Preprint, 11th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems
June 6-8, 2016. NTNU, Trondheim, Norway

Exploring Design Principles of Gene Regulatory Networks via Pareto Optimality ⋆

Irene Otero-Muras * **Julio R. Banga** *

* *BioProcess Engineering Group, IIM-CSIC, Spanish Council for Scientific Research, Eduardo Cabello 6, 36208 Vigo, Spain (e-mail: ireneotero@iim.csic.es, julio@iim.csic.es).*

Abstract: One central problem in systems and synthetic biology is to characterize the biological functions of regulatory network motifs. Here we consider recent model-based exploration approaches used to identify motifs capable of performing a specific biological task. In this work, we propose an optimization based strategy where the motivation is twofold: on the one hand, to introduce efficiency and optimality in the search, by using global mixed integer nonlinear optimization methods. On the other hand, to incorporate multiple design objectives (Pareto optimality), in order to cope with realistic trade-offs observed in nature. The potential of this approach is illustrated through an example where we explore the design principles underlying stripe-forming motifs.

Keywords: Global Optimization, Gene Regulatory N'etwork, Synthetic Biology, Systems Biology, Multiobjective Optimization

1. INTRODUCTION

A gene regulatory (or transcriptional) network consists of a collection of DNA segments and their interactions which together regulate biological functions by controlling the expression levels and temporal patterns in which gene products appear (Karlebach and Shamir, 2008). Transcriptional networks in living cells are complex, and one of the challenges of systems biology is to uncover their structural design principles.

A gene regulatory network can be described by a graph, where nodes correspond to genes, and edges indicate the transcriptional regulation of one gene by the protein product of another gene. Milo et al. (2002) developed an algorithm to detect patterns of interconnections occurring in real networks more often than in randomized ones, identifying a first set of *network motifs* or basic building blocks. To understand how specific functional outcomes or cellular behaviours emerge from particular interactions of genes and proteins, increasing effort is devoted to analyze the functionalities of these motifs and their interconnections (Ingram et al., 2006).

Different analytic and numeric approaches make use of dynamic models of biochemical networks to explore the mappings between the spaces of topologies and parameters and the space of model behaviour (Otero-Muras et al., 2014). The goal is to find patterns, structural and/or parametric features associated to biological functions, like the capacity for bistability and oscillations (Mincheva and Craciun, 2008; Otero-Muras et al., 2012), adaptive responses (Ma et al., 2009) or the formation of stripes (Munteanu et al., 2014; Rodrigo and Elena, 2011). Numerical approaches are usually based on the exhaustive exploration of the topology spaces. Cotterell and Sharpe (2010) proposed to link topologies together into a non-directed graph based on topological similarity, and then analyze the shape of a resultant *complexity atlas* to determine the *core topologies* for a given function.

In this work, we propose an optimization based approach to find patterns capable of specific biological tasks. In contrast to exhaustive exploration, which is computationally expensive and becomes practically unfeasible for increasing levels of complexity, our method aims to introduce efficiency in the search, exploiting the potential of Global Mixed Integer Programming solvers (Otero-Muras and Banga, 2014).

Moreover, our approach is multiobjective, allowing not only to find circuits with a specific functionality, but optimally performing with respect to a set of predefined criteria. This is motivated by the fact that the levels of complexity found in biological circuits cannot be explained by the accomplishment of a given function alone, suggesting multiple simultaneous (and potentially conflicting) goals. For example, even if a simple negative feedback is enough to generate oscillations, many oscillators found in nature contain both negative and positive feedback loops. Tsai et al. (2008) demonstrated through a computational study a number of advantages conferred by the presence of positive feedback in oscillators, namely period tunability, improved robustness and reliability.

Here we will also illustrate how the usefulness of multicriteria optimization in synthetic biology design goes beyond obtaining a set of optimal trade-offs (Pareto front). We

⋆ We acknowledge funding from the Spanish MINECO (and the European Regional Development Fund) project SYNBIOFACTORY (grant number DPI2014-55276-C5-2-R).

will show how the analysis of those Pareto solutions can lead to a more systematic inference and understanding of the underlying design principles. This concept is somewhat similar to the automated innovization approach recently used in engineering design (Deb et al., 2014). Innovization attempts to extract innovative design principles through analysis of optimization results.

As a proof of concept for our method we consider the problem of finding stripe forming motifs in 3-gene configurations, and compare our results with previously published studies based on exhaustive search (Munteanu et al., 2014).

2. METHODS

2.1 Modeling framework

Gene regulatory networks are represented as directed graphs, with nodes corresponding to genes, and edges indicating their interactions. One arrow from gene A to gene B indicates the transcriptional regulation of B by the transcription factor encoded by A.

The dynamics of a gene regulatory network can be described through a system of Ordinary Differential Equations representing the mass balances of the species involved. Detailed models might include promoters, RNA polymerase, mRNA, proteins and complexes among species. Simpler models take into account time scale separation, lump transcription and translation into a single step (Kepler and Elston, 2001) and consider only the levels of the transcription factor proteins encoded by the network genes. In this way, for a n-gene network, the state vector of the ODE model, $z(t) \in \mathbb{R}^n$ contains the levels of the n proteins at time t.

In this work we use a connectionist model (Mjolsness et al., 1991) to describe gene regulation. This model is biologically-verified and extensively employed in the study of developmental gene networks (Munteanu et al., 2014). Within this framework the regulation from gene G_i to gene G_j is characterised by two numbers: an integer $y_{ij} \in \{-1, 0, 1\}$, coding for inhibition (-1), no action (0), and activation (1), and a strictly positive weight $w_{ij} \in \mathbb{R}_{>0}$. We can construct two matrices $Y \in \mathbb{Z}^{n\times n}$ and $W \in \mathbb{R}_{>0}^{n\times n}$ containing respectively the gene-gene interaction indices and the weights.

The effective regulating input to a gene G_i is given by:

$$\chi_i = \sum_{j=1}^{n} \omega_{ji} z_j + \alpha_i I$$

where $\omega_{ji} = y_{ji} w_{ji}$, and the term $\alpha_i I$ reflects the effect of external inputs (in case the gene G_i is only affected by internal gene-gene interactions, the coefficient $\alpha_i = 0$). The transcription rate is proportional to the sigmoidal-filtering of the total contribution, such that the balance for the protein z_i encoded by G_i reads:

$$\dot{z}_i = \frac{1}{1 + exp(a - b(\chi_i))} - \delta z_i \qquad (1)$$

where parameters a and b control the steepness and location of the threshold value of the regulation function, and δ is the protein degradation rate constant.

As an example, we consider the three gene network in Fig. 1 with genes A, B and C, where the net internal interaction matrix is given by:

$$\Omega = \begin{pmatrix} 0 & 0 & 0 \\ \omega_{AB} & \omega_{BB} & \omega_{CB} \\ \omega_{AC} & \omega_{BC} & \omega_{CC} \end{pmatrix}$$

and the gene A is induced by an external input I. The ODE system describing the dynamics of the network reads:

$$\begin{aligned} \dot{A} &= \frac{1}{1 + exp(a - b(I))} - \delta A \\ \dot{B} &= \frac{1}{1 + exp\left(a - b(\omega_{AB}A + \omega_{BB}B + \omega_{CB}C)\right)} - \delta B \\ \dot{C} &= \frac{1}{1 + exp\left(a - b(\omega_{AC}A + \omega_{BC}B + \omega_{CC}C)\right)} - \delta C \end{aligned} \qquad (2)$$

Configurations with $\omega_{AB}\omega_{CB}\omega_{BC} < 0$ give rise to incoherent feedforward loops. In Fig. 1, the incoherent feedforward loop of type one IFF1 ($\omega_{AB} > 0, \omega_{AC} > 0$ and $\omega_{BC} < 0$) and the incoherent feedforward loop IFF3 ($\omega_{AB} < 0, \omega_{AC} > 0$ and $\omega_{BC} > 0$) are depicted.

Starting from this model (Fig. 1) Munteanu et al. (2014) investigated in a recent work 3-gene configurations capable of translating a morphogen gradient into a single stripe pattern. They considered a monotonically increasing input along a one-dimensional tissue of N isogenic cells, i.e., N circuits with the same values of Y, W and the parameters a and b, and only varying the input I.

They found two incoherent feedforward motifs (IFF1 and IFF3) as core topologies for stripe formation.

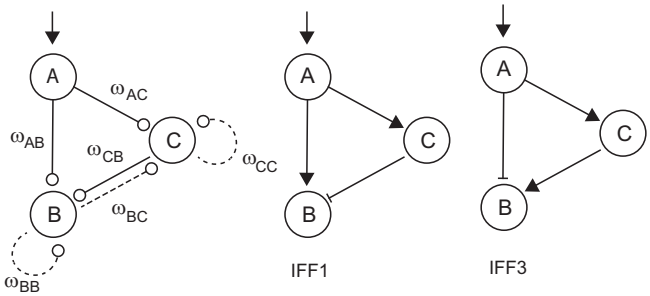

Fig. 1. Graph of the connectionist 3-gene model with gene A induced by an external input. The middle and right circuits correspond to incoherent feedforward loops of type 1 and 3 respectively.

2.2 Mixed Integer Non Linear Programming formulation

The search of a circuit performing a specific behaviour through the space of n-gene circuits can be formulated as an optimization problem.

Within the modeling framework previously described, a circuit (structure and parameters) can be characterized by two vectors: $w \in \mathbb{R}_{>0}^{r}$ containing the weights (its elements are taken columwise from W), and a vector of integer variables $y \in \mathbb{Z}^r$ determining the interactions (its elements are taken columwise from Y). Parameters that are fixed in the ODE model describing the network dynamics are included in a vector $k \in \mathbb{R}^k$.

We can encode the specific desired performance of the circuit by a suitable function $J_1(\dot{z}, z, w, y, k)$, such that

it reaches a minimum when the desired functionality is achieved (importantly, the solution does not need to be unique). As we said in the introduction, we are interested in circuits which, in addition to a given predefined behaviour (oscillations, switch-like behaviour, pulse generation...) perform optimally with respect to additional criteria such as protein production cost and/or robustness against perturbations.

Once we define the set of objectives $J_i(\dot{z}, z, w, y, k)$ for $i = 1, \ldots, s$, a multiobjective MINLP optimization 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:

$$\min_{w,y} \quad J_1(\dot{z}, z, w, y, k), J_2(\dot{z}, z, w, y, p), \ldots, J_s(\dot{z}, z, w, y, k) \tag{3a}$$

subject to:

i) the circuit dynamics in the form of ODEs with the state variables z and additional parameters k:

$$f(\dot{z}, z, w, y, k) = 0, \quad z(t_0) = z_0, \tag{3b}$$

in case of a tissue of N isogenic cells, we consider

$$f_i(\dot{z}_i, z_i, w, y, k) = 0, \quad z_i(t_0) = z_{i_0}, \text{for } i = 1, \ldots, N \tag{3c}$$

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

$$h(z, w, y, k) = 0, \tag{3d}$$

$$g(z, w, y, k) \leq 0, \tag{3e}$$

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

$$w_L \leq w \leq w_U, \tag{3f}$$

$$y_L \leq y \leq y_U. \tag{3g}$$

In order to evaluate the solutions of the multiobjective optimization problem, we need to introduce the notion of Pareto optimality (Miettinen, 2012; Sendin et al., 2010). Given two pairs (w^*, y^*), (w^{**}, y^{**}), we say that the vector $J(w^*, y^*)$ dominates $J(w^{**}, y^{**})$ if $J(w^*, y^*) \leq J(w^{**}, y^{**})$ for all $i = 1, \ldots, s$ with at least one strict inequality. A feasible circuit defined by (w^*, y^*) is a Pareto optimal solution of the multiobjective optimization problem if it is not dominated by other feasible circuits. The set of all Pareto optimal solutions is known as the Pareto front.

Here, we formulate the search for stripe forming motifs a Multiobjective MINLP optimization problem.

First, we define an objective function which captures appropriately the target functionality, in this case, the formation of a stripe in response to a monotonically increasing/decreasing input along the tissue. In Fig. 2 the desired response (a pulse in the level of protein B) is depicted. In order to quantify the quality of the pulse we compute the area between the B level curve and the x-axis (cell index) in three different regions with areas $R1$, $R2$ and $R3$, with maximum value denoted by $\overline{R1}$, $\overline{R2}$ and $\overline{R3}$ respectively. We consider as a perfect stripe the pulse depicted in Fig. 2, for which the function:

$$J_1(w, y) = -((\overline{R1} - R1)/\overline{R1} + R2/\overline{R2} + (\overline{R3} - R3)/\overline{R3})/3 \tag{4}$$

reaches its minimum value, $J_1 = -1$.

As a second optimization goal, we consider the protein production cost, which we assume proportional to the steady state levels of C over the one dimensional tissue:

$$J_2(w, y) = -\sum_{i=1}^{N} C_{ss}(i) \tag{5}$$

Economy in the production of proteins has been considered to be relevant in regulatory systems (Zaslaver et al., 2004).

The topology and parameters of 3-gene circuit configurations according to Fig. 1, are determined by:

i) 6 continuous variables: $w_{AB}, w_{BB}, w_{CB}, w_{AC}, w_{BC}$ and w_{CC};
ii) 6 integer variables: $y_{AB}, y_{BB}, y_{CB}, y_{AC}, y_{BC}$ and y_{CC} in $\in \{-1, 0, 1\}$.

which are the decision variables of the optimization problem.

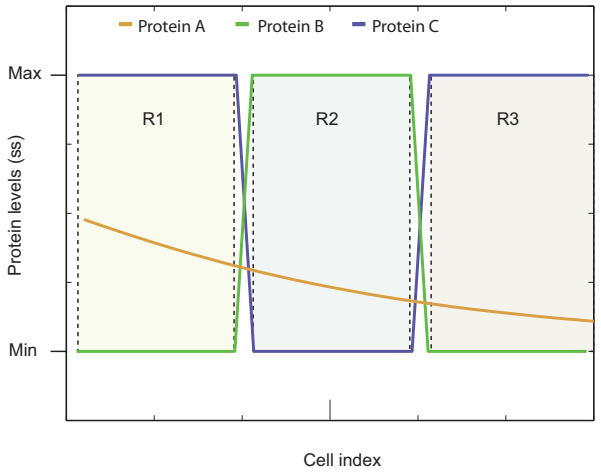

Fig. 2. Stripe in the level of B protein as a response to a monotonically decreasing input. For a perfect stripe, the area $R2$ reaches its maximum value, while the areas $R1$ and $R3$ are zero.

2.3 Solution of the MO-MINLP problem

The dynamics of gene regulatory networks are highly nonlinear, and this leads to an optimization problem that is non convex and multi-modal. Therefore, global optimization methods are required. Moreover, the search spaces can very large, and combine real and integer (and/or binary) variables.

Deterministic optimization approaches ensure convergence to the global optimum within a desired tolerance, but the computational burden is in general very high. Stochastic methods, by the contrary, offer no guarantee of convergence to the global minimum in a finite number of iterations but can provide good solutions in reasonable computational times. In this context, hybrid global MINLP solvers have been shown to be efficient for the design of gene regulatory networks (Otero-Muras and Banga, 2014).

In particular, three hybrid methods combining stochastic global search with the local mixed-integer sequential quadratic programming (`MISQP`) by Exler and Schit-

tkowski (2007) were successfully used for biocircuit design: enhanced scatter search `eSS` by Egea et al. (2010), mixed-integer tabu search (`MITS`) by Exler et al. (2008) and mixed-integer ant colony optimization (`ACOmi`) by Schlueter et al. (2009).

Due to the high nolinearity of the transcription network models, and the existence of integer decision variables, the expected Pareto front can be discrete and possibly non-convex.

In this work we use the ε-constraint strategy (Miettinen, 2012; Sendin et al., 2010), where the MOP is reduced to a a number of MINLP, and each MINLP is obtained by minimising one of the objectives and converting the rest of criteria to inequality constraints. Different solutions can be obtained by changing the upper bounds on the objectives not minimised.

This methodology has two important advantages in the context of gene regulatory networks: the methodology works well for discrete and non-convex Pareto fronts and, in addition, it allows exploiting the MINLP solvers introduced above. The proposed optimization process is composed of the following steps, considering two objective functions J_1 and J_2:

1. Search for the optima of each of the individual objectives:
$$(w_1^*, y_1^*), \quad (w_2^*, y_2^*).$$

2. Compute the individual objective bounds as:
$$\underline{J_1} = J_1(w_1^*, y_1^*), \ \overline{J_1} = J_1(w_2^*, y_2^*),$$
$$\underline{J_2} = J_2(w_2^*, y_2^*), \ \overline{J_2} = J_2(w_1^*, y_1^*).$$

3. Select the objective function to be minimized, denoted in what follows as the primary objective (without loss of generality let us take J_1 as the primary objective).

4. For the non-minimized objective J_2, generate a vector
$$\varepsilon = [\varepsilon_1, \ldots, \varepsilon_i, \ldots, \varepsilon_m]$$
such that $\varepsilon_1 \leq \underline{J_2}$, $\varepsilon_m \geq \overline{J}_2$ and $\varepsilon_1 < \varepsilon_2 < \ldots < \varepsilon_m$.

5. Solve the MINLP:
$$\min_{w,y} \quad J_1(\dot{z}, z, w, y, k)$$
subject to:
$$\varepsilon_k \leq J_2(\dot{z}, z, w, y, k) < \varepsilon_{k+1}$$
for $k = 1, \ldots, m-1$ by means of a MINLP solver.

6. Evaluate the solutions obtained and construct the Pareto front with the non dominated optimal ones.

For the stripe forming motifs case study we consider the score of the pulse as the primary objective (J_1), and the protein production cost as the secondary objective (J_2).

First, we solve the MINLP for the objective function J_1 as defined in Eq. (4), in order to obtain the optimum (w_1^*, y_1^*). Then, we take into account the second objective J_2 in Eq. (5) and search for the circuit which produces a stripe at a minimum cost, in order to get (w_2^*, y_2^*).

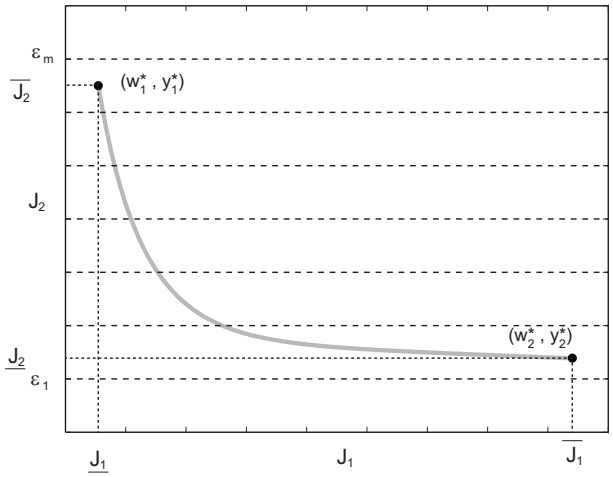

Fig. 3. Scheme of the ε-constraint strategy for a biobjective problem with primary objective J_1.

In second place, we compute the values of the protein cost at the two optima, obtaining the individual objective bounds $\underline{J_2} = J_2(w_2^*, y_2^*)$ and $\overline{J_2} = J_2(w_1^*, y_1^*)$.

Finally, we define a grid for the objective J_2, and minimize the stripe score J_1 in each interval, solving the corresponding constrained MINLP.

2.4 Innovization

Once the set of optimal trade-offs is obtained, we analyze the Pareto front of non-dominated solutions for a systematic inference of design principles, in the same vein of automated innovization approaches recently used in engineering design (Deb et al., 2014).

3. RESULTS AND DISCUSSION

Here we present the results of our approach applied to search for stripe forming motifs. The goal is to explore design patterns in 3-gene regulatory networks conferring to a tissue of N isogenic cells the capability to form a stripe of gene expression in response to a morphogen gradient. As in (Munteanu et al., 2014), the morphogen gradient is simulated by an input into gene A, which takes the form $I = Md^c$, with M being the morphogen concentration at the left boundary of the tissue, d is the reduction of morphogen concentration in each subsequent cell of the morphogen gradient and c is the increasing cell index $c = 1, 2, \ldots, N$ (we consider $M = 5$, $d = 0.982$, and $N = 30$). In order to obtain the system response to the morphogen gradient, we compute the steady state levels of the proteins in every cell starting by Eq. 2 with the corresponding value of c and initial condition $A(0) = B(0) = C(0) = 0.1$. The profiles of the proteins A, B and C are obtained by plotting the corresponding steady state levels as a function of the cell index c as depicted in Fig. 2.

Here we assume for simplicity no autoregulation of B and C. The topology of the network is determined by 4 real variables w_{AB}, w_{CB}, w_{AC} and w_{BC}, and 4 integer variables y_{AB}, y_{CB}, y_{AC} and $y_{BC} \in \{-1, 0, 1\}$.

Solving the MO-MINLP problem. Taking these variables as decision variables for our optimization problem,

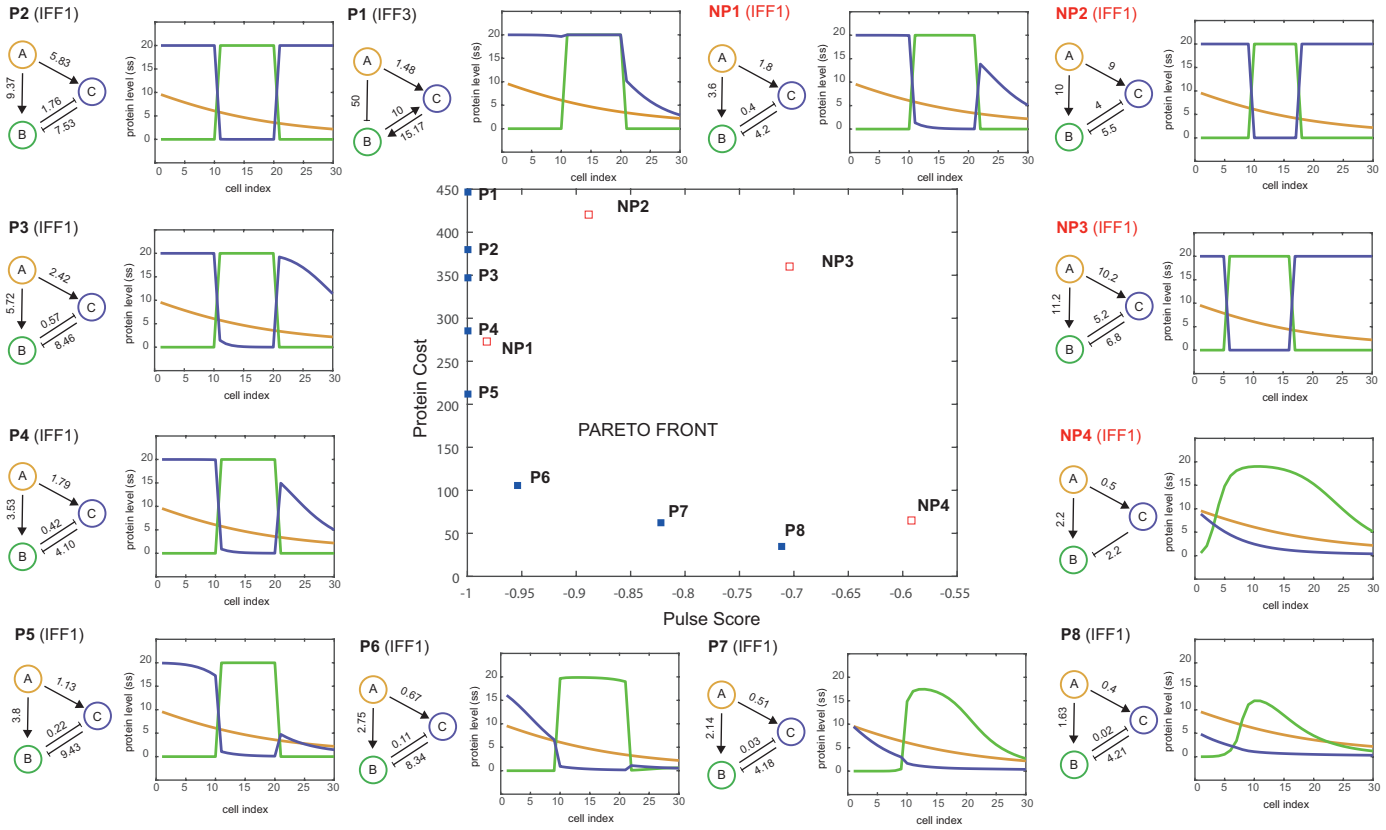

Fig. 4. Pareto Front obtained as a solution of the multiobjective MINLP problem (P1-P8). The corresponding circuit (topology and parameters) and output signal (stripe) is indicated. Circuits NP1-NP4 (Munteanu et al., 2014) correspond to dominated points.

we solve first a single objective MINLP minimizing the stripe score J_1 as defined in Eq. (4). The bounds for the real variables are set to $w_L = 1^{-3}, w_U = 50$. Then, we search for the circuit topology and parameters providing a pulse at the minimum cost. In this way we found respectively the points $P1$ and $P8$ in Fig. 4.

Following the ε-constraint strategy introduced in the previous section, we take the stripe score (J_1) as the primary objective and set a step size of 50 in the secondary objective (protein cost). We solve a constrained MINLP for every interval by means of a global MINLP solver introduced in the previous section (here we have used `eSS` by Egea et al. (2010)), obtaining a set of non-dominated points $P2, \ldots, P7$ in Fig. 4.

The complete Pareto front is illustrated in Fig. 4, where the topology and parameters for each circuit $P1, \ldots, P8$ is depicted together with the associated system's response. For comparison purposes, we also depict four circuits from (Munteanu et al., 2014) corresponding to dominated points $NP1, \ldots, NP4$.

Design principles through innovization approach. By analyzing the set of trade-off optimal solutions, a number of observations arise.

First, we can conclude that there is a trade-off between quality of the stripe and protein production cost defined respectively by Eqs. (4) and (5).

Second, all the circuits in the Pareto front are incoherent feedforward loops. As reported by Munteanu et al. (2014), in existing studies consisting of computational exploration and experimental explorations, incoherent feedforward motifs appear as the minimal structures for single-stripe formation under a morphogen gradient.

Third, IFF3 and IFF1 feedforward structures appear among the non-dominated solutions. This result is in agreement with the results reported by Munteanu et al. (2014), where both analytic and computational approaches were used to establish IFF1 and IFF3 as core topologies for stripe formation.

Finally, the topology of the optimal circuits change following a structured logic as we move along the Pareto front: for high values of protein cost we obtain a IFF3, for medium values of protein cost we obtain IFF1 structures reinforced with negative regulation from B to C, and for low values of protein cost we obtain IFF1 structures with very low regulation from B to C.

4. CONCLUSIONS

In this work we propose an optimization based approach to explore design principles of gene regulatory networks. The approach is multiobjective, exploits the efficiency of global MINLP solvers, and applies the fundamentals of innovization (Deb et al., 2014) for a systematic inference and understanding of design principles from trade-off solutions. As a proof of concept, we have applied our

method to the problem of finding stripe forming motifs in 3-gene structures. On the one hand, the results obtained are coherent with previously published studies (Munteanu et al., 2014) reporting two incoherent feedforward motifs, IFF1 and IFF3, as stripe forming structures. On the other hand, our results elucidate new aspects of stripe forming patterns, revealing how different connectivities affect the trade-off between quality of the stripe and protein production cost. Among ongoing and future work directions, we outline the following:

To consider positive feedbacks (autoregulation of B and C) in the analysis of 3-gene stripe forming motifs (the role of positive feedback in stripe forming motifs has been also studied by Munteanu et al. (2014)), and evaluate their effect in the trade-off optimal set of solutions.

To introduce the effect of stochastic noise, by combining our approach with stochastic simulation of the gene regulatory network dynamics. Munteanu et al. (2014) have shown a preliminary study where autoregulation affected to the border of the stripe as noise increase.

To explore stripe forming motifs in 4-gene and higher order configurations, exploiting the capability of the multiobjective MINLP approach for increasing levels of complexity (Otero-Muras and Banga, 2014). Here it is important to remark that, in contrast to exhaustive exploration, our method can efficiently handle medium-large order systems (Otero-Muras and Banga, 2014). Other functionalities of interest in systems and synthetic biology including oscillators, switches and capacity for adaptation, will also be considered.

To incorporate dominated solutions to provide additional useful information in the innovization process (Chichakly and Eppstein, 2013).

ACKNOWLEDGEMENTS

We acknowledge funding from the Spanish MINECO (and the European Regional Development Fund) project SYNBIOFACTORY (grant number DPI2014-55276-C5-2-R)

REFERENCES

Chichakly, K. and Eppstein, M.J. (2013). Discovering design principles from dominated solutions. *IEEE Access*, 1, 275–289.

Cotterell, J. and Sharpe, J. (2010). An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. *Mol Syst Biol*, 6, 425.

Deb, K., Bandaru, S., Greiner, D., Gaspar-Cunha, A., and Tutum, C.C. (2014). An integrated approach to automated innovization for discovering useful design principles: Case studies from engineering. *Applied Soft Computing*, 15, 42–56.

Egea, J.A., Marti, R., and Banga, J.R. (2010). An evolutionary method for complex-process optimization. *Comput Oper Res*, 37, 315–324.

Exler, O., Antelo, L.T., Egea, J.A., Alonso, A.A., and Banga, J.R. (2008). A tabu search-based algorithm for mixed-integer nonlinear problems and its application to integrated process and control system design. *Comput Chem Eng*, 32, 1877–1891.

Exler, O. and Schittkowski, K. (2007). A trust region sqp algorithm for mixed integer nonlinear programming. *Optim Lett*, 1, 269–280.

Ingram, P.J., Stumpf, M.P., and Stark, J. (2006). Network motifs: structure does not determine function. *BMC Genomics*, 7, 108.

Karlebach, G. and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*, 9, 770–780.

Kepler, G.K. and Elston, T.C. (2001). Stochasticity in transcriptional regulation: Origins, consequences and mathematical representations. *Biophys J*, 81, 3116–3136.

Ma, W., Trusina, A., El-Samad, H., Lim, W., and Tang, C. (2009). Defining network topologies that can achieve biochemical adaptation. *Cell*, 138, 760–773.

Miettinen, K. (2012). *Nonlinear multiobjective optimization*, volume 12. Springer Science & Business Media.

Milo, R., Shen-Orr, S., Itkovitz, S., Kashtan, K., Chklovskii, D., and Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, 298, 824–827.

Mincheva, M. and Craciun, G. (2008). Multigraph conditions for multistability, oscillations and pattern formation in biochemical reaction networks. *Proceedings of the IEEE*, 1281–1291.

Mjolsness, E., Sharp, D., and Reinitz, J. (1991). A connectionist model of development. *J Theor Biol*, 152, 429–453.

Munteanu, A., Cotterell, J., Sole, R., and Sharpe, J. (2014). Design principles of stripe-forming motifs:the role of positive feedback. *Sci Rep*, 4, 5003.

Otero-Muras, I. and Banga, J.R. (2014). Multicriteria global optimization for biocircuit design. *BMC Syst Biol*, 8, 113.

Otero-Muras, I., Banga, J., and Alonso, A. (2012). Characterizing multistationarity regimes in biochemical reaction networks. *PLoS ONE*, 7(7), e39194. doi:10.1371/journal.pone.0039194.

Otero-Muras, I., Yordanov, P., and Stelling, J. (2014). A method for inverse bifurcation of biochemical switches: inferring parameters from dose response curves. *BMC Syst Biol*, 8, 114.

Rodrigo, G. and Elena, S.F. (2011). Structural discrimination of robustness in transcriptional feedforward loops for pattern formation. *PLoS ONE*, 6, e16904.

Schlueter, M., Egea, J.A., and Banga, J.R. (2009). Extended ant colony optimization for non-convex mixed integer nonlinear programming. *Comput Oper Res*, 36, 2217–2229.

Sendin, J.O.H., Exler, O., and Banga, J.R. (2010). Multiobjective mixed integer strategy for the optimisation of biological networks. *IET Syst Biol*, 236–248.

Tsai, T.Y.C., Choi, Y.S., Ma, W., Pomerening, J.R., Tang, C., and Ferrell, .J.E. (2008). Robust, tunable biological oscillators from interlinked positive and negative feedback loops. *Science*, 321, 126–129.

Zaslaver, A., Mayo, A., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., MG, S., and Alon, U. (2004). Just-in-time transcription program in metabolic pathways. *Nat Genet*, 36, 486–491.