

# An Introduction to Control of Synthetic Bio-molecular Systems

Domitilla Del Vecchio

**Abstract**—This tutorial presents an introduction to synthetic biology from a control systems perspective. It provides a description of the main objectives of synthetic biology, of the state of the art in such a field, of enabling technologies, and finally of the main techniques and challenges in the analysis and design of synthetic bio-molecular networks.

## I. I

Biologists have long employed phenomenological and qualitative models in order to help discover the components of living systems and to describe their behaviors. On the other hand, the analysis in living organisms of the dynamical properties of complex molecular reaction networks composed of interacting genes, mRNA, proteins, and metabolites requires a more quantitative and systems-level knowledge. Thus, in recent years the field of *systems biology* has emerged, whose focus is the quantitative analysis of cell behavior, with the goal of unraveling the basic dynamic processes, feedback control loops, and signal processing mechanisms underlying life. Complementary to systems biology is the engineering discipline of *synthetic biology*. The goal of synthetic biology is to extend or modify the behavior of organisms, and control them to perform new tasks [3], [5], [11]. Through the *de novo* construction of simple elements and circuits, the field aims to foster an engineering discipline for obtaining new cell behaviors in a predictable and reliable fashion. In the process, synthetic biology plays a role in improving the quantitative and qualitative understanding of basic natural phenomena. In fact, one approach to the testing of mathematical models of biological systems is to design and construct instances of the system in accordance to hypothesized models. Discrepancies between expected behavior and observed behavior highlight either research issues that need more studying, or knowledge gaps and inaccurate assumptions in models.

While tools from controls and dynamical systems theory, such as systems identification and robustness analysis, have been put to great use in systems biology for the analysis of naturally occurring biological systems, the use of this theory for the design of synthetic biological circuits is still emerging. The pioneering work of several biologists and physicists [4], [6], [10], [12], [18] shows the potential and the need for such tools when tackling the challenges of biological design. The experimental results of [6] and of [18] on a negatively auto-regulated gene agree with the mathematical predictions obtained by using straightforward feedback control analysis. However, more complicated systems such as the oscillators

built in [10] and in [4] do not provide experimental results that match well the theoretical predictions. In particular, intrinsic and extrinsic noise sources [19] seem to disrupt the oscillating behavior of the repressilator [10]. In [4], the oscillations are only damped, which suggests that the parameters of the constructed system may not be inside the theoretically computed range of parameters that guarantee oscillations. From these results, the need emerges for *robust, model based design*.

Control and systems theory have much to offer to synthetic biology. But, conversely, one may look forward to new theoretical advances in control systems inspired by biological research. In particular, the standard control system paradigm of modeling a system as an input/output (dynamic) map may need to be revised when signals are carried by the physical displacement of molecules. Accordingly, the role of inputs and outputs may change when systems are interconnected. For discussion on this topic, the reader is referred to [13], [21], [22], [25].

## II. E T

The discovery of mathematical logic in gene regulation [17] and the early achievements in genetic engineering in the 1970s, such as recombinant DNA technology, set the stage for today's synthetic biology. Recent advances in molecular biology provide the ability to translocate and fuse promoters, operators, binding sites, and genes in almost any fashion on a size-wise-compatible plasmid through a procedure called *cloning*. Most importantly, a key enabler to synthesize DNA in amounts large enough to be used for transfection (or transformation) and for various measurement procedures has been the *Polymerase Chain Reaction* (PCR). This molecular biology technique allows a small amount of DNA to be amplified exponentially [1].

Another key enabling technology has been the development of *in vivo* measurement techniques that allow to measure the amount of protein produced by a target gene. For instance, green fluorescent protein (GFP) is a protein with the property that it fluoresces in green when exposed to UV light. It is produced by the jellyfish *Aequoria victoria*, and its gene has been isolated so that it can be used as a *reporter gene*. Other fluorescent proteins, such as yellow fluorescent protein (YFP) and red fluorescent protein (RFP) are genetic variations of the GFP. The reporter gene is inserted (cloned) into the chromosome, very close to the location of the target gene, so both are controlled by the same promoter. Since the target gene and the reporter gene are transcribed at the same rate, by measuring the intensity of the reporter gene light

emitted one can estimate the concentration of the protein expressed by the target gene.

Just as fluorescent proteins can be used as a read out of a circuit, *inducers* function as external inputs that can be used to probe the system. Inducers function by disabling repressor proteins. Repressor proteins bind to the DNA strand and prevent RNA polymerase from being able to attach to the DNA and synthesize mRNA. Two commonly used inducers are IPTG and aTc. Isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) induces activity of beta-galactosidase, which is an enzyme that promotes lactose utilization, through binding and inhibiting the lac repressor. The anhydrotetracycline (atc) binds the wild-type repressor (TetR) and prevents it from binding the Tet operator.

For engineering a system with prescribed behavior, one has to be able to change the physical component features so as to change the values of the parameters of the model. This is now possible. For example, the binding affinity of a transcription factor to its site on the promoter can be weakened by base pairs substitutions. Protein decay rates can be increased by adding degradation tags at the end of the gene expressing the protein (<http://parts.mit.edu/registry/index.php/Help:Tag>). Promoters can be designed to accept multiple transcription factors (called combinatorial promoters) to implement regulation functions that accept several inputs [2]. This can be attained by combining the operator sites of several simple promoters [4].

### III. S

Enabled by the recent technological developments briefly summarized in Section II, a number of simple synthetic circuits with prescribed behaviors have been designed and built in *E. coli*. Naturally occurring transcriptional networks are very complex, however biologists have been discovering recurrent patterns of interconnection that appear frequently. These patterns are called *motifs* [16]. Thus, synthetic biologists have been focusing mainly on synthetically reproducing these network motifs to study their behavior in isolation with the hope to (1) understand their role and features and to (2) create a number of understood building blocks that can be interconnected to create more complex networks with predictable behavior. At the heart of this approach that constructs more complicated systems starting from simpler building blocks is the concept of *modularity*. This concept is briefly discussed in Section IV.

**A self repressed gene.** Negative autoregulation occurs when a transcription factor represses its own transcription [6]. This system has been fabricated and two major findings resulted from the measurements: negative autoregulation speeds the response time [18], and negative autoregulation promotes robustness to fluctuations in production rates [6]. By using standard linear control theory, one can immediately predict that an increased negative feedback increments the robustness of the equilibrium point with respect to fluctuations. The interesting part is that this result has been confirmed by experiments performed on a synthetic negative feedback loop cloned on a plasmid and then transformed

in a bacterium. This fact is encouraging, as it means that the adopted modeling framework may be good enough to suggest design guidelines for circuitry to be implemented in the biological substrate.

**The toggle switch.** A genetic toggle switch is a bistable system in which reliable switches between the two steady-states are induced through an input signal. Any such genetic toggle switch typically needs particular behavioral characteristics in order to be considered a true “memory component”. First, the toggle switch must exhibit bistability over a wide range of parameter values (transcriptional rates, translational rates, decay constants, etc.) that tend to fluctuate in a living cell. Second, the two steady-states must be highly tolerant of random fluctuations in molecular-species concentrations, so that noise-induced transitions between the two states are virtually non-existent [12]. The design of [12] employs two repressible promoters that mutually repress each other. In [12], an analysis is proposed based on the nullcline shape. Within appropriate parameter ranges, one unstable and two stable steady-states exist. At the stable steady-states, one of the repressors is dominant over the other, while the other repressor is shut down. A switch of the dominance-toggle is induced by externally repressing the dominant repressor so it cannot bind any longer to the target promoter for the other gene. The effect is to boost the expression of the formerly low-expressed repressor, which thus returns to its higher constitutive expression rate. The external repression mechanism is experimentally accomplished by means of an inducer.

**The relaxation oscillator.** A relaxation oscillator can be obtained by virtue of the competition arising between a strong self-activating gene *A* that activates a repressor *R* and the repression of *A* by the repressor *R*. A genetic realization of this oscillator was proposed and fabricated by [4]. In order to obtain the parameter space that guarantees oscillations, a simple analysis based on a hybrid model was proposed by [4]. The analysis of a two dimensional model, which exploits the Poincaré-Bendixson Theorem, is proposed in [7]. The data obtained by the experiments in [4] show damped almost sinusoidal oscillations. The fact that the oscillations are damped means that the equilibrium point is stable (other behaviors are ruled out by the Poincaré-Bendixson Theorem). In [7], the analysis of a four dimensional model including the m-RNA dynamics is also proposed. The results suggest that the observed experimental oscillations are consistent with a parameter set close to a supercritical Hopf bifurcation [24] corresponding to a stable equilibrium point. This gives a clear suggestion of the required parameter change to obtain sustained oscillations.

**The Repressilator.** Elowitz and Leibler [10] constructed the first operational oscillatory genetic circuit in *E. coli* consisting of three repressors arranged in ring fashion, and coined it the “repressilator”. The repressilator exhibits sinusoidal, limit cycle oscillations in periods of hours, which are slower than the cell-division life cycle. Therefore, the state of the oscillator is transmitted between generations from mother to daughter cells. Motivated by the analysis

of cyclic feedback gene systems, Hastings [14] and Mallet-Paret and Smith [15] developed results which show that if the equilibrium point is unstable, the  $\omega$ -limit set of any bounded trajectory is a periodic orbit. A detailed parametric analysis for the repressilator model, using Hasting's Theorem, can be found in [9].

Thus, one can search for parameter values to guarantee the instability of the equilibrium point. This procedure was followed by [10] in the design of the repressilator. The experimentally obtained oscillations are not dampened, suggesting that the parameter set was properly chosen and thus validating both the model and the analysis performed. However, the oscillations appear to be very noisy compared to that of the relaxation oscillator in [4]. The causes of such a noisy behavior are currently under investigation [19].

#### IV. C M C S T :T M A

In traditional systems theory, a system is usually modeled as an input/output device with internal dynamics. Such an input/output abstraction has been also implicitly employed to analyze the circuits described in the previous section. This input/output description has been very useful in several other fields of engineering, including mechanical and electrical engineering, for composing systems and for deriving properties of an interconnection by the properties of the composing systems. Such an abstraction, however, tacitly assumes that the input/output response and internal dynamics of a system does not change upon interconnection. Such a property of a system is commonly referred to as *modularity*. As it has been noticed by [25], viewing interconnections as input-to-output assignments and viewing signal transmission as unidirectional impose constraints that are not present in the physics of a system. Such constraints may be appropriate in special situations occurring in signal processing and electronics, mainly because such engineering systems have been on purpose designed to obtain unidirectional signal propagation. Natural physical and biological systems are not necessarily describable using such constraints. This is especially true in transcriptional networks, in which proteins are typically signal carriers that connect an upstream system to a downstream system. A protein that travels from an upstream system to a downstream one to, for example, act as a transcription factor of a target gene cannot participate to the network of interactions that characterize the upstream system. As a consequence, the fact that the protein travels to a downstream system to carry the signal will affect the behavior of the upstream system. From a systems and signals viewpoint, we can interpret this phenomenon by saying that a signal generated by the downstream system will travel upstream and affect the dynamics of the upstream system. We refer to this type of signal that travels from downstream to upstream as *retroactivity* [13], [21], [22]. The amount of such a retroactivity will change depending on the features of the interconnection and of the downstream system. For example, if the affinity of the promoter binding sites of the protein is low, one can expect that the fact that the protein

acts as a signal carrier from upstream to downstream will not affect much the upstream system.

To formally model such a retroactivity phenomenon, it has been proposed [8] to define a system  $S$  to have internal state  $x$ , two types of inputs (I), and two types of outputs (O): an input “ $u$ ” (I), an output “ $y$ ” (O), a *retroactivity to the input* “ $r$ ” (O), and a *retroactivity to the output* “ $s$ ”. In such a formalism, achieving low retroactivity effect becomes the control-theoretic problem of *disturbance attenuation* [8]. Other authors have proposed instead to re-define the inputs and outputs of components in a bio-molecular system so as to attain zero retroactivity [21]. However, the feasibility of such an approach is still under discussion among researchers as it may rely on assumptions that not always are met in the biological substrate. Other approaches consider OPAMP [23] like schemes in order to attain low “output impedance” [20], while other researchers propose non-inverting amplifier-based schemes to decrease the retroactivity effect [8]. Such schemes, however, have still to be validated by *in vivo* implementation. Many open questions are left still to answer: To what extent can we perform modular design (as it is performed in electronics) by proper circuit design? What biological mechanisms can we exploit to achieve such a goal? Do natural biological systems employ mechanisms such as insulation to successfully propagate signals? Are there any input/output descriptions that result in lower retroactivity? These seem to be central questions to answer in order to develop a biological engineering discipline.

#### R

- [1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, 2002.
- [2] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall, 2006.
- [3] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss. Synthetic biology: New engineering rules for an emerging discipline. *Molecular Systems Biology*, 2:1–14, 2006.
- [4] M. R. Atkinson, M. A. Savageau, J. T. Meyers, and A. J. Ninfa. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell*, 113:597–607, 2003.
- [5] D. Baker, G. Church, J. Collins, D. Endy, J. Jacobson, J. Keasling, P. Modrich, C. Smolke, and R. Weiss. ENGINEERING LIFE: Building a FAB for biology. *Scientific American*, pages 44–51, June 2006.
- [6] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405:590–593, 2000.
- [7] D. Del Vecchio. Design and analysis of an activator-repressor clock in *E. coli*. In *American Control Conference*, pages 1589–1594, 2007.
- [8] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag. Modules and retroactivity: Theoretical framework, analysis, and design. In *International Conference on Systems Biology*, Long Beach, 2007.
- [9] H. El Samad, D. Del Vecchio, and M. Khammash. Repressilators and promoters: Loop dynamics in synthetic gene networks. In *American Control Conference*, pages 4405–4410, 2005.
- [10] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:339–342, 2000.
- [11] D. Endy. Foundations for engineering biology. *Nature*, 438:449–453, 2005.
- [12] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of the genetic toggle switch in *Escherichia Coli*. *Nature*, 403:339–342, 2000.
- [13] E. D. Gilles. Network theory for chemical processes. *Chemical Engineering & Technology*, 21:121–132, 1999.
- [14] S. Hastings, J. Tyson, and D. Webster. Existence of periodic solutions for negative feedback cellular systems. *J. of Differential Equations*, 25:39–64, 1977.

- [15] J. Mallet-Paret and H.L. Smith. The Poincaré-Bendixson theorem for monotone cyclic feedback systems. *J. of Dynamics and Differential Equations.*, 2:367–421, 1990.
- [16] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002.
- [17] J. Monod and F. Jacob. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. In *Cold Spring Harb. Symp. Quant. Biol.* 26, pages 389–401, 1961.
- [18] N. Rosenfeld, M. B. Elowitz, and U. Alon. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.*, 323:785–793, 2002.
- [19] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz. Gene regulation at the single-cell level. *Science*, 307:1962–1965, 2005.
- [20] G. De Rubertis and S. W. Davies. A genetic circuit amplifier: Design and simulation. *IEEE Trans. on Nanobioscience*, 2:239–246, 2003.
- [21] J. Saez-Rodriguez, A. Kremling, H. Conzelmann, K. Bettenbrock, and E. D. Gilles. Modular analysis of signal transduction networks. *IEEE Control Systems Magazine*, 24:35–52, 2004.
- [22] J. Saez-Rodriguez, A. Kremling, and E. D. Gilles. Dissecting the puzzle of life: modularization of signal transduction networks. *Computers and Chemical Engineering*, 29:619629, 2005.
- [23] Donald L. Schilling and Charles Belove. *Electronic Circuits: Discrete and Integrated*. McGraw Hill, 1968.
- [24] S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer-Verlag, 1990.
- [25] J.C. Willems. Behaviors, latent variables, and interconnections. *Systems, Control and Information*, 43:453–464, 1999.