Mixed Integer Multiobjective Optimization Approaches for Systems and Synthetic Biology *

Irene Otero-Muras * Julio R. Banga*

* Bioprocess Engineering Group, IIM-CSIC, Spanish National Research Council, 36208 Vigo, Spain (e-mail: {ireneotero,julio}@iim.csic.es).

Abstract: In this work we tackle a number of computational challenges in systems and synthetic biology exploiting optimization based approaches. Our framework combines three important capabilities: multiple optimization objectives (taking into account trade-offs between conflicting goals), simultaneous exploration of topology and parameter spaces (through a mixed integer modeling framework) and high computational efficiency.

We illustrate the capacities of the mixed integer multiobjective framework in three different applications: i) automated design of synthetic bistable genetic switches, ii) exploring design principles underlying biochemical bistable switches in living cells iii) advanced identification of cellular process models from experimental data.

© 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Global optimization, Multiobjective optimization, Gene Regulatory Networks, Design principles, Model Identification, Synthetic Biology, Gene regulatory switch.

1. INTRODUCTION

Optimization methods are widely used in Systems and Synthetic biology for topics ranging from model building, metabolic engineering and biocircuit design (Banga, 2014).

In (Otero-Muras and Banga, 2014) we set up the basis of an optimization framework for the automated design of biocircuits using multiobjective optimization, illustrating the advantages of a multicriteria approach through a number of examples. Importantly, the approach exploits mixed integer nonlinear optimization methods, allowing the simultaneous exploration of parameter and topology spaces.

Within this framework, we tackled two different problems: forward and reverse design of biocircuits. In the context of synthetic biology, we implemented a software tool for the automatic design of biocircuits with predefined target functions from libraries of standard parts, accounting for the gene regulatory dynamics (Otero-Muras et al., 2016). Regarding reverse design, we applied methods to automatically infer design principles of gene regulatory circuits to several relevant functionalities, including biological oscillators (Otero-Muras and Banga, 2016b), pattern formation, adaptation to stimuli and change fold detection (Otero-Muras and Banga, 2017).

Here, we advance the automated design framework further following three directions:

i) Forward design: combining automated design with novel methods for effective multistability detection (Otero-Muras et al., 2017b), to design gene bistable switches from a library of standard parts.

ii) Reverse design: providing insights about the design principles of biological bistable switches.

iii) Advanced identification of models in systems biology: helping to tackle structural uncertainty in the identification of gene regulatory and signaling pathways from experimental data.

2. METHODS

In this section we describe the generalities of our multi-objective mixed-integer optimization (MO-MINLP) framework. Extensions and specificities for each application will be introduced in the following sections.

Mixed-integer modeling framework. First, the dynamics of the biochemical network is encoded in a mixed integer framework. Any kinetics and model granularity are permitted, provided that the following assumptions are satisfied:

A1 A network in the search space is completely characterized by a vector of integer variables (topology) and a set of real variables (parameters).

A2 the dynamics of the system is described by a system of ordinary differential equations of the form:

\[ \dot{z}(t) = f(z, y, x, k), \quad z(0) = z_0 \quad (1) \]

where:

- \( z \in \mathbb{R}^N \) is the vector of state variables (levels or concentrations of the species involved);
- \( x \in \mathbb{R}^R \) is the vector of tunable parameters (real variables);
- \( y \in \mathbb{Z}^M \) is the vector of integer variables containing the model structure/architecture;
\( \bullet \ k \in \mathbb{R}^K \) is a vector of fixed (real) parameters.

**Multiobjective optimization problem formulation.** Once the set of objectives to optimize \( J_i(\dot{z}, z, w, y, k) \) for \( i = 1, \ldots, s \) is defined, the automated design problem is formulated as finding a vector \( w \in \mathbb{R}^r \) of continuous variables and a vector \( y \in \mathbb{Z}^r \) of integer variables which minimize the vector \( J \) of \( s \) objective functions:

\[
\min_{w,y} J_1(\dot{z}, z, w, y, k), J_2(\dot{z}, z, w, y, p), \ldots, J_s(\dot{z}, z, w, y, k)
\]

subject to:

i) the circuit dynamics:

\[
f(\dot{z}, z, w, y, k) = 0, \quad z(t_0) = z_0.
\]

ii) additional requirements in the form of equality and inequality constraints:

\[
h(z, w, y, k) = 0,
\]

\[
g(z, w, y, k) \leq 0,
\]

iii) upper and lower bounds for the real and integer decision variables:

\[
w_L \leq w \leq w_U,
\]

\[
y_L \leq y \leq y_U.
\]

Solving the resultant MO-MINLP optimization problem consists of finding the so-called Pareto front of optimal solutions: a feasible circuit defined by \((w^*, y^*)\) is Pareto optimal if for any feasible \((w^{**}, y^{**})\), \( J(w^*, y^*) \leq J(w^{**}, y^{**}) \) for all \( i = 1, \ldots, s \) with at least one strict inequality (Sendin et al., 2010).

**Numerical solution of the MO-MINLP problem.** The multiobjective optimization problem is reduced into a set of MINLP optimization problems through the \( \varepsilon \)-constraint strategy (Miettinen, 2012), and each MINLP is solved by hybrid optimization methods. Since these problems are multimodal, global optimization approaches are needed. Here we use the eSS metaheuristic (Egea et al., 2010) in combination with the MISQP local solver (Exler and Schittkowski, 2007), since this hybrid approach shows a good efficiency for this type of problems (Otero-Muras and Banga, 2017). It is important to remark that hybrid solvers combine features from deterministic and stochastic methods.

**Definition of the optimization objectives.** One of the crucial aspects in the design is the right selection of objective functions for an efficient search of the desired functionality, such that the target behaviour is obtained at the minimum of the objective function. For example, in a previous work we defined an objective function based on the auto-correlation function which allowed for a very efficient detection of sustained oscillations (Otero-Muras and Banga, 2016b). Here we focus on epigenetic switches, exploiting recently developed bistability conditions to define objective functions for efficient detection of bistability (Otero-Muras et al., 2017b).

3. MINLP-MO FOR AUTOMATED DESIGN OF GENE CIRCUITS IN SYNTHETIC BIOLOGY

One of the challenges in Synthetic Biology is building reliable synthetic gene regulatory circuits with increasing levels of complexity. The MO-MINLP framework, as implemented in SYNBA\(\text{d}m\) (Otero-Muras et al., 2016), allows for the automatic design of genetic circuits with higher complexity, both in the number of regulatory regions and in the kind of tasks that these circuits can accomplish. Starting from a library of standard components and a set of design objectives (both the library and the objective functions can be selected among the built in ones or defined by the user), the algorithm provides the Pareto Front with the circuits that optimally trade-off the design objectives. In bioobjective problems, the knee point indicates a circuit with a good compromise between objectives and is proposed as a criterion for selecting a circuit for implementation (Otero-Muras and Banga, 2017). With this approach we designed, for example, circuits with optimal performance (with respect to a pre-defined target) at a minimal protein burden.

Here we introduce new objective functions for efficient design of transcriptional switches, based on recently developed bistability detection methods (Otero-Muras et al., 2017b). Starting from a library of components, the aim is to find switches that maximize the size of the bistability region, while keeping the ON and OFF states as close as possible to the target. We start from a library with eight different promoters (denoted by \( P_1 \ldots P_8 \)), four transcripts (denoted by \( R_1 \ldots R_4 \)) and two inducers \( IPTG \) and \( aTc \). In particular we search for switches in response to a gradual \( IPTG \) input. We maximize the size of the bistability region, while keeping the protein levels before and after the \( IPTG \) threshold as close as possible to the target (\( L_{ON} \) and \( L_{OFF} \) in Fig 1, respectively). Using SYNBA\(\text{d}m\) (Otero-Muras et al., 2016) we solve the multiobjective optimization problem as introduced in the methods section. In Fig 1 we show the response of the circuit corresponding with best compromise between both objectives.

![Fig. 1. Bistable circuit found by multiobjective optimization (A) and bifurcation diagram (B).](image-url)
4. MO-MINLP FOR INFERENCE OF DESIGN PRINCIPLES OF BIOLOGICAL NETWORKS

One of the goals of Systems Biology is unraveling the design principles of complex biological networks. Optimization can be used to find evolutionary traits, i.e., advantageous features that have been selected through evolution linked with certain functionalities. Since natural circuits are subject to trade-offs, it makes sense to formulate reverse design problems in a multiobjective optimization framework. In previous works, using a MO-MINLP approach we found, for example, architectural properties of transcriptional oscillators linked to optimal trade-offs between the tunability of the period (postulated by Tsai et al. (2008) as an evolutionary trait for a wide range of biological oscillators), and the robustness of the oscillator against molecular noise (Pájaro et al., 2017). We also explored motif architectures linked to other functionalities including rapid adaptation upon stimulus, change fold detection or pattern formation (Otero-Muras and Banga, 2017).

One important attribute of living cells is the capacity to switch between different states in response to internal or external stimuli, and this requires molecular pathways capable of converting graded signals into ON-OFF responses. A widespread switching mechanism relies on bistability, which has been intrinsically linked to positive feedback regulations (Hsu et al., 2016). We often find in nature architectures more complex than strictly needed for a switch. For example, Pfeuty and Kaneko (2009) found that, under certain circumstances, negative feedback, although not necessary for bistability, can confer advantageous properties, when appropriately combined with positive feedback.

Combining our optimization approach with recently developed methods to efficiently detect bistability (Otero-Muras et al., 2017b) one can explore design principles underlying biochemical switches in living cells. Chen (2006) demonstrated a trade-off between the operational speed of a switch and the maintenance of the bistability property, which is found consistent with experimental observations. Starting from a topology superstructure and selecting as objectives the robustness of the bistability property (in terms of the size of the bistability region) and the operational speed, we can search for topologies leading to good trade-offs between both properties (see Fig 2).

5. MO-MINLP FOR ADVANCED MODEL IDENTIFICATION OF BIOSYSTEMS

Working with mathematical models of cellular processes we usually deal with high levels of complexity and uncertainty. Complexity is inherent to biosystems due to their interconnected nature and nonlinear kinetics. Our limited knowledge and capacity of observation of processes occurring in living cells lead to uncertainties, both parametric and structural (including many conflicting mechanistic hypothesis). The advantage of mixed integer approaches is that they allow to explore simultaneously parameter and topology spaces, which is much more efficient than estimating parameters for each of the candidate architectures (Rodriguez-Fernandez et al., 2013). Here we tackle two different problems:

A. Find topologies compatible with observed qualitative behavior. Mixed integer optimization approaches can be used to infer relevant information about cellular pathways from qualitative experimentally observed behaviour, even with limited knowledge about the underlying mechanisms and scarce quantitative experimental data. Here we re-write a model of adipogenesis proposed by Park et al. (2012) in mixed integer framework and formulate an optimization problem with 3 binary, 3 integer and 26 real decision variables to find conditions under which the system leads to an irreversible switch, in agreement with the irreversible transition observed from the preadipocytes into adipocytes capable of lipid accumulation. The 3 binary variables account for the presence or not of three feedback loops (illustrated in Fig 3 A), and the integer variables are levels of cooperativity. Since we want to ensure that the switch is irreversible we impose by means of the lower and upper bounds that one of the bifurcation occurs at a negative value of the control parameter. Through optimization we obtained results coherent with the analysis by Park et al. (2012) (based on stochastic simulations), in terms of levels of cooperativity and active feedbacks loops. In Fig 3 B we show the bifurcation diagram for the adipogenesis model with 3 active feedback loops ($y_1 = y_2 = y_3 = 1$) and the minimum cooperativity levels found for bistability ($n_1 = 2, n_2 = 3, n_3 = 3$ are the Hill coefficients corresponding to the interactions indicated in Fig 3 A).

B. Find circuits compatible with a set of quantitative experimental data. There is an extensive work being done to tackle mechanistic uncertainty in model development in systems biology (Sumaker et al., 2013). Here we take as a case study the system proposed by May (2010) to explore the interplay of the transcription factors (TF) GATA1, GATA2 and PU1 during erythroid differentiation. They start from a TF network architecture with a number of unknown connections, and, for each of the 32 possible architectures, solve a parameter optimization problem to find the best fit to the experimental data (time course expression levels of GATA1, GATA2 and PU1). With this analysis they could find a set of architectures providing a good fit, and discard approximately a half of the architectures (for which it was impossible to find parameter sets providing good fits). Applying our mixed integer optimization within a multistart strategy, we find the same
Fig. 3. Adipogenesis (A) molecular model proposed by Park et al and (B) corresponding bifurcation diagram showing an irreversible switch.

good architectures with much less computational effort. In Fig 4 we illustrate one of the obtained architectures with the corresponding fit. In this case, all the candidate structures show the same complexity. In case of candidates with different complexities, an additional objective like the Akaike criterion can be considered within a multiobjective framework.

Fig. 4. Erythroid differentiation, motif basal superstructure (A) and corresponding data fit (B).

REFERENCES


