Combinatorial Optimisation to Design Gene Regulatory Networks

Guillermo Rodrigo[1], Javier Carrera[1] and Alfonso Jaramillo[2]

[1] Inst. Biologia Molecular y Celular de Plantas, CSIC-UPV, 46022 Valencia, Spain
[2] Lab. Biochimie, Ecole Polytechnique, 91128 Palaiseau, France
http://synth-bio.org

Abstract. We propose a methodology to design gene regulatory networks with targeted dynamics based on combinatorial optimisation that poses new challenges for constraints programming. We use genetic programming techniques to evolve from scratch a transcriptional circuit with unconstrained number of genes that works as a logic gate or as an oscillator. Our circuits are defined by a set of non-linear differential equations describing the protein concentrations. At each evolutive step we could add or remove a concentration and its corresponding differential equation. This corresponds to adding or removing a gene. We can also modify the functional form of a differential equation or modify some kinetic parameter. This corresponds to mutation events in the regulatory or coding sequence. We define as the scoring function the distance between the circuit dynamics and the targeted behaviour We explore the space of all possible transcriptional regulation networks, where at each step we would add/subtract new interactions or modify kinetic parameters, to find the optimal circuit with specified system behaviour. We apply our methodology to the design of genetic devices having a desired switching or oscillatory behaviour. Our circuits could be constructed experimentally by assembling biological parts with appropriate kinetic properties. This is not possible in general and the designer, who will only have a small set of available biological parts, will be forced to evolve some of its parts. This introduces a parameter range for each part that it will propagate into an evolvability range for each designed circuit. Those ranges are best described by using constraints and the evolution process could implement model-checking prior to the evaluation of the dynamics.

Keywords: Biological Systems, Regulatory Networks, Genetic Programming, Combinatorial Optimisation.

1 Introduction

One of the most intriguing aspects of networks of complex systems is their temporal dynamics. Very often in complex systems the dynamics does not follow from the network topology. Among the chief examples of complex networks are the genetic transcription networks. The study of those networks has important applications in understanding the circuitry of living systems. There has been a tremendous work on elucidating the network topology of transcription networks [1]. Studies of recurrent network motifs showed that their dynamics could provide useful functions [2,3]. These reverse-engineering studies are very useful to plan the forward-engineering of synthetic circuits. The new development of standardized genetic parts [4] will allow designing much complex networks in a modular way according to some specifications by the assembly of those parts. Usually genetic parts are taken from wild-type organisms. Nevertheless, some experimental work has been performed on building synthetic parts such as promoters with altered operator sites [5,6,7], modified ribosome binding sites [8], or codon-optimized coding regions [9]. The *de novo* design of protein has engineered new coding regions with specified functions and sometimes they have no similarity with any natural sequence [10,11,12]. In addition, most synthetic promoters are regulated by a single transcription factor, but there has also been some work on the design of promoters regulated by two transcription factors [13,14].

The design of artificial genetic networks [15,16,17] has boosted the emerging field of synthetic biology [18]. Still most of the work has been done using rational design techniques, limiting the computational facilities to the solving of dynamical equations. It would be extremely useful to be able to use computational methods to aid in the optimization and design of new circuits. For instance, we could use a catalogue of genetic circuits with optimized transfer functions as educated guesses to aid in the design of a given genetic circuit. Previous work has already used evolutionary methods to design circuits able to oscillate, although they were composed of electronic components [19]. Another work [20] did focus on biological networks, using protein species and a post-translational regulation to design several types of circuits, although this type of regulation is difficult to implement experimentally. Here, we propose to address transcriptional regulatory interactions, neglecting post-translation regulations, to implement genetic networks that could eventually be synthesized.

Our computational algorithm (Genetdes) searches the space of artificial genetic networks to find the optimal circuits with a targeted temporal behaviour [21]. During our simulation, we add or subtract genes, change kinetic constants or the operator-binding logic function at promoters. Each generated circuit is evolved in time and we use the average deviation to an expected temporal function as scoring function. We use Monte Carlo Simulated Annealing [22] method to do the optimization in the space of all possible genetic circuits. Our circuits could be constructed experimentally by assembling biological parts with appropriate kinetic properties. This is not possible in general and the designer, who will only have a small set of available biological parts, will be forced to evolve some of her parts. This introduces a parameter range for each part that it will propagate into an evolvability range for each designed circuit. Those

ranges are best described by using constraints and the evolution process could implement model-checking prior to the evaluation of the dynamics.

2 Methods

2.1. Mathematical model

The dynamics of transcriptional regulatory networks can be depicted by systems of nonlinear first-order ordinary differential equations. We have considered an effective model of protein concentrations for the transcriptional regulations. The dynamics of a transcription factor concentration (Yi) follows the differential equation

$$d[Y_i]/dt = \alpha_i R_i - \beta_i [Y_i] + \gamma_i , \quad (\mathbf{1})$$

where α_i is the transcription-translation rate of gene i, β_i the corresponding degradation rate, and γ_i the basal rate. The function R_i defines the regulatory factor for the promoter of gene i, defined by

$$R_i = 1/(1+([Y_j]/K_{ij})^{nij}) , \quad (\mathbf{2})$$

where K_{ij} is the regulatory coefficient and n_{ij} is the Hill coefficient (chosen positive for repressions and negative for activations) for the transcription factor j.

2.2. Fitness function

We compute the fitness function as the deviation of the circuit dynamics (y) respect to the targeted dynamics (z) as

$$J = \int |y-z| \chi \, dt , \quad (\mathbf{3})$$

where χ is a weighting factor used to only compute a region of interest (e.g., to avoid transients or to impose an oscillatory dynamics). In that way, we construct a minimization problem, where we evolve networks to behave close to the specified dynamics (z).

We use several transfer functions to specify the target behaviour. Each transfer function gives the behaviour of the system for a given input state. In that way, four transfer functions are required to design a circuit working as a logic gate of two inputs, as there are four entries in the corresponding truth table. On the other hand, to design an autonomous oscillator we need just one transfer function. Therefore, for

each transfer function we compute the fitness of the system, and the global fitness function is the sum of all them.

However, the landscape proposed by that fitness function has not large biological referents as these systems have to be robust as well as functional. Thus, we extend the fitness function to

$$F = (1-r)\, J + r\, J' , \qquad (4)$$

where J is the fitness function given by the equation 3, J' is a new term to count the robustness of the system and r is the degree of robustness for our design. In this work we just study the robustness under parameter perturbations. However, further works will study the robustness under topological perturbations, which will give important issues for understanding the evolution of biological systems. Therefore, we compute J' as the average value of all fitness functions (here we compute 10) after perturbing randomly all the model parameters.

2.3. Optimization procedure

We use Monte Carlo Simulated Annealing [22] to optimize transcription circuits in the space of topologies and parameters. We define a mutation operator to evolve the circuit. This operator consists of two moves, both with a probability of occurrence. The first move is in the parameter space. We select randomly a parameter of the model and we perturb it within the corresponding range of values. The second move is in the topology space. There are five possibilities: (i) change the logic function of a binary promoter, (ii) add a new regulation, (iii) remove a regulation, (iv) add a new gene in the circuit or (v) remove a gene from the circuit. We can specify the probabilities to do these moves according to our design purposes. In addition, for convergence purposes, the probability to do a parameter move is taken much higher than the one to do a topology move (e.g., 0.99). In that way, for each evolved topology we explore the parameter space.

In case of no initial network specification, we start from a disconnected circuit where the number of genes of the circuit is equal to the number of inputs plus the number of outputs. We take a Metropolis criterion to accept a mutation, using an exponential cooling scheme. As each mutation only involves a small change in the network it could be possible to obtain an analytical approximation to the dynamics. This would speed up our methodology in at least one order of magnitude.

3 Design of small networks

We have applied our methodology to design genetic devices implementing a given behaviour. We focus in designing small functional modules, which could later be assembled arriving to large and sophisticated networks. We have targeted digital behaviours. Our devices consist on genetic circuits having the concentration of two and one transcription factors as input and output respectively. We have targeted AND,

OR, NAND and NOR gates, and in Fig. 1 we show the designed circuit with AND behaviour. u1 and u2 are the input transcription factors and y is the output corresponding to the concentration of a gene product. To compute the objective function we have averaged the score obtained with each transfer function corresponding to every entry of the truth table. We have evaluated the score by computing (3) during 100 minutes, which provides one order of magnitude more time that the transient needed to attain the steady state. We have computed a score for transfer function and we have averaged it. However, for visualization purposes, we have plotted a temporal dynamics where the input transcription factors concentrations u1 and u2 take all possible Boolean values of a two-input truth table. Inputs can be activators or repressors according to the chosen promoter during the simulation.

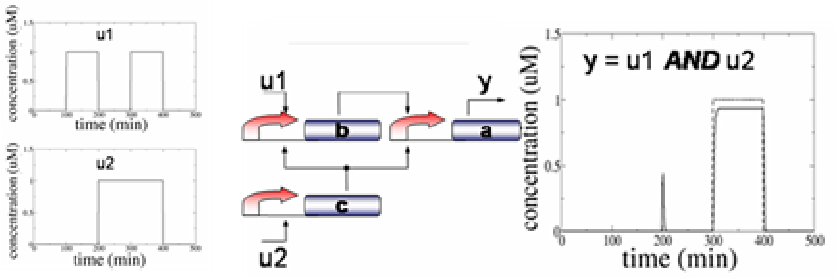

Fig. 1. Transcriptional network composed of three genes (a, b and c) designed to behave as an AND gate. On the left, time evolution of transcription factor inductors u1 and u2 corresponding to (u1, u2)=(0, 0), (0, 1), (1, 0) and (1, 1) for times 0-100, 100-200, 200-300 and 300-400 minutes respectively. On the centre, the circuit obtained with our methodology. On the right, the network dynamics (solid line) superimposed to the targeted behaviour (dashed line).

We have also designed circuits showing an oscillatory behaviour. Towards this end, we have targeted a sinusoidal function. We have considered a weighting factor to compute the score such that it was 1 only in the neighbourhood of a maximum or minimum of the targeted sinusoidal function. This was done to improve the convergence. Fig. 2 shows the optimal genetic network and its time behaviour. We plot as a dashed line the targeted sinusoidal function and as a solid line the corresponding time evolution of the output gene expression. This forward engineering approach has allowed us to design a large set of oscillators and to study evolutionary principles on natural occurring circadian clocks [23]. In addition, we have studied the behaviour of such networks when forcing with external cyclic stimuli at different periods [24].

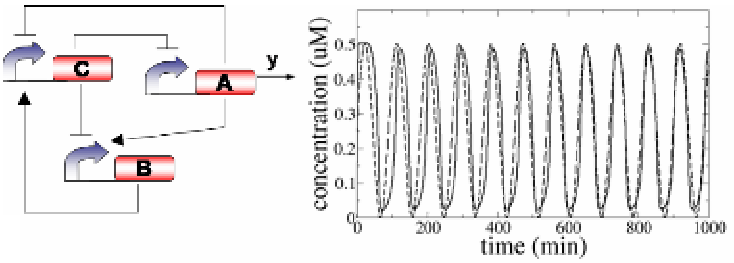

Fig. 2. Transcriptional network designed to show an oscillatory behaviour. The dashed line on the right plot denotes the targeted dynamics.

It is not possible to estimate the complexity of our evolution because it depends on a heuristic optimization process, which will change for each system analyzed. the search throughout the space of genetic networks. Our algorithm writes and reads in SBML level 2 [25], which allows to interface it with a large number of other software. In the species definition, if the specie is an input then its boundary condition will be set to true (false otherwise). Each reaction (transcription-translation) has 1 product and at most 2 reactants. To describe this we need the corresponding kinetic parameters and two additional variables specifying the logical function at the promoter and whether a gene is considered a reporter.

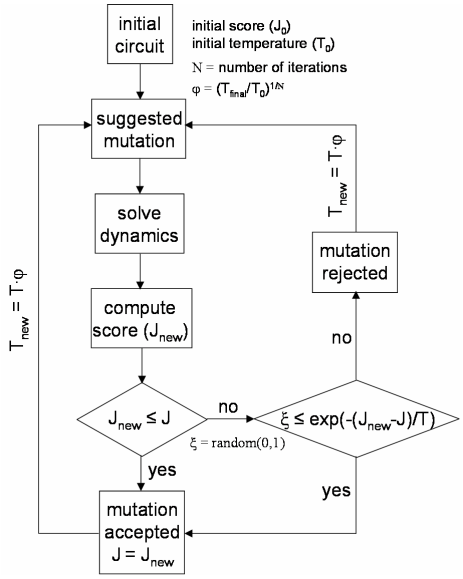

Fig. 3. Scheme showing the flux of Genetdes.

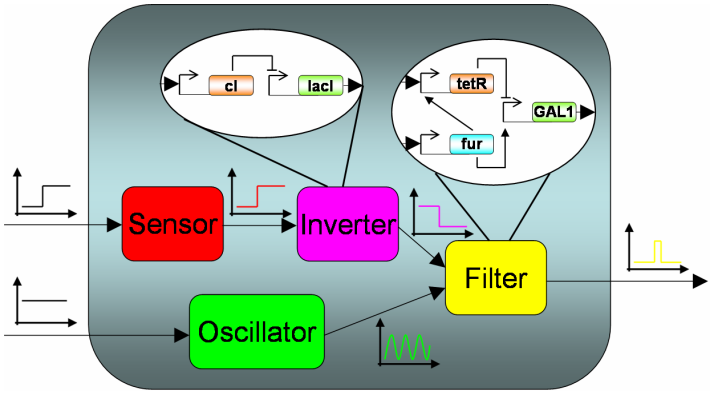

Fig. 4. Example of a design of a complex system using simple functional devices.

4 Discussion

One question to address, from a systems/synthetic biology point of view, is whether natural genetic networks are understandable as systems of devices. Have natural circuits a selective pressure for a given network motif or for a given function? If there could exist a selective pressure for given network modules behaviour, then some circuits within a module could get rewired by evolution while maintaining their functionality. For instance, it could be that some AND circuits would occasionally appear in evolution substituted by another AND circuit. On the other hand, natural gene circuits may not rely on functional modules, but on a complex intertwined network of interactions, as it usually happens with evolutionary design. In this later case, maybe the only way to design a system of devices would be by using an evolutionary design procedure. We could then use directed evolution of gene circuits or in a combination with a computational procedure. In that way, in further work we will expand our methodology to design systems with complex behaviours from a library of functional devices (see Fig. 4).

The implementation of a circuit in a given cellular context usually requires a constant fine-tuning of the model to obtain a successful prototype. In that way, the fact of having a repository of already characterized parts is very useful when implementing a circuit. Therefore, we have developed a software (Asmparts) to assemble part models to construct large systems [26]. We have combined Asmparts with Genetdes to construct and optimize genetic networks.

Acknowledgments. GR acknowledges a graduate fellowship from the Conselleria d'Educacio de la Generalitat Valenciana (BFPI 2007/160). This work was supported by the EU grant BioModularH2 (FP6-NEST 043340), the Spanish Ministry of

Education and Science (TIN 2006-12860) and the Structural Funds of the European Regional Development Fund (ERDF).

References

1. Babu, M. M. & Teichmann, S. A. (2003) Evolution of transcription factors and the gene regulatory network in Escherichia coli. Nucleic Acids Res. 31, 1234-1244.
2. Mangan, S. & Alon, U. (2003) Structure and function of the feed-forward loop network motif. Proc. Natl. Acad. Sci. USA 100, 11980-11985.
3. Brandman, O., Ferrell Jr., J. E., Li, R. & Meyer, T. (2005) Interlinked fast and slow positive feedback loops drive reliable cell decisions. Science 310, 496-498.
4. Endy, D. (2005) Foundations for engineering biology. Nature 438, 449-453.
5. Buchler, N. E., Gerland, U. & Hwa, T. (2003) On schemes of combinatorial transcription logic. Proc. Natl. Acad. Sci. USA 100, 5136-5141.
6. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J. & Phillips, R. (2005) Transcriptional regulation by the numbers: models. Current Opinion in Genetics & Development 15, 116-124.
7. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kulhman, T. & Phillips, R. (2005) Transcriptional regulation by the numbers: applications. Current Opinion in Genetics & Development 15, 125-135.
8. Basu, S., Mehreja, R., Thiberge, S., Cheng, M. T. & Weiss, R. (2004) Spatiotemporal control of gene expression with pulse-generating networks. Proc. Natl. Acad. Sci. USA 101, 6355-6360.
9. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. (2005) A synthetic multicellular system for programmed pattern formation. Nature 434, 1130-1134.
10. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. Science 302, 1364-1368.
11. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. (2003) Computational design of receptor and sensor proteins with novel functions. Nature 423, 185-190.
12. Jaramillo, A., Wernisch, L., Hery, S. & Wodak, S. J. (2002) Folding free energy function selects native-like protein sequences in the core but not on the surface. Proc. Natl. Acad. Sci. USA 99, 13554-13559.
13. Joung, J. K., Koepp, D. M. & Hochschild, A. (1994) Synergistic activation of transcription by bacteriophage l-cI protein and E. coli cAMP receptor protein. Science 295, 1863-1866.
14. Mayo, A. E., Setty, Y., Shavit, S., Zaslaver, A. & Alon, U. (2006) Plasticity of the cis-regulatory input function of a gene. Plos Biol. 4, e45.
15. Elowitz, M. B. & Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403, 335-338.
16. Gardner, T. S., Cantor, C. R. & Collins, J. J. (2000) Construction of a genetic toggle switch in E. coli. Nature 403, 339-342.
17. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. (2003) Development of genetic circuit exhibiting toggle switch or oscillatory behavior in Escherichia Coli. Cell 113, 597-607.
18. Andrianantoandro, E., Basu, S., Karig, D. K. & Weiss, R. (2006) Synthetic biology: new engineering rules for an emerging discipline. Mol. Syst. Biol. doi:10.1038/msb4100073.
19. Mason, J., Linsay, P. S., Collins, J. J. & Glass, L. (2004) Evolving complex dynamics in electronic models of genetic networks. Chaos 14, 707-715.
20. Francois, P. & Hakim, V. (2004) Design of genetic networks with specified functions by evolution in silico. Proc. Natl. Acad. Sci. USA 101, 580-585.

21. Rodrigo, G., Carrera, J. & Jaramillo, A. (2007) Genetdes: automatic design of transcriptional networks. Bioinformatics 23, 1857-1858.
22. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983) Optimization by simulated annealing. Science 220, 671-680.
23. Rodrigo, G., Carrera, J. & Jaramillo, A. (2007) Evolutionary mechanisms of circadian clocks. Cent. Eur. J. Biol. 2, 233-253.
24. Rodrigo, G., Carrera, J. & Jaramillo, A. (2007) In silico evolution of the oscillatory response under light-dark cycles. Biochimie, in press.
25. Hucka, M., et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19, 524-531.
26. Rodrigo, G., Carrera, J. & Jaramillo, A. (2008) Asmparts: assembly of biological model parts. Syst. Synth. Biol., in press.