2010

Keywords: Optimization Power-law formalism Evolution Design principles Operation principles

ABSTRACT

Cells are natural factories that can adapt to changes in external conditions. Their adaptive responses to specific stress situations are a result of evolution. In theory, many alternative sets of coordinated changes in the activity of the enzymes of each pathway could allow for an appropriate adaptive readjustment of metabolism in response to stress. However, experimental and theoretical observations show that actual responses to specific changes follow fairly well defined patterns that suggest an evolutionary optimization of that response. Thus, it is important to identify functional effectiveness criteria that may explain why certain patterns of change in cellular components and activities during adaptive response have been preferably maintained over evolutionary time. Those functional effectiveness criteria define sets of physiological requirements that constrain the possible adaptive changes and lead to different operation principles that could explain the observed response. Understanding such operation principles can also facilitate biotechnological and metabolic engineering applications. Thus, developing methods that enable the analysis of cellular responses from the perspective of identifying operation principles may have strong theoretical and practical implications. In this paper we present one such method that was designed based on nonlinear global optimization techniques. Our methodology can be used with a special class of nonlinear kinetic models known as GMA models and it allows for a systematic characterization of the physiological requirements that may underlie the evolution of adaptive strategies.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Cells are natural factories that can adapt to changes in external conditions (Causton et al., 2001; Gasch et al., 2000; Mitchell et al., 2009). Their adaptive responses are a result of evolution through different mechanisms that include random mutation, gene duplication, gene transfer, etc. (Koonin, 2009). During steady-state growth conditions, the cell works within normal operating ranges that are characterized by fluxes and metabolite levels moving within more or less narrow ranges (Watson, 1970; Wiebe et al., 2008).

As the conditions in the medium change, the operating range of cells may also change. If environmental changes are spurious, there are internal control mechanisms that play a fundamental role in maintaining the operating range of cells about its initial value. However, when the environmental changes are relevant or sustained, an adaptive response is mounted by the cells. Such adaptive responses occur during heat shock, oxidative stress, or other stresses. If those situations are prevalent in the evolutionary history of the cell, specific behaviors and mechanisms that facilitate cell adaptation through changes in gene expression and protein activity and assure cell viability are selected for. Such behaviors lead to a fine tuning of metabolic fluxes and concentrations (Vilaprinyo et al., 2006). The specificity of the adaptive response mounted by each cell type in response to a given stress depends both on the challenges it responds to and on the evolutionary history of the cell or organism (Bedford and Hartl, 2009; Kashiwagi et al., 2006; Teusink et al., 2009; Wilkins, 2007).

For example, the heat shock caused by a sudden rise in the temperature of the growing media triggers an ordered response in yeast that causes an arrest in cell cycle and specific changes in the coordinated activity of several metabolic pathways (Trotter et al., 2001). These changes help the cell to synthesize protective molecules that permit its adaptation and survival (Causton et al., 2001; Eisen et al., 1998; Gasch et al., 2000; Jenkins, 2003). In principle, many alternative sets of coordinated changes in the activity of pathways could allow for an appropriate adaptive readjustment of metabolism. However, experimental and theoretical measurements of the actual responses show that these follow fairly

^{*} Corresponding author. Tel.: +34 608533249; fax: +973702426. *E-mail address*: albert.sorribas@cmb.udl.cat (A. Sorribas).

^{0168-1656/\$ –} see front matter $\ensuremath{\mathbb{O}}$ 2010 Elsevier B.V. All rights reserved. doi:10.1016/j.jbiotec.2010.01.026

well defined patterns that are consistent with an evolutionary optimization of this response with respect to different physiological and functional effectiveness criteria (El-Samad et al., 2005; Kurata et al., 2006; Molina-Navarro et al., 2008; Vilaprinyo et al., 2006). Thus, it is important to identify functional effectiveness criteria that may explain why certain patterns of change in cellular components and activities during adaptive response have been preferably maintained over evolutionary time (Coelho et al., 2009; Han, 2008; Salvador and Savageau, 2003, 2006; Savageau, 1971, 1974a,b, 1976; Savageau et al., 2009). Such criteria are necessarily derived from the analysis of systemic properties that emerge from the integrated molecular behavior of the cellular components, and they may include robustness, dynamic stability, minimization of intermediates, minimization of biosynthetic cost, temporal responsiveness, etc. (Chang and Sahinidis, 2005; Coelho et al., 2009; Salvador and Savageau, 2003, 2006; Savageau et al., 2009). The functional effectiveness criteria for a response define sets of physiological constraints that shape that response and lead to different operation principles that could explain why the cells adapt in a certain way at the molecular level (Bedford and Hartl, 2009; Braunstein et al., 2008; Vilaprinyo et al., 2006; Voit and Radivoyevitch, 2000; Voit, 2003).

Although the operational principles of cellular responses are a result of evolution, they can be applied to and validated in biotechnological applications. Metabolic engineering manipulates naturally evolved organisms in order to obtain increased amount of new products (Bailey et al., 1990, 1996; Bailey, 1991, 1999, 2001; Hatzimanikatis and Liao, 2002). This manipulation often involves a process of optimization that searches for the best modified strain with respect to the initial optimization criteria (Goodman, 2008). Thus, developing methods that permit analysis of cellular responses from the perspective of identifying operational principles may have strong theoretical and practical implications. Often, this goal can only be achieved through methods that involved the creation, analysis and comparison of mathematical models for the processes and responses one is interested in studying (Alvarez-Vasquez et al., 2004; Klipp et al., 2005; Sims et al., 2004; Voit, 2003).

In this work, we discuss and extend a method that can be used to identify and study the operation principles of cellular response at the molecular level, by characterizing feasibility regions for those responses (Guillén-Gosálbez and Sorribas, 2009; Pozo et al., submitted for publication). Such feasibility regions encompass all possible ranges in enzyme activity that allow for an appropriate response by the cell after an environmental challenge. This method may help in both, understanding the evolution of such responses and guiding manipulations of gene expression in metabolic engineering applications. The proposed method for identifying feasibility regions uses a recently developed global optimization method (Guillén-Gosálbez and Sorribas, 2009; Pozo et al., submitted for publication). Here, the capabilities of that optimization method are enhanced through an iterative and systematic search strategy that identifies all the parameter regions containing admissible solutions that are compatible with the considered physiological constraints. The general framework presented here has the potential for solving problems of great interest in systems biology studies. As an example we analyze a mathematical model created to represent the heat shock response of the yeast Saccharomyces cerevisiae.

2. Methods

2.1. Generalized Mass Action models

Generalized Mass Action (GMA) models are a special class of models defined within the general framework of Biochemical Systems Theory (BST) (Voit, 2000). These models use the power-law formalism to obtain a representation of the different processes involved in the target system. For a network with p processes (enzyme reactions, transport systems, etc.), n internal metabolites, and m external parameters or independent variables, a GMA model is defined as follows:

$$\frac{dX_i}{dt} = \sum_{r=1}^p \mu_{ir} v_r = \sum_{r=1}^p \mu_{ir} \left(\gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}} \right) \quad i = 1, \dots, n$$
(1)

In Eq. (1), the parameters μ_{ir} account for the stoichiometry of the process, i.e. the number of molecules of X_i produced by or used in reaction v_r (for instance +1, +2 for production, or -1, -2, etc., for degradation). The parameters of the power-law representation of each reaction are the apparent rate-constant γ_r and the kinetic-order f_{ir} , defined as (Savageau, 1969a,b, 1976):

$$f_{ir} = \left(\frac{\partial \nu_r}{\partial X_i}\right)_0 \frac{X_{i0}}{\nu_{r0}} \tag{2}$$

The subscript 0 stands for the operating point where the powerlaw representation is derived. Appropriate parameter values for a given system can be estimated using different procedures. As this is a broad subject, the reader is referred to the recent review by Chou and Voit (2009). In the following, we shall assume that a parameter set has been obtained and that the GMA model can be used for characterizing the properties of the system.

GMA representations integrate information about network stoichiometry and regulation (kinetic-orders) into a dynamic mathematical model. These models can be used for computing both the transient and steady-state responses of metabolites and fluxes to changes in the environment of the model. Due to their structure and to the available methods, GMA models are well suited for evaluating parameter sensitivities and for developing optimization techniques (Chang and Sahinidis, 2005; Marin-Sanguino et al., 2007; Polisetty et al., 2008; Torres et al., 1996, 1997; Voit, 1992). Thus, that representation is especially useful as a framework for systems biology applications and provides a description of processes that is more accurate than the one provided by other techniques based on the stoichiometric matrix alone, such as Flux Balance Analysis (FBA) (Lee et al., 2006). This added accurateness comes at the price of needing more information to estimate parameter values for GMA models.

2.2. Characterization of the effect of changes in enzyme activities

Given a GMA model, changes in enzyme activities can be implemented by changing the value of the rate-constant for the processes in which the enzymes are involved.¹ For simplicity, we can write

$$\frac{dX_i}{dt} = \sum_{r=1}^p \mu_{ir} \nu_r = \sum_{r=1}^p \mu_{ir} \left(k_r \gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}} \right) \quad i = 1, \dots, n$$
(3)

where k_r indicates the change-fold over the original enzyme activity (which is γ_r). Thus, in the reference state, $k_r = 1$. Accordingly, a vector (k_1, k_2, \ldots, k_p) would correspond to a specific pattern of fold changes in enzyme activities. For this vector, the change in the

¹ Enzyme activities can be explicitly included in the model as independent variables. However, for constant levels of enzyme activity, doing so is equivalent to changing the rate-constant directly. If the model includes gene regulation and modulatory changes in protein activity, enzymes should be explicitly included as internal variables in the model. Mimicking changes in the medium can be done either by changing the values of an external variable or by changing the values of rate constants for the processes that are responsible for sensing those changes.

system steady-state can be easily computed numerically by solving the steady-state equation

$$0 = \sum_{r=1}^{p} \mu_{ir} \left(k_r \gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}} \right) \quad i = 1, \dots, n$$
 (4)

We shall use Eqs. (3) and (4) to analyze the effect of different activity patterns on the systemic performance of the model and evaluate how this performance influences the overall physiological outcome of the response.

2.3. Criteria for functional effectiveness in cellular metabolism

Changes in the reference steady-state as a consequence of a change in the enzyme activity pattern can be compared to a series of functional effectiveness criteria (Vilaprinyo et al., 2006; Voit and Radivoyevitch, 2000). Those criteria, which define the boundaries of internal change that the cell must go through in order to adapt and survive, are matched against the internal changes that are caused by the changes in enzyme activity of a given response profile. While some of those criteria may be quite general, others may be case-specific and may have different quantitative thresholds in different cases (Salvador and Savageau, 2003, 2006; Vilaprinyo et al., 2006). We now briefly discuss some of the criteria that have been used in the literature. These are useful for discussing operative ranges, evolution, and optimization of metabolic processes.

2.3.1. Change in metabolic fluxes

Changes in the rate of synthesis for key metabolites are important indices of functional effectiveness. For example, if a system regulates production of a metabolite in response to the cellular demand for that metabolite, an increase in the demand should lead to an increase in the production (e.g. Alves and Savageau, 2000 and references therein). The specific flux criteria are dependent on the system one is interested in and should be considered in the context of the whole system and not as isolated processes. For example, in the adaptive response to heat shock an increase in ATP production that causes depletion of NADPH or a dramatic decrease in glycolytic flux may be inappropriate in the general context of the adaptive response (Vilaprinyo et al., 2006).

In GMA models, steady-state fluxes can be easily computed for each condition using the following equation

$$v_{r_{ss}} = k_r \gamma_r \prod_{j=1}^{n+m} X_{j_{ss}}^{f_{rj}} \quad r = 1, \dots, p$$
(5)

where subscript *ss* indicates the relevant steady-state values corresponding to the new conditions. As stated before, the steady-state solution for metabolites can be obtained by numerically solving Eq. (4).

In larger networks that involve different branch points and regulatory effects, it is possible to obtain similar increases in a given set of fluxes with different patterns of modified enzyme activities. Thus, this criterion, by itself, will seldom be enough to assess the adaptive value of a set of changes and fully explain the observed operation principles for the system.

2.3.2. Metabolite accumulation

Changes in steady-state fluxes may often lead to changes in metabolite levels. From a practical point of view, either in biotechnological applications or in natural systems, one may argue that accumulation of intermediary metabolites may cause undesirable cross regulation side effects and tax the finite solvability capacity of the cell (see Alves and Savageau, 2000 and references therein). Thus, minimization of intermediate metabolite accumulation will be typically regarded as an important effectiveness criterion of an adaptive response, except for those cases in which metabolite accumulation might play an important role (for instance accumulation of trehalose in the heat shock response). Changes in metabolite levels are given by the steady-state solution to Eq. (4).

2.3.3. Overall changes in enzyme activities

Changes in enzyme activity are easy to simulate. However, it is often difficult to assess in a real situation whether those changes are indirect and due to the modulation of either gene expression or stability of mRNA (Garcia-Martinez et al., 2007; Romero-Santacreu et al., 2009), or direct and due to modulator effects on the activity of the protein. The later can arise via reversible covalent modification of specific residues or via changes in the conformation of the protein in response to a new set of physical chemical parameters in the medium. Changes in gene expression are costly in terms of metabolic resources (Wagner, 2005). They lead to mRNA and protein synthesis, which are among the most expensive metabolic activities of a cell. Thus, minimization of fold change can be considered an important functional effectiveness criterion (Raiford et al., 2008). If one assumes that changes in protein activity during the long term adaptive response of a cell are mostly due to changes in gene expression then, to a first approximation, one can estimate the cost of a given set of changes in enzyme activity by adding up all the k_r values. One possible way to account for both up- and downregulations consists of defining a "biological" cost of a response that is mathematically given by $\sum_{\forall i} |\ln(k_i)|$.

2.3.4. Parameter robustness

Parameter robustness is an important criterion as it refers to the system's sensitivity to slight differences in parameter values (Aldana et al., 2007; Coelho et al., 2009; Kitano, 2004; Kitano, 2007; Morohashi et al., 2002; Savageau, 1971). Systems with large parameter sensitivities may indicate the existence of processes that are more responsive to noise. Thus, they could be considered as less well adapted than systems that are more sensitive to parameter changes. Although low parameter sensitivities may arise from poorly identified parameters, one can argue that, in most cases, low sensitivity is a desirable property in well-adapted systems. This criterion has been extensively used in identifying design principles and in evaluating model adequacy and behavior (Cascante et al., 1995; Coelho et al., 2009; Curto et al., 1997; de Atauri et al., 2000; Voit, 2000).

2.3.5. Temporal responsiveness

Temporal responsiveness is another criterion that is important for systemic performance. Systems with inadequate temporal responsiveness may not survive to reach a new steady state, independently of the adequacy of their steady-state responses. In general, evaluating this criterion requires numerical simulations, except for the case where we are only interested in studying the dynamics in the neighborhood of a steady-state solution. In such a case one can linearize the system of equations about the steady state and obtain analytical solutions for the transient behavior of the dependent variables (Hlavacek and Savageau, 1998).

Unlike the other criteria that were discussed so far, using temporal responsiveness as a criterion for optimization poses many problems. In the context of globally optimizing metabolic systems, there are indeed very few methods capable of handling the dynamic constraints required to assess the temporal responsiveness. In fact, the strategies proposed so far are only applicable to specific types of models, and usually optimization uses large amounts of CPU time, even when tackling small problems with few variables and constraints (Chachuat et al., 2006; Chang and Sahinidis, 2005; Esposito and Floudas, 2000; Papamichail and Adjiman, 2002, 2004; Singer and Barton, 2006). This limitation can be overcome by performing the assessment of the temporal responsiveness in the post-optimal analysis of the solutions found. Hence, once a feasible solution is identified, the evaluation of its temporal responsiveness can add an extra criterion for deciding the relevance of such solution. In terms of evolution, this may be important as a given optimum can involve dynamic properties that will make the solution unfeasible in practice.

2.3.6. Steady-state stability

Dynamic stability is a criterion that evaluates the ability of a given system for returning to a steady-state after a perturbation. A stable system can accommodate fluctuations and will be able of maintain a reference state. Evaluation of steady-state stability should be a complementary criterion for testing the appropriateness of a proposed change in the system (Savageau, 1974a, 1975, 1998). In the optimization of metabolic systems, this criterion can be included in the optimization model itself (Chang and Sahinidis, 2005) or it can be assessed in the post-optimal analysis of the solutions for Eq. (4).

3. Feasibility regions in biochemical pathways: definition and their practical significance

3.1. Definitions

A feasibility region is a region in parameter space whose internal membership is defined by the sets of all parameter values that are compatible with specific physiological constraints (Dayarian et al., 2009; Guillén-Gosálbez and Sorribas, 2009). Here, without loss of generality, we shall concentrate on the special case of feasibility regions defined by changes in enzyme activities, that is the set of vectors representing the fold change in enzyme activities: (k_1, k_2, \ldots, k_p) , that are compatible with a set of functional effectiveness criteria (constraints). These functional criteria must be assessed through mathematical models, such as the GMA representation, that allow predicting the biological performance of a system in a specific environment. In mathematical terms, performing a feasibility analysis entails conducting a systematic search for determining the set of values of some variables of the biological model for which the overall formulation remains feasible. In this context, linear models usually fail to capture the whole complexity of the biological system, so it is necessary to use nonlinear formulations.

Hence, finding the boundaries for this class of feasibility regions requires obtaining global optimal solutions for nonlinear optimization problems. One of the important limitations of standard nonlinear optimization techniques is that they cannot guarantee the global optimality of the solutions found when they are applied to nonlinear problems that have non-convexities. Non-convexities, such as bilinear terms, fractional terms, etc., are very common in many engineering problems. In the context of our analysis, these non-convexities arise from the kinetic equations required to link the concentration of the metabolites with the velocities of the reactions that take place in the metabolic network.

There are currently several global optimization methods that can handle non-convex problems and provide solutions that are globally optimal within a desired tolerance (Tawarmalani and Sahinidis, 2002). Most of these methods are general purpose, that is to say, they can be applied to a wide range of problems regardless of the type of non-convexities embedded in the model. However, their performance can change drastically from one application to another depending on the specific structure of the problem to be solved (for a detailed review of these methods see Grossman and Biegler, 2004). A possible way of expediting the search for global solutions for nonlinear non-convex problems consists of exploiting the structure of the involved non-convexities. The major classes of non-convex problems studied so far include concave minimization (Hansen et al., 1992) and problems with linear fractional and bilinear terms (Quesada and Grossman, 1995), and a method for problems with signomial parts (Porn et al., 2008). Different optimization strategies have also been suggested for S-system and GMA models within BST (Chang and Sahinidis, 2005; Hatzimanikatis et al., 1996; Marin-Sanguino et al., 2007; Polisetty et al., 2008; Voit, 1992). Recently, a highly efficient global optimization technique for GMA models has been developed by our group. Technical aspects of this optimization are discussed elsewhere (Guillén-Gosálbez and Sorribas, 2009; Pozo et al., submitted for publication). We shall use this technique in the feasibility analysis presented here.

3.2. Characterization of feasibility regions in GMA models

The method for finding the feasibility regions was first introduced by Guillén-Gosálbez and Sorribas (2009). Here, we briefly review it and discuss the different steps and their importance. Mainly, steps 2–3 are critical for reducing the search space and obtain a useful result. After reviewing the method, we shall apply it to two practical cases showing its utility both for optimization and evolutionary studies. Finally, we shall stress the role of the set of physiological constraints in defining the feasibility region. As stated before, the strategy presented relies on the use of global optimization methods that are customized for this particular application.

A feasibility region for a particular problem can be identified through the following steps:

- 1. Define a set of constraints that must be fulfilled by any solution (limits for fluxes, concentrations, gene expression, etc.). At this point, collaboration with experts in the biological problem is fundamental.
- 2. Define the search space for the fold change of each enzyme. (i.e., the lower and upper bounds, k_i^{LO} and k_i^{UP} that define the interval within which the fold change must fall). Based on experimental information, one can restrict the search space for practical purposes. Thus, if measurements in microarray experiments show that during the studied response a given gene is over expressed between 5 and 8-fold, we could consider allowing changes from 1 up to 20-fold from the basal condition for that gene. By making the range so large, one covers for other plausible values that may also be linked to alternative adaptive solutions.
- 3. Find the maximum and minimum (*bound contraction*) values for changing each enzyme that are compatible with the set of constraints defined in 1 and with the limits established in 2. This is achieved by defining k_i as the objective function. Note that these optimizations provide bounds for all the variables k_i (i.e., $k_i^{\text{LO}*}$ and $k_i^{\text{UP}*}$) that will fall within those first considered in step 2. Thus, we can assure that outside the obtained bounds for k_i no other combination of changes in the enzymes produces a valid solution. In mathematical terms, we have that a solution will be unfeasible if there exists at least one k_i that satisfies that $k_i \notin [k_i^{\text{LO}*}, k_i^{\text{UP}*}]$.
- 4. Define a grid of values for each k_i : $(k_i^1, k_i^2, \ldots, k_i^{n_i})$, using the minimum and maximum values obtained in the previous step. Typically, we have divided the allowed range in 10 sections. The bound contraction step shortens the search region, which results in a more efficient search of the feasibility region.
- 5. Consider a set of the hyper-rectangles, each of which is defined by lower and upper limits imposed on the values of each k_i . For instance, a particular hyper-rectangle would be defined by fold changes that are between 2.5 and 3.7 for enzyme 1, between 10 and 12.3 for enzyme 2, and so on.

- 6. Find the global optimum using any of the velocities as the objective function. This will give a set of fold changes $(k_{11}, k_{21}, \ldots, k_{p1})$ for which this optimum is attained. At this stage, as the goal is to find admissible solutions, we can select any of the velocities as the objective function. The results of the feasibility analysis will be the same independently of this choice. At this point, all we need is to be sure that at least a solution exist that is compatible with the set of constraints. Hence, it is not strictly required to globally optimize the model in each hyper-rectangle, since a feasible solution suffices for the purpose of the analysis. One can wrongly conclude from this observation that it is possible to conduct the feasibility analysis using a local optimization method. This is not true, as the task of the algorithm is not only to identify feasible solutions in each hyperrectangle, but also to discard regions in which no feasible point exists. Standard local optimization methods cannot accomplish the latter task, as they can fail even in solving convex problems (Tawarmalani and Sahinidis, 2002), in such a way that one will never be sure if the convergence problems that will arise when attempting to optimize empty hyper-rectangles will really indicate the absence of feasible solutions. One possible way to circumvent this issue is to rely on a lower bounding problem, which is one of the main ingredients of any global optimization technique, capable of providing a valid lower bound on the global optimum of the model. Particularly, the feasibility analysis requires the use of a linear lower bounding problem, since linear and mixed-integer linear programming techniques (LP and MILP, respectively) can indeed identify problems with no feasible solutions. Hence, our method exploits the fact that LP and MILP techniques will only fail when attempting to solve models of small/medium size that are really unfeasible (i.e., do not contain any feasible point). Note that in this context the main task of the lower bounding problem is not to provide a tight bound on the global solution of the model, as is the case in standard global optimization methods, but to detect empty hyper-rectangles from unfeasible models. Thus, the particular features of our feasibility analysis justify the need for a customized global optimization method.
- 7. Identify and annotate the hyper-rectangle that contains this solution. This will be the one whose lower and upper limits contain the values of k_i associated with the optimal solution identified in the current iteration.
- 8. Repeat steps 4–7 by excluding the hyper-rectangle containing the optimal solution obtained in step 5 by adding an integer cut to the lower bounding problem. This is repeated until no further solution is found to be compatible with the remaining hyperrectangles (i.e., until the lower bounding problem turns out to be unfeasible).
- 9. Analyze the obtained results and compare the feasible region with actual experimental data. If the feasibility region contains the observed data, this is an indication that the considered set of constraints may explain the adaptive response. Alternative constraints can be introduced and a new feasibility region can be obtained by starting again the analysis at point 2. In the next section we will discuss the interpretation of results obtained with different sets of constraints.

This procedure is illustrated in Fig. 1. For simplicity, we show results for two enzymes only. However, at each optimization, all the enzymes are allowed to change values (see below). The constraints considered in each optimization are those defined in step 1 plus the limits on the values of k_i that define each hyper-rectangle. By following the procedure described above, a region of feasibility is bounded and defined by a set of feasible hyper-rectangles. These regions can be further refined by increasing the granularity of the hyper-rectangles within the region(s) of feasibility. Note that any of the hyper-rectangles that define the feasible regions contains at least one admissible solution.

3.3. Utility of feasibility regions characterization

The feasibility regions determined through the previously described methodology can be most useful in two situations:

1. *In biotechnological applications*. In practical applications, it may be impossible to attain the optimal solution that would



Fig. 1. Strategy for finding a feasible region. In the first step, the global optimum is identified and the hyper-rectangle where it occurs is annotated (a). In the next step, this hyper-rectangle is discarded and the new optimum is located (b). The process is repeated (c) until no new optimum is obtained (d). Here we show results for two of the hypothetical enzymes but the search is done for all simultaneously.



Fig. 2. Schematic representation of the central metabolism of yeast. Details are discussed in previous papers (Polisetty et al., 2008; Vilaprinyo et al., 2006; Voit and Radivoyevitch, 2000). Basal values for enzyme activities and the resulting steady state are given in Tables 1 and 2.

1

correspond to an optimization analysis. Thus, one can define a minimum percentage of the optimum that would identify a practical cost-beneficial strategy. Feasible regions that are compatible with that threshold can be obtained using the method proposed in this work. The desired minimal limit can be mathematically represented by a simple inequality constraint. Once the feasibility region is determined the user can select the most appropriate values for practical implementation. The feasible regions will contain a global optimum for attaining this practical threshold as well as many other subobtimal solutions. Note that all the identified solutions, including the sub-optimal ones, would be guaranteed to attain the minimum increase in the objective function considered in the analysis.

2. *In evolutionary studies.* Feasible regions that are compatible with physiological requirements can be identified in studies about evolution of responses. If the model captures the features of the system that are important for the response, one would expect to find the actual adaptive response within this region. An iterative analysis considering different physiological constraints may help in identifying which of these constraints are more important as selective pressures for evolving an appropriate response, avoiding the spandrel effects. Furthermore, comparison of actual data with optimal solutions can help in understanding the selective pressures in a given case.

3.4. Examples

3.4.1. Metabolic model

As an example for showing the applicability of the method described above, we shall consider a simplified conceptual model for the basal metabolism of yeast that is derived from previous models of the same pathways (Curto et al., 1995; Polisetty et al., 2008; Voit and Radivoyevitch, 2000).

This model, summarized in Fig. 2, accounts for glycolisis, the synthesis of glycogen and trehalose, the branching from fructose-1,6-P to glycerol, and the branching from glycolisis to the pentose phosphate metabolism. For convenience, we consider simplified reactions by lumping together a number of processes. For example, we consider an aggregated process leading to trehalose and glycogen. Numerically, we shall consider that the flux into trehalose is a fraction of the total flux for this branch (Vilaprinyo et al., 2006; Voit and Radivoyevitch, 2000). For more details on the simplifications, assumptions, and experimental evidences used to build this model the reader is referred to the paper by Voit and Radivoyevitch (2000). The different processes are modeled using the power-law formalism as:

Process	Velocity	Power-law representation	Steady-state rat	e
HXT	v_1	$0.9023X_2^{-0.2344}X_6$	17.73	
GLK	v_2	$3.1847X_1^{0.7464}X_5^{0.0253}X_7$	17.73	
PFK	v_3	$0.5232 X_2^{0.7318} X_5^{-0.3941} X_8$	15.946	
TDH	v_4	$0.011 X_3^{0.6159} X_5^{0.1308} X_9 X_{14}^{-0.6088}$	15.06	(6)
РҮК	v_5	$0.0947 X_3^{0.05} X_4^{0.533} X_5^{-0.0822} X_{10}$	30.00	(0)
TPS + GLY	v_6	$0.0009X_2^{0.7318}X_{11}$	0.014	
G6PDH	v_7	$1.76898 X_2^{0.0526} X_{15}^{0.9646}$	1.77	
GOL	v_8	$0.103209 X_3^{0.05} X_4^{0.533} X_5^{-0.0822} X_{12}$	1.772	
ATPase	v_9	$0.937905X_5^1X_{13}$	26.55	

The stoichiometric matrix corresponding to the model in Fig. 2 is given by:

$$N = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1/2 & 0 \\ 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 2 & 1 & -1 & 0 & 0 & -1 \end{pmatrix}$$
(7)

Multiplying the stoichiometric matrix by the vector of velocities $\mathbf{V} = (v_1, \dots, v_9)'$, we would obtain the set of differential equations for the model in GMA form²:

$$\dot{\mathbf{X}} = \mathbf{N} \cdot \mathbf{V} \tag{8}$$

² Here we use the notation $\dot{X}_i = dX_i/dt$.

Table 1 Basal enzyme activities.

Symbol	Name	Value
X ₆	Glucose uptake (HXT)	19.7 mM min ⁻¹
X7	Hexokinase (GLK)	68.5 mM min ⁻¹
X_8	Phosphofructokinase (PFK)	31.7 mM min ⁻¹
X_9	Glyceraldehyde-3-phosphate	49.9 mM min ⁻¹
	dehydrogenase (GAPD or, as	
	alternative name, TDH)	
X ₁₀	Pyruvate kinase (PYK)	3440 mM min ⁻¹
X ₁₁	Polysaccharide production	14.31 mM min ⁻¹
	(glycogen + trehalose)	
X ₁₂	Glycerol production (GOL)	203 mM min ⁻¹
X ₁₃	ATPase	25.1 mM min ⁻¹
X_{14}	NAD ⁺ /NADH ratio	0.042

a 4 a 4 a 1 a 7 4 6 4 x 0 0 2 5 3 x

Table 2

Steady-state values of the considered model at the basal conditions.

Symbol	Name	Basal concentration (mM)
X_1	Internal glucose	0.0345
X_2	Glucose-6-phosphate	1.011
X_3	Fructose-1,6-diphosphate	9.144
X_4	Phosphoenolpyruvate (PEP)	0.0095
<i>X</i> ₅	ATP	1.1278

time, which can be unpractical. One possible application of the feasibility method proposed here is to explore possible changes in enzyme activity leading to acceptable solutions, say a given percentage over the basal value or a given percentage below the optimum value. As an example, we will consider the optimization of ethanol production using the reference model. Because of the

The complete mathematical model is given by:

-0.2344

÷

$$\begin{aligned} \dot{X}_{1} &= 0.9023X_{2}^{-0.0311}X_{6}^{-} - 3.1847X_{1}^{0.1464}X_{5}^{0.0233}X_{7} \\ \dot{X}_{2} &= 3.1847X_{1}^{0.7464}X_{5}^{0.0253}X_{7} - 0.5232X_{2}^{0.7318}X_{5}^{-0.3941}X_{8} - 0.0009X_{2}^{0.7318}X_{11} - 1.76898X_{2}^{0.0526}X_{15}^{0.9646} \\ \dot{X}_{3} &= 0.5232X_{2}^{0.7318}X_{5}^{-0.3941}X_{8} - 0.011X_{3}^{0.6159}X_{5}^{0.1308}X_{9}X_{14}^{-0.6088} - 0.0516X_{3}^{0.05}X_{4}^{0.533}X_{5}^{-0.0822}X_{12} \\ \dot{X}_{4} &= 2 \times \left(0.011X_{3}^{0.6159}X_{5}^{0.1308}X_{9}X_{14}^{-0.6088}\right) - 0.0947X_{3}^{0.05}X_{4}^{0.533}X_{5}^{-0.0822}X_{10} \\ \dot{X}_{5} &= 2 \times \left(0.011X_{3}^{0.6159}X_{5}^{0.1308}X_{9}X_{14}^{-0.6088}\right) + 0.0947X_{3}^{0.05}X_{4}^{0.533}X_{5}^{-0.0822}X_{10} - 3.1847X_{1}^{0.7464}X_{5}^{0.0243}X_{7} - \\ 0.0009X_{2}^{0.7318}X_{11} - 0.5232X_{2}^{0.7318}X_{5}^{-0.3941}X_{8} - 0.937905X_{5}^{1}X_{13} \end{aligned}$$

The basal enzyme activities used in the models are shown in Table 1. The steady-state calculated from these values and the model parameters given in Eq. (9) is shown in Table 2.

3.4.2. Feasible regions for a significant increase in ethanol production

Optimization of cellular processes is an important goal in biotechnology. However, optimal solutions obtained with a model will seldom be practically realizable. In most cases, significant increases in flux would imply modifying many enzymes at the same

simplific	ations	introdu	iced in	the	model,	the	rate	of s	synthesi	s of
ethanol i	s the s	ame as	that fo	r the	synthe	sis of	f pyrı	ivat	te.	

First, as a reference for comparison, we explore the maximum rate of ethanol production that can be achieved if changes are allowed in all enzyme activities. The results of the optimization analysis using the method described elsewhere (Guillén-Gosálbez and Sorribas, 2009) are shown in Table 3. For comparative purposes, we obtain the optimal solution with different allowed ranges for enzyme activity changes. A nearly linear increase in ethanol production is achieved as we allow higher increases in the enzymes.

Table 3

Maximization of ethanol production.

No constraints Maximum fold change in any enzyme			VNADPH (5% maximum variation) Maximum fold change in any enzyme				VNAPDH, VATP (5% maximum variation) Maximum fold change in any enzyme					
	5	10	15	20	5	10	15	20	5	10	15	20
Fold change	e values at the	e different op	timum									
HXT	5	10	15	20	5	10	15	20	1.01	1.01	1.01	1.01
GLK	1.16	10	10.56	16.03	1.16	2.16	7.6	20	0.27	0.27	0.27	0.27
PFK	5	10	15	20	5	10	15	20	1.14	1.14	1.14	1.14
TDH	5	10	15	5.27	5	9.99	15	20	2.58	2.58	2.58	2.58
PYK	5	10	15	20	5	10	15	20	0.34	0.34	0.34	0.34
TPS	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
G6PDH	0.2	0.2	0.2	0.2	0.94	0.94	0.94	0.94	0.95	0.95	0.95	0.95
GOL	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.22	0.22	0.22	0.22
ATPase	5	10	15	20	5	10	15	20	0.94	0.94	0.94	0.94
Steady-stat	e values for n	netabolites (n	nM) correspon	ding to the dif	ferent optim	um						
Glu	0.23	0.03	0.05	0.04	0.23	0.25	0.08	0.03	0.21	0.21	0.21	0.21
Glu-6-P	1.23	1.24	1.24	1.24	1.21	1.22	1.23	1.23	0.95	0.95	0.95	0.95
F-1,6-P	10.41	10.43	10.44	91.01	10.29	10.38	10.40	10.42	2.05	2.05	2.05	2.05
PEP	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.09	0.09	0.09	0.09
ATP	1.42	1.43	1.43	1.43	1.40	1.41	1.42	1.42	1.29	1.29	1.29	1.29
Steady-stat	Steady-state values for fluxes (mM min ⁻¹) corresponding to the different optimum											
VATP	336.1	673.8	1011.5	1348.9	332.8	670.2	1008.2	1345.9	63.4	63.4	63.4	63.4
V _{TRE}	0.00027	0.00027	0.00028	0.00028	0.00027	0.00027	0.00027	0.00027	0.00022	0.00022	0.00022	0.00022
V _{NADPH}	0.36	0.36	0.36	0.36	1.68	1.68	1.68	1.68	1.68	1.68	1.68	1.68
$V_{\rm GLY}$	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	1.33	1.33	1.33	1.33
VETHANOL	168.0	336.9	505.8	674.6	166.3	335.2	504.1	672.9	31.9	31.9	31.9	31.9

The different scenarios are defined by allowing a maximum fold change increase for any of the enzymes of 5, 10, 15, and 20-fold. Optimal enzyme patterns are obtained without any other restriction (left) and with a maximum allowable change in the rate of NADPH synthesis of 5% about its basal value (center) and a maximum allowable change in the rate of NADPH and ATP synthesis of 5% about its basal value (right). Steady-state values of metabolites and relevant fluxes resulting from the optimal change are also shown for comparison.



Fig. 3. Feasibility analysis of the maximum ethanol production when only HXT (k_1) and PFK (k_3) are allowed to change. Values inside the left tables indicates optimum ethanol production within each cell. Cells are defined by the values of k_1 and k_3 indicated in the right tables. In each case, the minimum and maximum value defining each cell are shown in those tables. Color code shows the decreasing ethanol production that can be attained in different conditions. Blue color and 0 values indicate unfeasible combinations. (a) Optimization constrained to prevent an accumulation of intermediary metabolites that is over 10 times their basal value. (b) Optimization constrained to prevent an accumulation of the references to color in this figure legend, the reader is referred to the web version of the article.)

When no restriction is considered (Table 3 left), with a 20-fold change one can reach a velocity of 674.6 mM min⁻¹ that is much higher than the basal value of 30.0 mM min⁻¹. While HXT, PFK, PYK, and ATPase should increase 20-fold, GLK and TDH require a lower increase. In all four scenarios, optimization of ethanol production should require lowering TPS, G6PDH, and GOL activities. Imposing limits on the changes of NADPH production leads to a similar result, but now the activity of G6PDH is almost unchanged (Table 3 center). It is important to stress that in all the cases the synthesis of ATP increases by a large amount, from a basal value of 46.07 mM min⁻¹ to 1348.9 mM min⁻¹ when the maximum fold change allowed is 20. If we restrict both the increase in NADPH and ATP production, then the maximum attainable ethanol production drastically decreases (Table 3 right).

In the previous examples, all the enzymes were allowed to change. A more realistic approach that could be translated into wet lab experiments should analyze the practical possibilities of increasing ethanol production when only a small number of enzymes are changed. For illustrative purposes, based on the previous analyses of this problem (Guillén-Gosálbez and Sorribas, 2009; Polisetty et al., 2008; Vilaprinyo et al., 2006), we select HXT and PFK. We shall maintain all the other enzymes fixed at their basal activity. As we are looking for solutions that do not compromise cell viability, we shall enforce the condition that all the internal metabolites should not change more than 10-fold from their basal values (Polisetty et al., 2008). Under such constrains, the maximum rate of ethanol production that can be obtained in the model is 100.52 mM min⁻¹ (Fig. 3a).

Now, we will obtain the feasibility region under the same constrains. This is an alternative to just finding the optimal solution and it may help in discussing the changes that can be implemented in practice. First, we obtain the feasibility region without limitation in additional fluxes. In this case, the feasibility region has admissible ranges between 0.21 and 5.38 for changing HXT, and a range between 0.2 and 15 for PFK. Outside these limits, no feasible solution can be obtained (Fig. 3a). While it is reasonable to expect that ethanol production would increase by increasing HXT, our results show that increasing simultaneously HXT above 4.86fold and PFK above 3.16-fold (cell number 9 for HXT and cell 3 for PFK in Fig. 3a) leads to unfeasible solutions. This is so because intermediary metabolites would accumulate and the fitness of cells would decrease. Thus, a biotechnological implementation of a 5-fold change in HXT and a 5-fold change in PFK is expected to result in a failure in producing a viable strain. Our results also show that near optimum increases in ethanol production can be obtained in different conditions. As far as HXT activity is increased a minimum of 6.12-fold, we can obtain an almost optimal ethanol increase with different increases in PFK activity. For instance, we could decide a 3-fold increase in PFK and an 8-fold increase in HXT to obtain an almost optimal solution.

Planning a biotechnological strategy for increasing the production of a given metabolite must consider all the implications of the planned changes in the overall cellular metabolism. As a second scenario, we have determined the feasible solutions by imposing the additional constraint of maintaining the rate of NADPH within a 5% of its basal level (Fig. 3b). Now, the feasible region has been drastically reduced and the possible increase in ethanol production is almost minimal when compared to its basal value. Thus, in those cases in which maintaining the rate of NADPH unchanged is an important limitation, it is impossible to find a strategy involving changes in HXT and PFK capable of producing a significant increase in ethanol production (Fig. 3b).

These results show the potential application of our feasibility analysis in practical applications. Following this procedure, we can efficiently obtain an overall picture of the attainable values and a clear estimation of the unfeasible changes. This may help in discussing the most convenient implementation and the expected increase one would obtain in the objective function.

3.4.3. Feasible regions for the adaptive response to heat shock in yeast

Understanding the evolution of adaptive responses was the main motivation for developing the feasibility method (Guillén-Gosálbez and Sorribas, 2009). As stated before, the goal of the feasibility analysis is to obtain admissible values for enzyme activity changes that drive the model to a new state in which a set of

constraints are satisfied. Here, two fundamental ingredients are required. First, we need a mathematical model that is accurate enough to represent the biological problem at hand. Second, a set of constraints must be defined, so that changes in enzyme activity can be evaluated for compatibility.

Both issues pose significant challenges. Useful mathematical models are hard to build and, in most cases, parameters values for those models are difficult to obtain. This limitation is common to any application as discussed in the optimization section. Finding a set of constraints for the response is not an easy task either, and a sound biological understanding of the problem is required. As an example of the potential use of the methodology presented here, we consider the model presented in Fig. 2. First, we shall perform a feasibility analysis taking into account the set of constraints C_1-C_6 suggested by Vilaprinyo et al. (2006) (see Table 4). These constraints were identified and used for a previous analysis of operating principles in the adaptive response of yeast to heat shock.

Taking these constrains into account, we first find the upper and lower admissible values for changing each enzyme. Mathematically, this corresponds to performing a *bound contraction* on some continuous variables of the non-convex problem. Specifically, those limits are obtained by solving an optimization problem that finds the maximum and minimum values of a given k_i for which admissible solutions are found. Results are shown in Fig. 4a. These results are a generalization of those in Fig. 2(D) in the paper of Vilaprinyo et al. (2006). As we are now using a systematic search, our results include those obtained previously by intensive compu-

Table 4

Physiological constraints for the feasibility analysis (see Vilaprinyo et al., 2006 for details).

Constraint	Value
C1	$V_{\rm ATP}$ > 180.6 mM min ⁻¹
C ₂	$V_{\rm TRE}$ > 0.03 mM min ⁻¹
C ₃	$V_{\rm NADPH}$ > 3.54 mM min ⁻¹
	Internal glucose > 0.04 mM
	G6P < 20.22 mM
C4	F16P < 22.86 mM
	PEP < 0.01 mM
	ATP < 6.77 mM
C ₅	Cost < 12.06
<i>C</i> ₆	$V_{\rm Glycerol}$ > 0.39 mM min ⁻¹

See Vilaprinyo et al. (2006) for details.

tations and are slightly wider as the previous analysis was done by considering only a set of discrete values. Furthermore, by using the new procedure, computational time is dramatically reduced to seconds.

In Fig. 4b, we plot the activity profiles corresponding to different experimental measurements (see details in Table I of Vilaprinyo et al., 2006). Note that all the experimental results are within the predicted values. Imposing two extra constraints (C_7 – C_8) on the changes in PFK and TPS relative to the rate of trehalose synthesis ($\Psi = (\Delta PFK \times \Delta TPS)/v_{TRE}$, $\Psi < 100$), and a minimum value of F-1,6-P of 8.16, (criteria C_7 , C_8 in Vilaprinyo et al., 2006), the limits for PFK are drastically reduced (Fig. 4c), although the resulting



Fig. 4. Result of the bound contraction procedure. In each case, the maximum and minimum admissible change folds for each enzyme are indicated. (a) Bounds with C_1-C_6 (Table 5), (b) experimental data plotted to show they are located in the admissible region found in (a), (c) Bounds with C_1-C_8 (see text for the definition of criteria C_7 and C_8), (b) experimental data plotted to show they are located in the admissible region found in (c). Experimental data are those of Tables 1 and 2 in Vilaprinyo et al. (2006). The search regions allowed for the change-fold in each enzyme are shown in Table 6.

Author's personal copy

A. Sorribas et al. / Journal of Biotechnology 149 (2010) 141-153



Fig. 5. Feasibility regions for a simultaneous change in two enzymes. In each case, all the other enzymes can change to compensate and make the changes compatible with the constraints C_1 – C_6 . Red rectangle identifies the limit for changing a given enzyme. For instance, in the case of PFK we have considered changes between 0.2 and 4. These conditions are the same considered in Vilaprinyo et al. (2006) and are maintained here for comparison. Red rectangles indicate admissible solutions. White regions are unexplored in that example. Blue points indicate experimental values described in Table 1 of Vilaprinyo et al. (2006). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

bounds still contain the observed results (Fig. 4d, see also Fig. 2(E) in Vilaprinyo et al., 2006).

The complete results of performing the feasibility analysis are presented in Fig. 5. As indicated above, we first obtain the limits for admissible solutions taking into account criteria C_1-C_6 and specific limits imposed on the allowable fold changes in each enzyme based of experimental results (Table 5). For clarity, in Fig. 5 we show oneby-one figures that show the simultaneous feasibility regions for two particular enzymes. It can be seen that some enzyme activities

Table 5Limits for the fold change in the different enzymes in the
feasibility analysis of Fig. 5.

Enzyme	Explored fold change
HXT	1 < <i>k</i> ₁ < 10
GLK	1 < k ₂ < 19
PFK	$0.25 < k_3 < 4$
TDH	0.25 < k ₄ < 6
РҮК	$0.25 < k_5 < 20$
TPS	$1 < k_6 < 19$
G6PDH	1 < <i>k</i> ₇ < 8

These limits were defined considering experimental results. In each, a wide region around the values observed after heat shock are selected (see Vilaprinyo et al., 2006 for details). can take a wide range of values within their allowable boundaries, while still fulfilling the imposed constraints. This is the case for TPS and GLK. For other enzymes, feasible changes are more restricted. For example, PFK and TDH cannot increase by more than 5-fold their basal values. Outside this range, the system cannot compensate the changes and the constraints are not met. This is also the case for PFK and PYK. Feasible solutions for changes in both enzymes are obtained only in a relatively narrow margin. Interestingly, experimentally measured changes from different experiments are found to be within the feasibility regions identified by our method (see Vilaprinyo et al., 2006 for details). This is consistent with the notion that the set of constraints defined for the response are relevant for the physiological adaptation of yeast.

3.4.4. On the importance of an appropriate set of constraints

The set of initial physiological constraints that are applied to the optimization procedure play a fundamental role in the feasibility analysis. Different sets of constraints are likely to produce different feasibility regions. The situation is exemplified in Fig. 6. Each of the represented regions would correspond to different sets of constraints. In this hypothetical situation, regions (1), (2), and (4) contain experimental results, while (3) and (5) do not. Thus, the constraint sets leading to regions (3) and (5) could be discarded as explanatory physiological constraints for that case. Constraints



Fig. 6. Hypothetical feasibility regions obtained from five different sets of constraints. Points represent experimental results on fold change for enzymes Ei and Ej in a given adaptive response (see text for details).

producing region (1) should be considered more restrictive than those of (2), although both explain the observed result. Finally, the set of constraints that produce region (3) partially explain some results but not others. In principle, this set of constraints would be less appropriate to describe the physiological requirements of the response than sets (1) and (2).

How would changing constraint sets affect the results in our analysis of yeast heat shock response? As an example, we have considered an alternative set of constraints to those used above (compare Tables 4 and 6). For illustrative purposes, in feasibility analysis using constraints from either Table 4 or Table 6, the activity of any enzyme is allowed to change between 0.2 and 20-fold. For simplicity, only the results for PFK, TDH and PYK are shown in Fig. 7. The two sets of constraints result in different feasibility regions that share some common values. Interestingly, the feasibility region obtained with the new set of constraints does not contain all the experimental values (see Table 1 in Vilaprinyo et al., 2006). This suggests that this second set of constraints does not adequately describe the physiological requirements that may have shaped the adaptive response of yeast to heat shock.

Table 6

Alternative set of constraints for evaluating heat shock response in yeast.

Constraint	Value
C ₁	$V_{\rm ATP} > 5B$
C ₂	$V_{\text{TRE}} > 30B$
C ₃	$V_{\rm NADPH} > 5B$
C ₄	3B < internal glucose < 5B
	15B < G6P < 20B
	2B < F16P < 5B
	2B < PEP < 5B
	3B < ATP < 6B
C ₅	Cost < 20
<i>C</i> ₆	$V_{\rm Glycerol} > B$

B indicates de corresponding basal value (Tables 1 and 2) for the flux or metabolite considered in each criterion.

4. Discussion

Understanding why metabolic pathways evolved to be as they are and how to optimize them are two closely related subjects. Studies in either field often use similar tools to compute the response of the whole system to changing conditions.

In optimization problems, control variables are manipulated by the experimenter and a predefined goal is pursued. This is often the case in metabolic engineering studies, where the general goal is that of modifying cells so that specific production targets can be reached (Hatzimanikatis et al., 1998). Typically, one considers optimizing the yield of a given process, maximizing flux through a pathway, etc. Then, optimization procedures are used on a mathematical model for the relevant processes in order to analyze which changes are the most likely to produce the desired result (Gianchandani et al., 2008). There is a wide scope of optimization methods that can be used for this task, based on different optimization strategies (Banga, 2008; Chang and Sahinidis, 2005; Marin-Sanguino et al., 2007; Nielsen, 2007; Polisetty et al., 2008; Schuetz et al., 2007; Vital-Lopez et al., 2006).

In evolutionary studies, however, we are faced with conserved changes that can appear in organisms by random mutations, by gene transfer, gene duplication, gene deletion, and other mechanisms. Natural selection may operate as a purifying mechanism that acts upon the systemic effect of these changes on the overall fitness



Fig. 7. Feasibility analysis obtained with two different sets of constraints. Constraint set 1 correspond to Table 5, constraint set 2 to Table 6. In both cases, enzymes are allowed to change between 0.2 and 20-fold over basal. Blue points indicate experimental measurements as presented in Table 1 of Vilaprinyo et al. (2006).

and leads to the fixation of new designs and operative patterns in a population, due to the differential reproduction of individuals. As a result, organisms often evolve towards some quasi-optimal regime under the conditions they live in. Such regime however may become quite sub-optimal if conditions change drastically. Those changes could lead to a new round of natural selection, this time with different physiological constraints. Thus, evolution in natural systems can be seen as a perpetual optimization-like process, with the parameter conditions that maximize survival and reproduction shifting over time.

In fact, one of the biggest current problems in this area is how to establish a connection between what researchers see as the actual functioning conditions of the molecular pathways that allow a cell to perform appropriately and the fitness of that cell. Causative genotype phenotype models (Martens et al., 2009; O'Connor and Mundy, 2009) are but a start in connecting the optimization of the molecular determinants of life and the fitness of organisms. We hypothesize that adaptive responses are to be found within feasible regions that allow the system to meet a set of physiological constraints that are required for cell survival (Guillén-Gosálbez and Sorribas, 2009; Vilaprinyo et al., 2006). The numerical thresholds considered in these constraints would shape the admissible changes in the system parameters so that the effect on global fitness can be sensed by natural selection. As a result, a specific adaptive response would evolve. Future work should deal with connecting the molecular aspects of the adaptive response to the direct survival ability.

From a practical point of view, there is a set of considerations that should be taken into account in optimization related studies of biological problems at the pathway level: (1) a model that can be used to compute fluxes, metabolite levels, the effect of changes in parameters, dynamic response, etc., is required; (2) stoichiometry-based models, such as Flux Balance Analysis models, are not sufficiently accurate to be used for characterizing quantitative changes. This is so because they do not account for regulatory interactions within the network and cannot be used to accurately calculate metabolite levels, dynamic changes, and other quantitative information (Nikolaev, 2009); (3) models that include information about the regulatory signals are essential for an accurate analysis; (4) kinetic information, even if it is only approximated, is required to define a quantitative model that may help in the analysis. Because of (4), at present we are still unable to create genome-wide models for metabolism, because not enough information is available. However, the obtained results show the importance of developing GMA-like models as a basis for a more complete analysis of system optimization and evolution. In this paper we have presented a methodology designed to address important practical questions, both in metabolic engineering applications and in studies of pathway evolution, through the use of a global nonlinear optimization technique and the characterization of feasibility regions. Although linear global optimization methods had been used before to search for survivability regions in Flux Balance Analysis models, those studies have the limitations described in points (1)-(4) of the previous paragraph. Our methodology overcomes those limitations and it can be applied to a special class of nonlinear differential equations models known as GMA models.³ If such a model is defined for a given metabolic problem, then our method allows for an exhaustive exploration of different evolutionary strategies and a systematic characterization of the physiological requirements that may underlie the evolution of adaptive strategies.

Acknowledgements

AS acknowledges the financial support from grant BFU2008-0196 (Ministerio de Ciencia e Innovación (MICINN, Spain)). GG-G expresses his gratitude for the financial support from the MICINN (projects DPI2008-04099, PHB2008-0090-PC, BFU2008-754 00196 and CTQ2009-14420-C02-01) and the Spanish Ministry of External Affairs (projects A/8502/07, HS2007-0006 and A/020104/08). RA is supported by the MICINN through the Ramon y Cajal program and grant BFU2007-62772/BMC. EV is supported by grant PI06/1649 from the Spanish Ministry of Health.

References

- Aldana, M., Balleza, E., Kauffman, S., Resendiz, O., 2007. Robustness and evolvability in genetic regulatory networks. J. Theor. Biol. 245, 433–448.
- Alvarez-Vasquez, F., Sims, K.J., Hannun, Y.A., Voit, E.O., 2004. Integration of kinetic information on yeast sphingolipid metabolism in dynamical pathway models. J. Theor. Biol. 226, 265–291.
- Alves, R., Savageau, M.A., 2000. Extending the method of mathematically controlled comparison to include numerical comparisons. Bioinformatics 16, 786–798.
- Bailey, J.E., 2001. Reflections on the scope and the future of metabolic engineering and its connections to functional genomics and drug discovery. Metab. Eng. 3, 111–114.
- Bailey, J.E., 1999. Lessons from metabolic engineering for functional genomics and drug discovery. Nat. Biotechnol. 17, 616–618.
- Bailey, J.E., Sburlati, A., Hatzimanikatis, V., Lee, K., Renner, W.A., Tsai, P.S., 1996. Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. Biotechnol. Bioeng. 52, 109–121.
- Bailey, J.E., 1991. Toward a science of metabolic engineering. Science 252, 1668–1675.
- Bailey, J.E., Birnbaum, S., Galazzo, J.L., Khosla, C., Shanks, J.V., 1990. Strategies and challenges in metabolic engineering. Ann. N. Y. Acad. Sci. 589, 1–15.
- Banga, J.R., 2008. Optimization in computational systems biology. BMC Syst. Biol. 2, 47.
- Bedford, T., Hartl, D.L., 2009. Optimization of gene expression by natural selection. Proc. Natl. Acad. Sci. U.S.A. 106, 1133–1138.
- Braunstein, A., Mulet, R., Pagnani, A., 2008. Estimating the size of the solution space of metabolic networks. BMC Bioinformatics 9, 240.
- Cascante, M., Curto, R., Sorribas, A., 1995. Comparative characterization of the fermentation pathway of Saccharomyces cerevisiae using biochemical systems theory and metabolic control analysis: steady-state analysis. Math. Biosci. 130, 51–69.
- Causton, H.C., Ren, B., Koh, S.S., Harbison, C.T., Kanin, E., Jennings, E.G., Lee, T.I., True, H.L., Lander, E.S., Young, R.A., 2001. Remodeling of yeast genome expression in response to environmental changes. Mol. Biol. Cell 12, 323–337.
- Chachuat, B., Singer, A.B., Barton, P.I., 2006. Global methods for dynamic optimization and mixed-integer dynamic optimization. Ind. Eng. Chem. Res. 45, 8373–8392.
- Chang, Y.J., Sahinidis, N.V., 2005. Optimization of metabolic pathways under stability considerations. Comput. Chem. Eng. 29, 467–479.
- Chou, I.C., Voit, E.O., 2009. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math. Biosci. 219, 57–83.
- Coelho, P.M., Salvador, A., Savageau, M.A., 2009. Quantifying global tolerance of biochemical systems: Design implications for moiety-transfer cycles. PLoS Comput. Biol. 5, e1000319.
- Curto, R., Voit, E.O., Sorribas, A., Cascante, M., 1997. Validation and steady-state analysis of a power-law model of purine metabolism in man. Biochem. J. 324 (Pt 3), 761–775.
- Curto, R., Sorribas, A., Cascante, M., 1995. Comparative characterization of the fermentation pathway of Saccharomyces cerevisiae using biochemical systems theory and metabolic control analysis: model definition and nomenclature. Math. Biosci. 130, 25–50.
- Dayarian, A., Chaves, M., Sontag, E.D., Sengupta, A.M., 2009. Shape, size, and robustness: feasible regions in the parameter space of biochemical networks. PLoS Comput. Biol. 5, e1000256.
- de Atauri, P., Sorribas, A., Cascante, M., 2000. Analysis and prediction of the effect of uncertain boundary values in modeling a metabolic pathway. Biotechnol. Bioeng. 68, 18–30.
- Eisen, M.D., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A. 95, 14863–14868.
- El-Samad, H., Kurata, H., Doyle, J.C., Gross, C.A., Khammash, M., 2005. Surviving heat shock: control strategies for robustness and performance. Proc. Natl. Acad. Sci. U.S.A. 102, 2736–2741.
- Esposito, W.R., Floudas, C.A., 2000. Deterministic global optimization in nonlinear optimal control problems. J. Global Optim. 17, 97–126.
- Garcia-Martinez, J., Gonzalez-Candelas, F., Perez-Ortin, J.E., 2007. Common gene expression strategies revealed by genome-wide analysis in yeast. Genome Biol. 8, R222.

³ It should be noted that ODE models written using other mathematical forms can be recasted into GMA models, increasing the generality of the method presented here.

- Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., Brown, P.O., 2000. Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell 11, 4241-4257.
- Gianchandani, E.P., Oberhardt, M.A., Burgard, A.P., Maranas, C.D., Papin, J.A., 2008. Predicting biological system objectives de novo from internal state measurements. BMC Bioinformatics 9, 43.
- Goodman, C., 2008. Engineering ingenuity at iGEM. Nat. Chem. Biol. 4, 13.
- Grossman, I.E., Biegler, L.T., 2004. Part II. Future perspective on optimization. Comput. Chem. Eng. 28, 1193-1218.
- Guillén-Gosálbez, G., Sorribas, A., 2009. Identifying quantitative operation principles in metabolic pathways: a systematic method for searching feasible enzyme activity patterns leading to cellular adaptive responses. BMC Bioinformatics 10, 386.
- Han, J.D., 2008. Understanding biological functions through molecular networks. Cell Res. 18, 224-237
- Hansen, P., Jaumard, B., Lu, S., 1992. Global optimization of univariate lipschitz functions. I. Surrey and properties. Math. Program. 55, 251–272. Hatzimanikatis, V., Liao, J.C., 2002. A memorial review of jay bailey's contribution in
- prokaryotic metabolic engineering. Biotechnol. Bioeng. 79, 504-508.
- Hatzimanikatis, V., Emmerling, M., Sauer, U., Bailey, J.E., 1998. Application of mathematical tools for metabolic design of microbial ethanol production. Biotechnol. Bioeng. 58, 154-161.
- Hatzimanikatis, V., Floudas, C.A., Bailey, J.E., 1996. Optimization of regulatory architectures in metabolic reaction networks. Biotechnol. Bioeng. 52, 485-500.
- Hlavacek, W.S., Savageau, M.A., 1998. Method for determining natural design principles of biological control circuits. J. Intell. Fuzzy Syst. 6, 87.
- Jenkins, G.M., 2003. The emerging role for sphingolipids in the eukaryotic heat shock
- response. Cell Mol. Life Sci. 60, 701–710. Kashiwagi, A., Urabe, I., Kaneko, K., Yomo, T., 2006. Adaptive response of a gene network to environmental changes by fitness-induced attractor selection. PLoS One 1, e49.
- Kitano, H., 2007. Towards a theory of biological robustness. Mol. Syst. Biol. 3, 137.
- Kitano, H., 2004. Biological robustness. Nat. Rev. Genet. 5, 826-837
- Klipp, E., Nordlander, B., Kruger, R., Gennemark, P., Hohmann, S., 2005. Integrative model of the response of yeast to osmotic shock. Nat. Biotechnol. 23, 975-982.
- Koonin, E.V., 2009. Darwinian evolution in the light of genomics. Nucleic Acids Res. 37.1011-1034.
- Kurata, H., El-Samad, H., Iwasaki, R., Ohtake, H., Doyle, J.C., Grigorova, I., Gross, C.A., Khammash, M., 2006. Module-based analysis of robustness tradeoffs in the heat shock response system. PLoS Comput. Biol. 2, e59.
- Lee, J.M., Gianchandani, E.P., Papin, J.A., 2006. Flux balance analysis in the era of metabolomics. Brief Bioinform. 7, 140-150.
- Marin-Sanguino, A., Voit, E.O., Gonzalez-Alcon, C., Torres, N.V., 2007. Optimization of biotechnological systems through geometric programming. Theor. Biol. Med. Model. 4. 38.
- Martens, H., Veflingstad, S.R., Plahte, E., Martens, M., Bertrand, D., Omholt, S.W., 2009. The genotype-phenotype relationship in multicellular pattern-generating models-the neglected role of pattern descriptors. BMC Syst. Biol. 3, 87.
- Mitchell, A., Romano, G.H., Groisman, B., Yona, A., Dekel, E., Kupiec, M., Dahan, O., Pilpel, Y., 2009. Adaptive prediction of environmental changes by microorganisms. Nature 460, 220-224.
- Molina-Navarro, M.M., Castells-Roca, L., Belli, G., Garcia-Martinez, J., Marin-Navarro J., Moreno, J., Perez-Ortin, J.E., Herrero, E., 2008. Comprehensive transcriptional
- analysis of the oxidative response in yeast. J. Biol. Chem. 283, 17908–17918. Morohashi, M., Winn, A.E., Borisuk, M.T., Bolouri, H., Doyle, J., Kitano, H., 2002. Robustness as a measure of plausibility in models of biochemical networks. J. Theor. Biol. 216, 19-30.
- Nielsen, J., 2007. Principles of optimal metabolic network operation. Mol. Syst. Biol. 3.126
- Nikolaev, E.V., 2009. The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems. Metab. Eng.
- O'Connor, T.D., Mundy, N.I., 2009. Genotype-phenotype associations: Substitution models to detect evolutionary associations between phenotypic variables and genotypic evolutionary rate. Bioinformatics 25, i94-i100.
- Papamichail, I., Adjiman, C.S., 2004. Global optimization of dynamic systems. Comput. Chem. Eng. 28, 403-415.
- Papamichail, I., Adjiman, C.S., 2002. A rigorous global optimization algorithm for problems with ordinary differential equations. J. Global Optim. 24, 1-33.
- Polisetty, P.K., Gatzke, E.P., Voit, E.O., 2008. Yield optimization of regulated metabolic systems using deterministic branch-and-reduce methods. Biotechnol, Bioeng 99, 1154-1169
- Porn, R., Bjork, K.M., Westerkund, T., 2008. Global solution of optimization problems with signomial parts. Discrete Optim. 5, 108-120.
- Pozo, C., Sorribas, A., Jiménez, L., Guillén-Gosálbez, G., submitted for publication. Outer approximation-based algorithm for biotechnology studies in systems biology. Comput. Chem. Eng.
- Quesada, I., Grossman, I.E., 1995. A global optimization algorithm for linear fractional and bilinear programs. J. Global Optim. 6, 39-76.
- Raiford, D.W., Heizer Jr., E.M., Miller, R.V., Akashi, H., Raymer, M.L., Krane, D.E., 2008. Do amino acid biosynthetic costs constrain protein evolution in Saccharomyces cerevisiae? J. Mol. Evol. 67, 621-630.

- Romero-Santacreu, L., Moreno, J., Perez-Ortin, J.E., Alepuz, P., 2009. Specific and global regulation of mRNA stability during osmotic stress in Saccharomyces cerevisiae. RNA.
- Salvador, A., Savageau, M.A., 2006. Evolution of enzymes in a series is driven by dissimilar functional demands. Proc. Natl. Acad. Sci. U.S.A. 103, 2226-2231.
- Salvador, A., Savageau, M.A., 2003. Quantitative evolutionary design of glucose 6phosphate dehydrogenase expression in human erythrocytes. Proc. Natl. Acad. . Sci. U.S.A. 100, 14463–14468.
- Savageau, M.A., Coelho, P.M., Fasani, R.A., Tolla, D.A., Salvador, A., 2009. Phenotypes and tolerances in the design space of biochemical systems. Proc. Natl. Acad. Sci. U.S.A. 106. 6435-6440.
- Savageau, M.A., 1998. Demand theory of gene regulation. I. Quantitative development of the theory. Genetics 149, 1665-1676.
- Savageau, M.A., 1976. Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology. Addison-Wesley, Reading, MA.
- Savageau, M.A., 1975. Optimal design of feedback control by inhibition: dynamic considerations. J. Mol. Evol. 5, 199-222.
- Savageau, M.A., 1974a. Optimal design of feedback control by inhibition. J. Mol. Evol. 4, 139–156.
- Savageau, M.A., 1974b. Genetic regulatory mechanisms and the ecological niche of Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 71, 2453-2455.
- Savageau, M.A., 1971. Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. Nature 229, 542-544.
- Savageau, M.A., 1969a. Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation. J. Theor. Biol. 25, 370-379.
- Savageau, M.A., 1969b. Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. J. Theor. Biol. 25, 365-369
- Schuetz, R., Kuepfer, J., Sauer, U., 2007, Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol. Syst. Biol. 3, 119.
- Sims, K.J., Spassieva, S.D., Voit, E.O., Obeid, L.M., 2004. Yeast sphingolipid metabolism: clues and connections. Biochem. Cell Biol. 82, 45-61.
- Singer, A.B., Barton, P.I., 2006. Global optimization with nonlinear ordinary differential equations. J. Global Optim. 34, 159-190.
- Tawarmalani, M., Sahinidis, N.V., 2002. Convexification and global optimization. In: Anonymous (Ed.), Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications (Nonconvex Optimization and its Applications). Springer.
- Teusink, B., Wiersma, A., Jacobs, L., Notebaart, R.A., Smid, E.J., 2009. Understanding the adaptive growth strategy of lactobacillus plantarum by in silico optimisation. PLoS Comput. Biol. 5, e1000410.
- Torres, N.V., Voit, E.O., Glez-Alcon, C., Rodriguez, F., 1997. An indirect optimization method for biochemical systems: description of method and application to the maximization of the rate of ethanol, glycerol, and carbohydrate production in Saccharomyces cerevisiae. Biotechnol. Bioeng. 55, 758-772.
- Torres, N.V., Voit, E.O., Gonzalez-Alcon, C., 1996. Optimization of nonlinear biotechnological processes with linear programming: Application to citric acid production by Aspergillus niger. Biotechnol. Bioeng. 49, 247-258.
- Trotter, E.W., Berenfeld, L., Krause, S.A., Petsko, G.A., Gray, J.V., 2001. Protein misfolding and temperature up-shift cause G1 arrest via a common mechanism dependent on heat shock factor in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A. 98, 7313-7318.
- Vilaprinyo, E., Alves, R., Sorribas, A., 2006. Use of physiological constraints to identify quantitative design principles for gene expression in yeast adaptation to heat shock. BMC Bioinformatics 7, 184.
- Vital-Lopez, F.G., Armaou, A., Nikolaev, E.V., Maranas, C.D., 2006. A computational procedure for optimal engineering interventions using kinetic models of metabolism. Biotechnol. Prog. 22, 1507-1517.
- Voit, E.O., 2003. Design principles and operating principles: the yin and yang of optimal functioning. Math. Biosci. 182, 81–92. Voit, E.O., 2000. Computational Analysis of Biochemical Systems. A Practical Guide
- for Biochemists and Molecular Biologists. Cambridge University Press, Cambridge, U.K.
- Voit, E.O., Radivoyevitch, T., 2000. Biochemical systems analysis of genome-wide expression data. Bioinformatics 16, 1023-1037.
- Voit, E.O., 1992. Optimization in integrated biochemical systems. Biotechnol. Bioeng. 40.572-582.
- Wagner, A., 2005. Energy constraints on the evolution of gene expression. Mol. Biol. Evol. 22, 1365-1374.
- Watson, T.G., 1970. Effects of sodium chloride on steady-state growth and metabolism of Saccharomyces cerevisiae. J. Gen. Microbiol. 64, 91-99.
- Wiebe, M.G., Rintala, E., Tamminen, A., Simolin, H., Salusjarvi, L., Toivari, M., Kokkonen, J.T., Kiuru, J., Ketola, R.A., Jouhten, P., Huuskonen, A., Maaheimo, H., Ruohonen, L., Penttila, M., 2008. Central carbon metabolism of Saccharomyces cerevisiae in anaerobic, oxygen-limited and fully aerobic steady-state conditions and following a shift to anaerobic conditions. FEMS Yeast Res. 8, 140-154.
- Wilkins, A.S., 2007. Between "design" and "bricolage": genetic networks, levels of selection, and adaptive evolution. Proc. Natl. Acad. Sci. U.S.A. 104 (Suppl. 1), 8590-8596.