19th European Symposium on Computer Aided Process Engineering – ESCAPE19
J. Jeżowski and J. Thullie (Editors)

A global optimization strategy to identify quantitative design principles for gene expression in yeast adaptation to heat shock

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

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

Abstract

In this paper, we present a new method that is able of identifying the optimal enzyme activity changes that allow a system to meet a set of physiological constraints. The problem is formulated as a nonlinear programming (NLP) model, and it is solved by a novel bi-level global optimization algorithm that exploits its mathematical structure.

Keywords: Optimization, Power-law, Evolution

1. Introduction

The emergence of design in biological systems was a mystery until natural selection was established as the driving force for their evolution. At the molecular level, the identification of design principles in these systems has led to a better understanding of their adaptation. This knowledge may be used in building new gene and metabolic networks that attain specific targets.

Quantitative evolutionary constraints play also an important role in the evolution of biological systems. Once the basic design is in place, the adaptive response of the cellular mechanism to different situations would be attained by tuning gene expression and enzyme activity. Understanding the evolution of adaptive strategies in different conditions is a major 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 restrictions. Those restrictions are the selective pressures over which natural selection works.

Within this general context, we develop a new approach that focuses on identifying optimal enzyme activity changes that satisfy a set of physiological constraints. The method presented allows to identify the possible evolutionary solutions that are expected to contain the actual adaptive response. This general framework focuses on the properties of a particular class of non-linear mathematical representation, 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 allows discussing the practical usefulness of the proposed method.

2. Modeling approach: GMA representation

Our method considers a general metabolic network with p fluxes, each of which can contribute to the change in the concentration of the pool of any of the n internal metabolites. The mathematical representation of such a network is:

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

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, which would include various parameters. From all the available formalisms, the so-called power-law formalism is one of the most convenient. In this formalism, each velocity is represented as:

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

In this representation, X_j accounts for the concentration of metabolite j, r is an apparent rate constant for flux r, and f_{rj} is the kinetic order of variable X_j in re- action 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]:

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

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

3. Mathematical formulation

The method presented relies on formulating a non-linear programming (NLP) model that is solved via global optimization techniques. NLP models based on the power-law formalism were first proposed by Voit [2]. In this context, the use of an S-system representation allows performing a transformation to logarithmic coordinates, so the original model can be converted into a linear formulation. Unfortunately, this technique cannot be generally applied with GMA models.

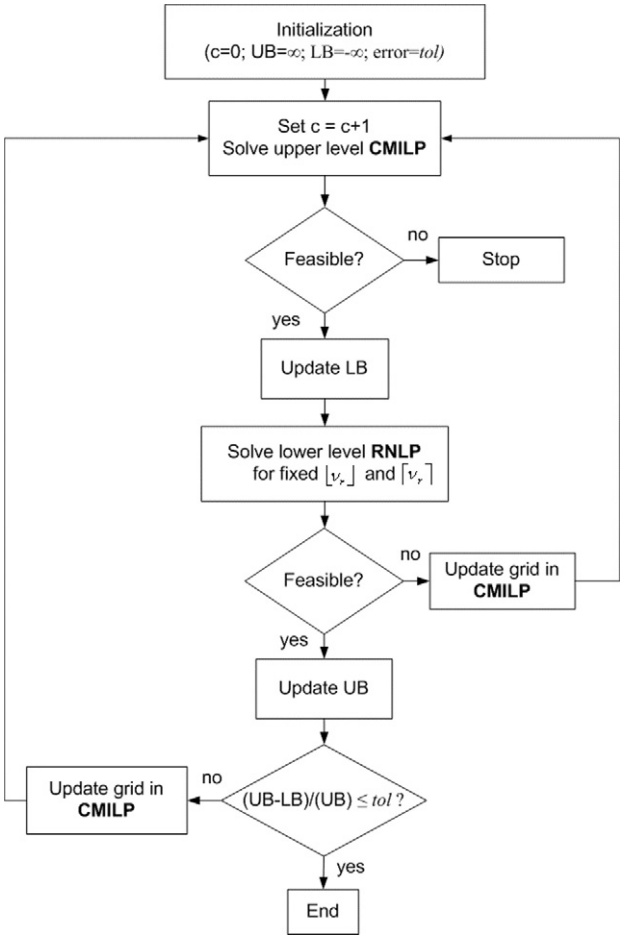

Figure 1. Proposed algorithm.

The task of the NLP problem is to seek those values of v_r, γ_r and X_j that maximize a given criterion and satisfy simultaneously the equations of the GMA representation. Such a model can be expressed as follows:

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

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

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

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

The complexity in solving **ONLP** is given by the non-convexities of the model, which usually give rise to multiple local optima, some of which may be far away from the global optimum. When performing a biological study, this limitation may result in

wrong conclusions as well as low quality predictions. Hence, to circumvent this issue, it is necessary to employ global optimization techniques that ensure the optimality of the solutions found within the desired optimality tolerance. The specific method used in this paper is inspired on the works of Bergamini and co-workers [3] and Polisetty et al. [4], and relies on hierarchically decomposing the problem into two levels, an upper level master problem **CMILP** an a lower level slave problem **RNLP**, between which the algorithm iterates until a termination criterion is satisfied (see Figure 1).

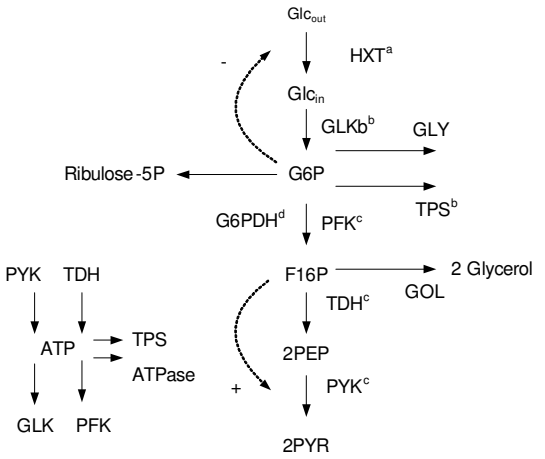

Figure 2. Scheme of the modeled pathways and ranges used for generation of the in silico gen expression profiles (GEPs)

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, 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.

4. Adaptive response of yeast to heat shock

The method presented in this work was applied to study the optimal adaptive response of yeast to heat shock. The model developed 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).

Table 1. Results obtained in the optimization of trehalose, NADPH and ATP.

Maximization objective	TRE	NADPH	ATP

Fluxes (rate of synthesis (mMmin-1)	2.08	46.21	1755.9
Gene expression fold-change with respect to pre-stress situation			
HXT	20.00	20.00	20.00
GLK	0.50	0.50	0.50
PFK	0.25	0.25	20.00
TDH	0.50	20.00	0.25
PYK	20.00	20.00	20.00
GLK-TPS+GOL	20.00	0.25	0.25
G6PDH	0.25	20.00	0.25
Glycerol production	0.25	0.25	0.25
ATPase	0.25	0.25	20.00

Considering the set of constraints identified by Vilaprinyo et al. [5], we ran an optimization procedure for maximizing different objective functions. First, we computed the optimal profile of enzyme activities so that the rate of trehalose synthesis was optimized. Second, we maximized the NADPH production. Finally, we optimized the synthesis of ATP. The mathematical model, which features 15 continuous variables and 43 constraints, was implemented in GAMS and solved with CPLEX 9.0 (master problem) in conjunction with CONOPT (slave problem) on an Intel 1.2 GHz machine (see Table 1). It took less than one CPU second to close a 1% optimality gap in all the cases. The solutions obtained agree with those presented in the literature as they are included in the optimal set identified in [5]. Note that the expected adaptive solution found by natural solution is expected to be close to these solutions, but it does not necessarily have to correspond to the global optimum.

5. Conclusions

A systematic method for identifying the optimal enzyme activity changes that allow a system to meet a set of physiological constraints has been introduced. The approach presented relies on formulating a nonlinear programming (NLP) problem that is solved by a novel bi-level global optimization algorithm.

The capabilities of our modeling framework and solution strategy have been illustrated in the study of the optimal adaptive response of yeast to heat shock. Our method has been proved to provide valuable insight into the evolution of adaptive responses to environmental changes. Furthermore, our strategy can be used in other applications such as the evaluation of parameter changes that are compatible with healthy and disease states.

6. 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 “Generalitat de Catalunya” (FI programs) is fully appreciated.

References

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

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

[3] 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.

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

[5] 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.