10th International Symposium on Process Systems Engineering - PSE2009
Rita Maria de Brito Alves, Claudio Augusto Oller do Nascimento and Evaristo Chalbaud Biscaia Jr. (Editors)

An Outer Approximation Algorithm for the Global Optimization of Regulated Metabolic Systems

Gonzalo Guillén-Gosálbez[a*], Carlos Pozo[a], Laureano Jiménez[a], Albert Sorribas[b]

[a] *Department of Chemical Engineering, University Rovira i Virgili, Tarragona, Spain*
[b] *Departament de Ciències Mèdiques Bàsiques, Universitat de Lleida, Lleida, Spain*

Abstract

Understanding the evolution of cellular metabolism requires a number of techniques able to deal with its complexity. Adaptive responses observed in evolutive studies are expected to consist of an optimal set of changes in enzymes activities fulfilling important physiological constraints. Within this context, we present a novel approach to identify enzyme activity regions that contain feasible biological responses in evolution. The framework presented also allows to optimize the enzyme activity changes required to maximize certain fluxes in biotechnological applications. The method relies on solving nonlinear programming models via global optimization techniques.

Keywords: Optimization, Power-law, Evolution.

1. Introduction

In natural cells, emergence of new designs results from evolution. The adaptive response of the cellular metabolism to different situations is attained by tuning gene expression and enzyme activity. Understanding the evolution of adaptive strategies is an important goal in Systems Biology.

The evolution of adaptive stress responses can be seen as a multi objective optimization problem. In that sense, the observed response represents an optimal (in some sense) combination of changes that ensure appropriate survival in the considered conditions. Evolution results in adaptations that are admissible solutions fulfilling important physiological constraints.

Within this general context, we introduce a novel approach that aims to identify enzyme activity regions containing feasible responses observed in evolution. The method introduced can also be employed to optimize biological systems in biotechnological applications. Our approach focuses on the properties of a particular class of non-linear models, the GMA (Generalized Mass Action) models that are based on the power-law formalism. The proposed algorithm is very efficient for realistic problems. The solutions found would represent the landscape in which evolutive solutions are expected. Comparison of our results and actual data shows the practical usefulness of the proposed method.

2. GMA representation

We shall consider a metabolic network that has p fluxes that can contribute to the change in the concentration of the pool of any of the n internal metabolites:

$$\frac{dX_i}{dt} = \sum_{r=1}^{p} \mu_{ir} v_r \qquad i = 1,...n$$

Here, μ_{ir} is a stoichiometric factor that indicates how many molecules of X_i are produced or used by the process v_r; it is a positive integer if the flux r produces X_i and it is a negative integer if the flux r depletes the pool of X_i. Each velocity can be represented by different functional forms, but, the so-called power-law formalism is one of the most convenient:

$$v_r = \gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}}$$

In this representation, X_j accounts for the concentration of metabolite j, γ_r is an apparent rate constant for flux r, and f_{rj} is the kinetic order of variable X_j in reaction r. Each kinetic order quantifies the effect of the metabolite X_j on flux r and corresponds to the local sensitivity of the rate v_r to X_j evaluated at the corresponding operating point. Using this representation, a Generalized Mass Action (GMA) model is defined as [1]:

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

In this expression, m indicates independent (external) metabolites.

3. Optimization model and solution strategy

Here, we present an optimization framework for GMA systems that will be later on taken as a basis for deriving the feasibility approach, which is the main contribution of this work. Non-linear optimization models based on the power-law formalism were first proposed by Voit [2]. In S-system representations, a transformation to logarithmic coordinates can be applied thus leading to linear optimization models. However, when the problem is represented by a GMA model, this technique cannot be applied.

In general, the problem of identifying the optimal values of v_r, γ_r and X_j that maximize a given criterion and satisfy at the same time the equations involved in the GMA representation can be posed as a nonlinear programming (NLP) as follows:

$$ONLP = \min \; U(v_r, \gamma_r, X_j)$$

$$s.t. \quad \sum_{r=1}^{p} \mu_{ir} v_r = 0 \qquad i = 1,...n$$

$$v_r = \gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}} \qquad r = 1,..., p$$

$$v_r, \gamma_r, X_j \in \Re_+$$

Model **ONLP** corresponds to a non-convex problem. Because of this, standard NLP techniques may get trapped in local solutions that are likely to be far away from the global optimum. This may lead to wrong conclusions when performing biological studies. To circumvent this limitation, we introduce a deterministic algorithm to globally optimize **ONLP** that is based on the works of Bergamini and co-workers [3] and Polisetty et al. [4]. The proposed method relies on hierarchically decomposing the problem into two levels, an upper level master problem **CMILP** and a lower level slave problem **RNLP**, between which the algorithm iterates until a termination criterion is satisfied (see Figure 1).

The master level of the algorithm entails the solution of a mixed-integer linear (MILP) problem, which is a relaxation of model **ONLP** (i.e., it rigorously overestimates the feasible region of **ONLP**), and therefore predicts a valid lower bound on its global optimum. In the lower level, the original problem is locally optimized in a reduced search space (i.e., model **RNLP**), thus yielding an upper bound on its global solution. The upper and lower level problems are solved iteratively until the bounds converge. Due to space limitations, technical details of the main features of the proposed algorithm are omitted. As mentioned before, this method can be employed in biotechnological applications in order to optimize a given bioprocess. In this work, as discussed in section 4, such method is employed to derive a tool to perform feasibility analysis in evolutive studies.

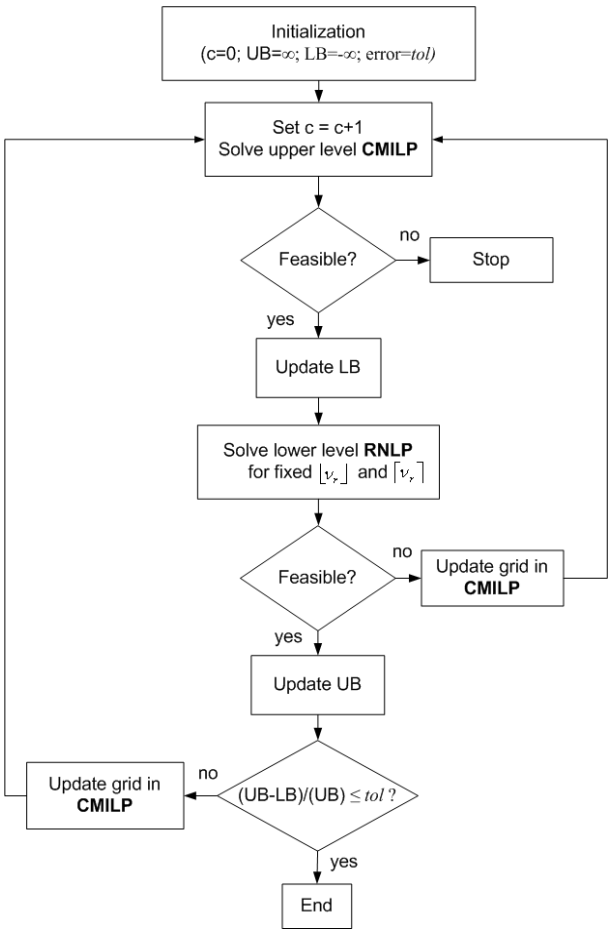

Figure 1. Proposed algorithm.

4. Feasibility approach

The algorithm previously presented can be used, after minor modifications, to identify regions that contain feasible solutions to the original problem **ONLP**, and discard others

in which no single feasible solution exists. Given the metabolic model, the goal is then to find the admissible changes at the level of enzyme activities that are compatible with a set of physiological and functional effective criteria.

From the mathematical point of view, this analysis requires the definition of a set of disjoint sets P_S^q ($P_S^q \cap P_S^{q'} = 0$ for all $q \neq q'$) such that their union contains the feasible space S of **ONLP** ($S \subseteq \bigcup_{q=1,...,Q} P_S^q$). In this work, for the sake of simplicity, we assume that each of these regions P_S^q is a hyper-rectangle described by a set of linear inequalities that impose lower and upper limits ($\lfloor \gamma^q \rfloor$ and $\lceil \gamma^q \rceil$, respectively) on the values of the apparent rate constants γ^q. Thus, we have:

$$P_S^q = \left\{ (v, \gamma, X) \in \Re_+^p \times \Re_+^p \times \Re_+^{n+m} : \lfloor \gamma^q \rfloor \leq \gamma^q \leq \lceil \gamma^q \rceil \right\} \qquad q = 1,...,Q$$

Hence, the feasibility analysis must determine whether these hyper-rectangles contain feasible solutions to **ONLP** or not.

The method devised to accomplish this task is based on the same ideas presented before and comprises two different levels. At the upper level, a master problem is solved to identify a region (i.e., hyper-rectangle) that may contain a feasible solution of **ONLP**. At the lower level, the prediction made by the master problem is checked by solving the original problem in a reduced search space. If a feasible solution is found, then integer cuts are added to the master problem in order to exclude the region containing such a feasible point. Otherwise, the master model is updated by refining its grid, until either a feasible solution is obtained in the lower level or the higher level problem turns out to be unfeasible.

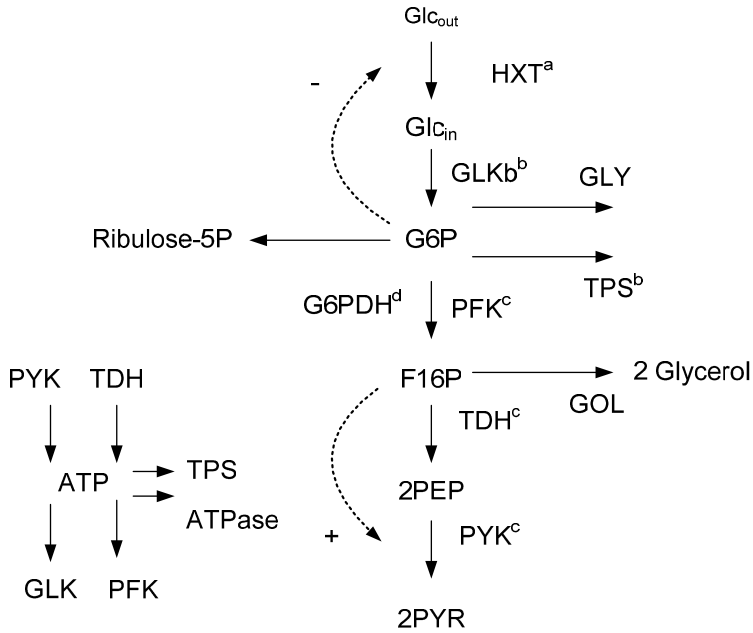

Figure 2. Scheme of the modeled pathways and ranges used for generation of the in silico gen expression profiles (GEPs).

5. Feasible adaptive response of yeast to heat shock

The capabilities of our method were illustrated through its application to the optimal adaptive response of yeast to heat shock (for a detailed description see [5]). Our model includes the core of the glycolytic pathway and the first step of the pentose phosphate pathway. It also accounts for the synthesis of glycogen, trehalose and glycerol, as shown in Figure 2. The notation used in this figure is as follows. Glcout: Extracellular Glucose; Glcin: Intracellular Glucose; G6P: Glucose-6-phosphate; F16P: Fructose-1,6-biphosphate; PEP: Phosphoenolpyruvate; PYR: Pyruvate; HXT: Hexose transporters (HXT1–4, HXT6–8, HXT12); GLK: Glucokinase/Hexokinase (GLK1, HXK1, HXK2); PFK: Phosphofructokinase (PFK1, PFK2); TDH: Glyceraldehyde-3-phosphate dehydrogenase (TDH1, TDH2, TDH3); PYK: Pyruvate kynase (PYK1, PYK2); GLY: Production glycogen; TPS: Trehalose 6-phosphate syntase complex (TPS1, TPS2, TPS3); G6PDH: Glucose 6-phosphate dehydrogenase (ZWF1).

The metabolic network was found to be specially sensitive to changes in two specific enzymes (i.e., PFK and TDH). For this reason, the feasibility analysis was performed on their domain, defining ten different sub-intervals for each of them. Hence, in this particular example, the feasibility analysis focuses on identifying, from the initial set of 100 hyper-rectangles, those containing feasible solutions to **ONLP** and those in which no feasible point exists.

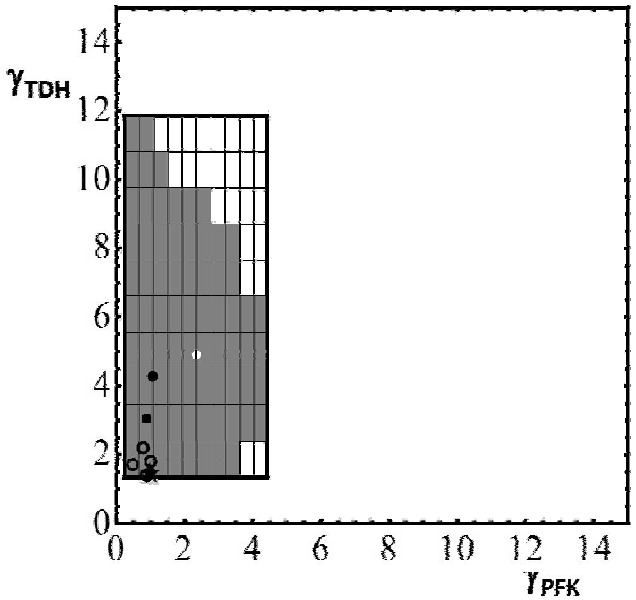

Figure 3. Feasibility analysis. White circle: Maximum rate of ATP synthesis; Black circle: Maximum rate of NADPH synthesis; Black square: Maximum rate of Trehalose synthesis; Black star: Minimum cost; Black empty circles: Experimental observations by Vilaprinyo et al. [5]

The algorithm was implemented in GAMS interfacing with CPLEX and CONOPT as main optimization packages. The total CPU time was less than one minute on an Intel 1.2 GHz machine. Results of this analysis are depicted in Figure 3.

In the figure, shady boxes represent hyper-rectangles that contain at least one feasible solution to the problem, whereas those in white have been proved to be unfeasible. For

comparison purposes, we have also depicted other solutions that are optimal in terms of some criteria: maximum rate of ATP synthesis, maximum rate of NADPH synthesis, maximum rate of Trehalose synthesis and minimum cost. This last metric (i.e., the cost) measures the overexpression of the enzymes.
As can be seen, experimentally observed responses [5] fall within the feasible region predicted by the algorithm. Interestingly, they allocate especially close to the minimum cost solution, that is, the one that would minimize the overexpression of the enzymes. Additionally, the maximum Trehalose rate solution is also near. This probably indicates some importance in the adaptation process.

6. Conclusions

This work introduced a systematic method for identifying the enzyme activity changes that allow a system to meet a set of physiological constraints while optimizing a parameter in the network. The approach presented relies on formulating nonconvex nonlinear problems that are solved via global optimization techniques.
The approach presented was applied to study the optimal adaptive response of yeast to heat shock. Experimental data fall well within the feasible region predicted. The closeness of those points to the minimum cost solution suggests that a conservative strategy where minimum changes are done is the preferred adaptive response. On the computational side, our method proved to be very efficient for medium size problems. The solutions found are intended to shed light on both, biotechnological and evolution studies.

7. Acknowledgments

Financial support received from the Spanish “Ministerio de Educación y Ciencia” (projects DPI2008-04099, PHB2008-0090-PC and BFU2008-00196), the Spanish “Ministerio de Asuntos Exteriores” (projects A/8502/07, HS2007-0006 and A/020104/08) and from “Generalitat de Catalunya” (FI programs) is fully appreciated.

References

Voit EO: Computational Analysis of Biochemical Systems. A Practical Guide for Biochemists and Molecular Biologists. Cambridge, U.K.: Cambridge University Press 2000.

Voit EO: Optimization in integrated biochemical systems. Biotechnol Bioeng 1992, 40(5):572-582.

Bergamini ML, Aguirre P, Grossmann IE: Logic-based outer approximation for globally optimal synthesis of process networks. Computers and Chemical Engineering 2005, 29:1914-1933.

Polisetty PK, Gatzke EP, Voit EO: Yield optimization of regulated metabolic systems using deterministic branch-and-reduce methods. Biotech Bioeng 2008, 99(5):1154-1169.

Vilaprinyo E, Alves R, Sorribas A: Use of physiological constraints to identify quantitative design principles for gene expression in yeast adaptation to heat shock. BMC Bioinformatics 2006, 7:184.