

# Synthetic Biology

Domitilla Del Vecchio\* and Richard M. Murray\*\*

\*Department of Mechanical Engineering, MIT  
Cambridge, 02139

\*\*Control and Dynamical Systems, Caltech  
Pasadena, 91125

ddv@mit.edu, murray@cds.caltech.edu

**Abstract**—The past decade has seen tremendous advances in DNA recombination and measurement techniques. These advances have reached a point in which *de novo* creation of biomolecular circuits that accomplish new functions is now possible, leading to the birth of a new field called synthetic biology. Sophisticated functions that are highly sought in synthetic biology range from recognizing and killing cancer cells, to neutralizing radioactive waste, to efficiently transforming feedstock into fuel, to control the differentiation of tissue cells. To reach these objectives, however, there are a number of open problems that the field has to overcome. Many of these problems require a system-level understanding of the dynamical and robustness properties of interacting systems and hence the field of control and dynamical systems theory may highly contribute. In this paper, we review the basic technology employed in synthetic biology, a number of simple modules and complex systems created using this technology, and discuss key system-level problems along with challenging research questions for the field of control theory.

**Index Terms**—biomolecular systems, gene expression, robustness, modularity.

## I. INTRODUCTION TO SYNTHETIC BIOLOGY

Synthetic biology is an emerging engineering discipline in which the biochemical and biophysical principles present in living organisms are used to engineer new systems [5]. These systems will have the ability of accomplishing a number of remarkable tasks, such as turning waste into energy sources, neutralizing radioactive waste, detecting environmental pathogens, or recognizing cancer cells with the aim of targeting them for deletion. While synthetic biology can be employed to create new functionalities, it can also enable the understanding of fundamental design principles of living systems. In fact, implementing a circuit with a prescribed behavior provides a powerful means to test hypotheses regarding the underlying biological mechanisms.

The functions of living organisms are controlled by biomolecular circuits, in which proteins and genes interact with each other through activation and repression interactions forming complex networks. A common signal carrier is the concentration of the active form of a protein, which can be controlled through a number of mechanisms, including gene expression regulation and post translational modification. Through the process of gene expression, proteins are produced by their corresponding genes, whose production rates can be activated or repressed by other proteins (transcription factors).

Once the proteins are produced they can be activated or inhibited, by other proteins or smaller molecules, through post-translation modification processes including covalent modification, such as phosphorylation, and allosteric modification [2]. We next describe some salient aspects of gene expression focusing, for simplicity, on prokaryotic systems.

A gene is a piece of DNA whose expression rate can often be controlled by a DNA sequence upstream of the gene itself, called promoter. The promoter contains the binding regions for the RNA polymerase, an enzyme that transcribes the gene into a messenger RNA molecule, which is then translated into protein by the ribosomes. The promoter also contains operator sites, which are binding regions where other proteins, called transcription factors, can bind. If these proteins are activators, they will help the RNA polymerase in binding the promoter to start transcription. By contrast, if these proteins are repressors, they will prevent the RNA polymerase from binding the promoter. These activation and repression interactions are highly nonlinear and often stochastic, therefore the most commonly used modeling frameworks include systems of nonlinear ordinary differential equations, stochastic differential equations, or the chemical master equation [16, 17].

The basic technique for constructing synthetic circuits is that of assembling, through the process of cloning, DNA sequences with prescribed combinations of promoters and genes such that a desired network of activation and repression interaction is created. For example, if we would like to create an inverter where protein A represses protein B, we can simply place the gene of B under the control of a promoter repressed by protein A. Currently, there is a library of parts that one can use to assemble a desired circuit this way. The set of parts includes promoters, gene coding sequences, terminators, and ribosome binding sites. Terminators are DNA sequences placed at the end of a gene to make the RNA polymerase terminate transcription, while ribosome binding sites are DNA sequences placed at the beginning of a gene, which establish the rate at which ribosomes will bind to the mRNA, determining the overall translation rate [13]. An area of intense research is expansion of the library by creating mutations of existing parts or by assembling new ones.

Once a DNA sequence is created that encodes the desired circuit, it is inserted in a living cell either on the chromosome itself or on DNA plasmids. When the circuit is inserted in the chromosome it will be in one copy, while when it is

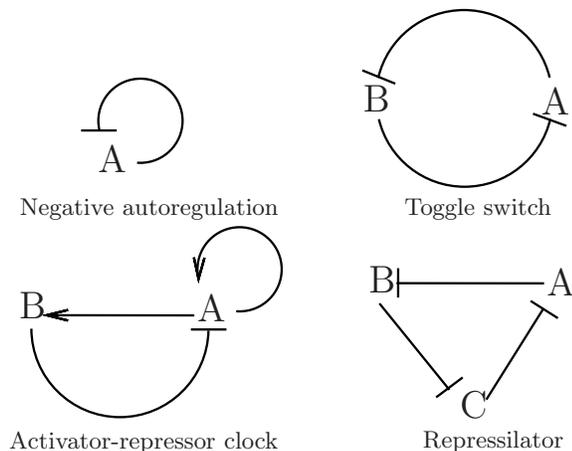


Fig. 1. Early gene circuits that have been fabricated in bacteria *E. coli*: the negatively autoregulated gene [6], the toggle switch [15], the activator-repressor clock [3], and the repressilator [12].

inserted in DNA plasmids, it will be in as many copies as the plasmid copy number. Plasmid copy number can vary from low copy (5-10 copies), to medium copy (20 copies), to high copy (about 100 copies). Once in the cell, the circuit will have the required resources to function, including RNA polymerase, ribosomes, amino acids, and ATP (the cell energy currency). In this sense, the cell can be viewed as a chassis for the synthetic circuits. The operation of the circuit can then be observed by monitoring the concentration of reporters, that is, of proteins that are easy to detect and quantify. These include fluorescent proteins, that is, proteins that exhibit bright fluorescence when exposed to light of a specific wave length. Examples include the green, red, blue, and yellow fluorescent proteins. These fluorescent proteins are mainly employed in two different ways to measure the amount of a protein of interest. One can fuse the gene of the fluorescent protein with the gene expressing the protein of interest. Alternatively, one can use the protein of interest as a transcription factor of the fluorescent protein. In both cases, the concentration of the fluorescent protein will provide an indirect measurement of the concentration of the protein of interest.

It is also possible to apply external inputs to a circuit to control the activity of transcription factors. This is accomplished through the use of inducers, which are small signaling molecules that can be injected in the cell culture and enter the cell wall. These inducers bind specific transcription factors and either activate them, allowing the transcription factor to bind the promoter operator sites, or inhibit them, reducing the transcription factor ability to bind the promoter operator sites.

## II. EXAMPLES OF SYNTHETIC BIOLOGY MODULES

A number of modules comprising two or three genes have been fabricated in the earlier days of synthetic biology [3, 6, 12, 15, 26]. We can group them into oscillators [3, 12, 26], mono-stable systems [6], and bistable systems called toggle

switches [15]. More recently, feedforward loops have also been fabricated [8].

**Oscillators.** The creation of circuits whose protein concentrations oscillate periodically in time has been a major focus. In fact, the ability of creating an oscillator has the potential of shedding light into the mechanisms at the basis of natural clocks, such as circadian rhythms and the cell cycle. Oscillator designs can be divided into two types: loop oscillators [12], in which repression/activation interactions occur in a loop topology, or oscillators based on the interplay between an autocatalytic loop and negative feedback [3, 26] (see Figure 1).

The design requirements of synthetic circuits are usually explored through models of varying detail, starting with the use of low-dimensional “toy models”, which are composed of a set of nonlinear ordinary differential equations describing the rate of change of the circuit’s proteins. These models allow application of a number of tools from dynamical systems theory to infer parameter or structural requirements for a desired behavior. After toy models are analyzed, larger scale mechanistic models are constructed, which include all the intermediate species taking part in the biochemical reactions. These models can be either deterministic or stochastic. Simulation is usually required for the study of these more complicated models and the Gillespie algorithm is often employed for stochastic simulations [16].

As an example of a toy model and related analysis, consider the activator-repressor clock of Atkinson et al. [3] shown in Figure 1. This oscillator is composed of an activator *A* activating itself and a repressor *B*, which, in turn, represses the activator *A*. Both activation and repression occur through transcription regulation. Denoting in italics the concentration of species, a toy model of this clock can be written as

$$\begin{aligned} \dot{A} &= \frac{\beta_A(A/K_a)^n + \beta_{0,A}}{1 + (A/K_a)^n + (B/K_b)^m} - \gamma_A A, \\ \dot{B} &= \frac{\beta_B(A/K_a)^n + \beta_{0,B}}{1 + (A/K_a)^n} - \gamma_B B, \end{aligned} \quad (1)$$

in which  $\gamma_A$  and  $\gamma_B$  represent protein decay (due to dilution and/or degradation). The functions  $(\beta_A(A/K_a)^n + \beta_{0,A})/(1 + (A/K_a)^n + (B/K_b)^m)$  and  $(\beta_B(A/K_a)^n + \beta_{0,B})/(1 + (A/K_a)^n)$  are called Hill functions and are the most commonly used models for transcription regulation [2]. The first Hill function in system (1) increases with *A* and decreases with *B* while the second one increases with *A*, as expected since *A* is an activator and *B* is a repressor. The key mechanism by which this system displays sustained oscillations is a supercritical Hopf bifurcation with bifurcation parameter the relative time scale of the activator dynamics with respect to the repressor dynamics [10]. Specifically, as the activator dynamics become faster than the repressor dynamics, the system goes through a supercritical Hopf bifurcation and a stable periodic orbit appears (Figure 2(b)).

**Mono-stable systems.** The mono-stable system engineered

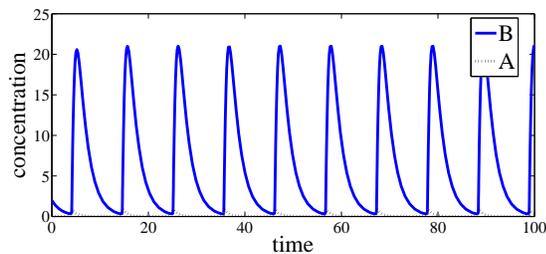


Fig. 2. Activator-repressor clock time trajectory.

through negative autoregulation was fabricated with the aim of understanding the role of negative feedback in attenuating biological noise. The results of Becskei and Serrano [6] clearly showed that negative autoregulation can reduce intrinsic noise. Furthermore, the results of Austin et al. [4] demonstrated that while low frequency noise is attenuated, noise at high frequency can be amplified by negative autoregulation in accordance with Bode’s integral formula [1].

**Bistable systems.** The toggle switch of Gardner et al. [15] was the first bistable system constructed. It constitutes the simplest circuit with memory, in which the state of the system can be switched from one equilibrium (low, high) to the other (high, low) by external inputs. Once the system state is switched to one of these two equilibria, it will stay there unless another external perturbation is applied.

**Feedforward loops.** While the early circuits described so far were fabricated mainly to investigate design principles for limit cycles and for robustness, many more circuits after those have been fabricated with the aim of solving concrete engineering problems. As an example, the incoherent feedforward circuit of Bleris et al. [8] was fabricated in bacteria *E. coli* with the aim of making protein production independent of DNA plasmid copy number. In fact, DNA copy number fluctuates stochastically with possibly large deviations from the nominal value. As a consequence, the concentration of proteins expressed from genes residing on a plasmid also fluctuates stochastically. In order to make protein concentration independent of an unknown DNA copy number, one could leverage principles for disturbance rejection such as integral control. While an explicit integral control action is particularly hard to implement through biological parts, incoherent feedforward loops are easier to implement and can accomplish the same disturbance rejection task. In these loops, the disturbance input affects the output through two branches, one in which the disturbance activates the output and a longer one in which the disturbance represses the output [2]. If these two branches are appropriately balanced, the steady state value of the output will be practically independent of the disturbance input, leading to disturbance rejection to constant or slowly changing disturbances.

### III. FROM MODULES TO SYSTEMS

One approach to creating systems that can accomplish sophisticated tasks is to assemble together simpler modules, such as those described in the previous section [23]. For example, the artificial tissue homeostasis circuit proposed by Miller et al. [19] is composed of several interconnected modules, including an activator-repressor clock, a toggle switch, a couple of inverters, and an “and” gate. Control of tissue homeostasis refers to the ability of regulating a cell type to a constant level in a multi-cellular community. This ability is central in several diseases such as cancer and diabetes, in which tissue homeostasis is misregulated. The design proposed by Miller et al. [19] illustrates how a synthetic biological circuit can be modularly created to accomplish this complicated regulation function.

Layered logic gates are often necessary in order to integrate multiple signals. Moon et al. [21] have constructed an “and” gate that integrates more than two signals by cascading pairs of “and” gates. Of course, problems of latency become more relevant as the number of layers increases and methods to mitigate these effects are being developed.

An application that requires the integration of multiple signals is the cell type classifier of Xie et al. [28]. Here, a synthetic gene circuit is created that integrates sensory information from a number of molecular markers to determine whether a cell is in a specific state, that is, cancer, and, in such a case, produces a protein output triggering cell death. The design of this circuit is based on the composition of three key modules. Specifically, a double inversion module senses high levels of a molecular marker, a single inversion module senses low levels of a molecular marker, and a logical “and” module finally integrates the outputs of the other two modules to produce the output protein.

Finally, biofuels is another high-impact application of synthetic biology [22]. Metabolic engineering has been employed for long time in order to engineer microbes to produce advanced biofuels with similar properties to petroleum-based fuels. One challenge in using microbes (or other living organisms) to convert feedstock into biofuel is that of overcoming the endogenous cell regulation to achieve sufficiently high yields such that advanced biofuels are economically advantageous. Specifically, engineered pathways are optimized on the basis of nominal operating conditions, but these conditions often change when microbes are in bioreactors. To mitigate this problem, synthetic gene circuits have been designed to sense the metabolic status of the host and regulate key points in the metabolic pathway to optimize yield [29].

### IV. MAIN SYSTEM-LEVEL CHALLENGES TO DESIGN

One major challenge in synthetic biology is the ability of going from simple modules to larger sophisticated systems [23]. Problems in advancing in this direction can be divided into two categories: “hardware” problems and system-level

problems. Hardware problems include issues such as the availability of enough orthogonal parts to allow scaling up the size of synthetic circuits. We do not expand on this here, and instead focus on system-level problems. These include issues such as *context-dependence* [9], that is, the fact that modules behave in a poorly predictable way once interacting together in the cell environment. This is a major obstacle to creating larger circuits that behave predictably.

Problems of context dependence can be further divided into three qualitatively different types: (a) inter-modular interactions, (b) interactions of synthetic circuits with the cell machinery, (c) perturbations in the external environment. We analyze each of them separately.

- (a) When modules are connected to each other to create larger systems, a protein in an upstream module is used as an “input” to a downstream module. This fact creates a “loading” on the upstream system due to the fact that the output protein cannot take part in the upstream module reactions whenever it is taking part in the downstream module reactions. As a consequence, the behavior of the upstream system changes compared to when the system functions in isolation [11, 25]. These loading effects have been called retroactivity to extend the notion of loading and impedance to biomolecular systems. Accordingly, solutions to mitigate this problem are being investigated [14, 18, 20].
- (b) Ideally, the cell should function as a “chassis” for synthetic biology circuits. In practice, this is not the case because the endogenous circuitry interacts with synthetic circuits even when parts that are orthogonal to the endogenous systems are employed. A major example of this interaction is the depletion of cellular resources, such as ATP, RNA polymerase and ribosomes, which are required for the operation of synthetic circuits. This depletion reduces cell fitness, with deleterious consequences also for synthetic circuits, a phenomenon called “metabolic burden” [7]. A more subtle phenomenon than purely reducing cell fitness is that synthetic circuits compete with each other for the same resources. This fact creates implicit and unwanted coupling among circuits with unpredictable consequences. Approaches to mitigate these problems are under investigation. One direction is the use of orthogonal RNA polymerase and ribosomes [27, 24]. A completely different, but complementary, direction is that of establishing implementable design principles that allow circuits to function robustly despite fluctuations in the resources they use.
- (c) The external environment where a cell operates has a number of physical attributes, which may also be subject to perturbations. These physical attributes include temperature, acidity, nutrients’ level, etc. Perturbations in these attributes often lead to poor cell fitness or to non-standard growth conditions, ultimately leading to synthetic circuits malfunctions.

## V. SUMMARY AND FUTURE DIRECTIONS

The future of synthetic biology highly depends on the ability of scaling up the complexity of design to create more sophisticated functions. While a number of issues, such as the availability of enough orthogonal parts, can be successfully addressed by (non-trivial) fabrication of new parts, issues such as context-dependence require a system-level dynamic understanding of circuits and their interactions. Here is where control and dynamical systems theory could greatly contribute. Control theory has proven critical to reason about and engineer robustness in a number of concrete applications including aerospace and automotive systems, robotics and intelligent machines, manufacturing chains, electrical, power, and information networks. Similarly, control theory could enable the understanding of principles that ensure robust behavior of synthetic circuits once interacting with each other in the cell environment, leading to the ultimate progress of synthetic biology.

A number of challenges need to be addressed for the successful application of control and dynamical systems theory to synthetic biology. The behavior of synthetic circuits is highly nonlinear and, as a consequence, control theoretic tools designed for understanding robustness in linear systems are not directly applicable. Understanding how to exploit the rich structure of biomolecular circuits to quantitatively reason about robustness to interconnections, competition for shared resources, and fluctuations of temperature and nutrients is likely to have a major impact. Even with this understanding, however, the question of how to implement robust designs with the currently available biomolecular mechanisms must be addressed. Stochasticity is another major problem since the behavior of synthetic circuits is intrinsically noisy. Unfortunately, the availability of analytical tools that allow quantification of how perturbations and uncertainty propagate through a nonlinear stochastic system is still limited and designers often resort to stochastic simulation. Finally, the values of the salient parameters of the available parts are poorly known. Physical attributes such as binding affinities, ribosome binding site strengths, promoter strengths, etc. are only known within very coarse bounds. These bounds are also usually determined based on a specific organism and in specific growth conditions, which may be different from the ones in which the circuit is ultimately running. Hence, a central question is how to design and implement a system such that the prescribed behavior is robust to all sources of perturbations described above within a large range of possible parameter values.

## REFERENCES

- [1] K. J. Åström and R. M. Murray. *Feedback Systems*. Princeton, 2008.
- [2] U. Alon. *An introduction to systems biology. Design principles of biological circuits*. Chapman-Hall, 2007.
- [3] 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.

- [4] D. W. Austin, M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, G. S. Saylor, N. F. Samatova, C. D. Cox, and M. L. Simpson. Gene network shaping of inherent noise spectra. *Nature*, 439:608–611, 2005.
- [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] W. E. Bentley, N. Mirjalili, D. C. Andersen, R. H. Davis, and D. S. Kompala. Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol Bioeng*, 35(7):668–81, 1990.
- [8] L. Bleris, Z. Xie, D. Glass, A. Adadey, E. Sontag, and Y. Benenson. Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Molecular Systems Biology*, 7:519, 2011.
- [9] S. Cardinale and A. P. Arkin. Contextualizing context for synthetic biology identifying causes of failure of synthetic biological systems. *Biotechnology Journal*, 7:856866, 2012.
- [10] D. Del Vecchio. Design and analysis of an activator-repressor clock in *E. coli*. In *Proc. American Control Conference*, pages 1589–1594, 2007.
- [11] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag. Modular cell biology: Retroactivity and insulation. *Molecular Systems Biology*, 4:161, 2008.
- [12] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:339–342, 2000.
- [13] D. Endy. Foundations for engineering biology. *Nature*, 438(24):449–452, 2005.
- [14] E. Franco, E. Friedrichs, J. Kim, R. Jungmann, R. Murray, E. Winfree, and F. C. Simmel. Timing molecular motion and production with a synthetic transcriptional clock. *Proc. Natl. Acad. Sci*, doi: 10.1073/pnas.1100060108, 2011.
- [15] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of the genetic toggle switch in *Escherichia Coli*. *Nature*, 403:339–342, 2000.
- [16] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81:2340–2361, 1977.
- [17] D. T. Gillespie. The chemical langevin equation. *The Journal of Chemical Physics*, 113:297–306, 2000.
- [18] S. Jayanthi and D. Del Vecchio. Retroactivity attenuation in biomolecular systems based on timescale separation. *IEEE Trans. Aut. Control*, 56:748–761, 2011.
- [19] M. Miller, M. Hafner, E. Sontag, N. Davidsohn, S. Subramanian, P. Purnick, D. Lauffenburger, and R. Weiss. Modular design of artificial tissue homeostasis: robust control through synthetic cellular heterogeneity. *PLoS computational biology*, 8:e1002579, 2012.
- [20] D. Mishra, P. Rivera-Ortiz, D. Del Vecchio, and R. Weiss. A load driver device for engineering modularity in biological networks. *Under Review*, 2013.
- [21] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C.A. Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491:249–253, 2012.
- [22] P. P. Peralta-Yahya, F. Zhang, S. B. del Cardayre, and J. D. Keasling. Microbial engineering for the production of advanced biofuels. *Nature*, 488:320–328, 2012.
- [23] P. Purnick and R. Weiss. The second wave of synthetic biology: from modules to systems. *Nature reviews. Molecular cell biology*, 10:410–22, 2009.
- [24] O. Rackham and J. W. Chin. A network of orthogonal ribosome-mrna pairs. *Nature Chemical Biology*, 1(3):159–166, 2005.
- [25] J. Saez-Rodriguez, A. Kremling, H. Conzelmann, K. Bettenbrock, and E. D. Gilles. Modular analysis of signal transduction networks. *IEEE Control Systems Magazine*, pages 35–52, 2004.
- [26] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456:516–519, 2008.
- [27] A. Wenlin and J. W. Chin. Synthesis of orthogonal transcription-translation networks. *Proc. Natl. Acad. Sci.*, 106(21):84778482, 2009.
- [28] Z. Xie, L. Wroblewska, L. Prochazka, R. Weiss, and K. Benenson. Multi-input rnai-based logic circuit for identification of specific cancer cells. *Science*, 333:1307–11, 2011.
- [29] F. Zhang, J. M. Carothers, and J. D. Keasling. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.*, 30:354–359, 2012.