10

Synthetic Networks

Jongmin Kim

CbsBioscience Inc., Daejeon, Korea

10.1 Introduction

In order to understand and utilize the diverse functionality displayed by biological organisms, it is first necessary to comprehend the regulatory network underlying such complex behavior. Fortunately, biological networks share certain properties of engineered networks [1], and thus are potentially amenable to automated design and characterization. Recent advances in both metabolic and genetic engineering have made feasible the investigation of novel biological functionality through the design and implementation of synthetic biological networks. Well-characterized 'parts' would be essential for streamlining synthetic network design processes, such that complex functionality can be created without reinventing all details of the molecules involved. One example of such an effort towards standardized parts for abstraction is the Registry of Standard Biological Parts (http://parts.mit.edu). Another important research venue is mathematical modeling including quantitative analysis, which allows for the circuit behavior to be explored with uncertain parameter sets and external disturbances. Today, several software tools are available to aid biochemical kinetic simulations [2]. In this chapter, the current understanding of cellular networks, synthetic network construction and the remaining challenges towards automating biochemical processes using synthetic circuitry are reviewed.

10.2 Cellular Network: Functional Design

Cells live in a complex environment and can sense many different signals, whether physical, chemical or biological. Cells also have the ability to process information for survival

Automation in Proteomics and Genomics: An Engineering Case-Based Approach Edited by Gil Alterovitz, Roseann Benson and Marco Ramoni

and reproduction, such as detecting nutrients and avoiding harmful chemicals, by using functional circuits composed of many interacting molecular species [1]. Hence, information processing through regulatory networks lies at the heart of all living systems. By taking a 'top-down' view of protein–protein interactions, signaling pathways and gene regulatory pathways, the basic architecture of biological networks has been analyzed [3]. The network description of cellular circuits allows the application of tools and concepts which have been developed in fields such as graph theory, physics, sociology and engineering [4]. Remarkably, biological networks share the design principles of engineered networks, namely modularity, robustness and recurring circuit elements. A module in a network is a set of nodes that have strong interactions and a common function [1]. Modules in engineering – and presumably also in biology – have special features that make them easily embedded in almost any system. The robustness of a cellular network design requires that the design must function under plausible fluctuations and interferences due to the components and to the environment [5]. Recurring network motifs for signal processing tasks, such as filtering out input noise, accelerating throughput of the network or temporal programming, can be found in biological networks [6]. The fact that a biological organism must function and compete for resources imposes severe constraints on the regulatory network design, which could have shaped the biological networks with characteristics analogous to human-engineered networks. These design principles of cellular networks will help delineate system architecture with limited data, such that researchers can focus on modular and robust patterns. Indeed, some of these patterns are already known as network motifs.

10.2.1 Network Motifs

Alon and colleagues studied the transcription network of *Escherichia coli* to identify meaningful patterns on the basis of statistical significance. The transcription network was compared to an ensemble of randomized networks, with similar characteristics such as the same number of nodes and edges but with random connections between nodes and edges. Patterns that occur in the real network significantly more often than in randomized networks were termed network motifs [6, 7]. One network motif is that of negative autoregulation, where a protein product binds to its own promoter and represses its own transcription. Negative autoregulation has two useful features – the speed-up of response time and robustness to fluctuation. The response time, which is defined as the time to reach halfway between the initial and final levels in a dynamic process, depends simply on degradation and dilution rates in unregulated transcription and translation processes. In order to achieve the same steady-state value, negative autoregulation employs a stronger promoter than its unregulated counterpart; therefore, the initial build-up of signals is fast with negative autoregulation, cutting down the response time. Moreover, the steady-state protein level is stable with negative autoregulation, albeit with fluctuations in the production rate. An important three-node motif – termed the feedforward loop – is defined by a transcription factor X that regulates a second transcription factor Y, such that both X and Y jointly regulate an operon Z (Figure 10.1a). Most of the feedforward loops are coherent; that is, the direct regulation of X on Z and indirect regulation of X on Z through Y are of the same sign. Mathematical analysis suggests that the coherent feedforward loop can act as a persistence detector, rejecting short pulses of activation signals from the general transcription factor responses. Consider the case where both X and Y transcription factors are required for the

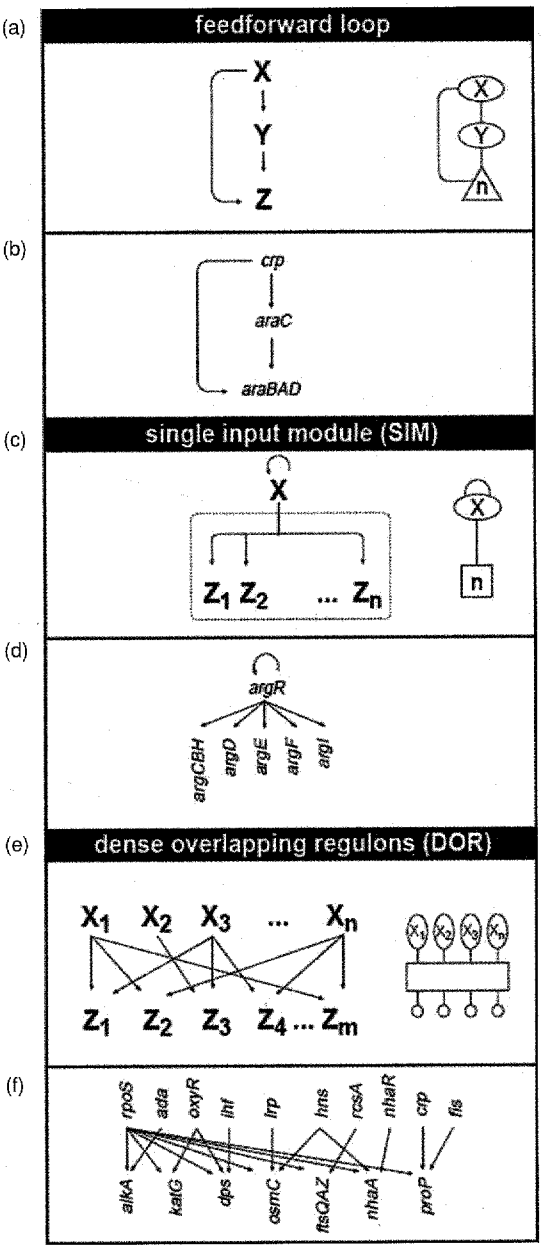

Figure 10.1 *Network motifs found in the* E. coli *transcriptional network. (a) Feedforward loop: a transcription factor X regulates a second transcription factor Y, and both jointly regulate one or more operons $Z_1 \ldots Z_n$; (b)* L*-arabinose utilization network; (c) SIM motif: a single transcription factor, X, regulates a set of operons $Z_1 \ldots Z_n$; (d) Arginine biosynthesis network; (e) DOR motif: a set of operons $Z_1 \ldots Z_m$ are regulated by a combination of a set of input transcription factors, $X_1 \ldots X_n$; (f) Stationary phase response network. (Reprinted by permission from Macmillan Publishers Ltd. Ref. [6])*

activation of Z in a feedforward loop. Upon arrival of activation signal for X, the activation of Z is delayed because Y takes time to build up to a threshold level. Thus, if the activation signal for X has a short duration, Y cannot reach the threshold level needed to activate Z. Response to signals such as nutrients that activate X incurs production cost for the final enzyme Z, but no significant benefit can be gleaned if the nutrients disappear by the time enzyme Z level is sufficiently high. A cost–benefit analysis indicates that a coherent feedforward loop offers more benefit over the simple regulation of X and Y on Z under a fluctuating environment where transient activation signal is common [8].

Two other larger motifs are called the single-input module (SIM) and the dense overlapping regulon (DOR). The SIM network motif – a simple pattern in which one regulator controls a group of genes – can generate temporal programs of expression, in which genes are turned on one by one in a defined order (Figure 10.1c). In contrast, the DOR network motif is a layer of overlapping interactions between operons and a group of input transcription factors, in which the signal inputs are integrated and the output genes are under a combinatorial control (Figure 10.1e). Other network motifs appear in a developmental transcription network [9], such as a positive feedback loop and a long cascade. A positive feedback loop can serve as a memory, locking in the cell fate if an early developmental signal ever reaches a threshold level. Long cascades are uncommon in sensory information processing due to significant delays, but prove useful in developmental timing that spans several cell generations. Transcription regulatory networks operate on the timescale of tens of minutes to hours, whereas signal transduction networks rely on protein–protein interactions to process sensory signals on the timescale of seconds to minutes. A more complete picture of cellular networks requires an analysis of the interaction of different network components operating at different timescales and searching for novel regulatory mechanisms operating on such interfaces.

Network motifs provide a powerful tool to understand cellular organization from a functional point of view, bypassing the biochemical details. The spontaneous evolution of modularity and network motifs has been demonstrated in computational evolution models of electronic circuits and neural networks [10]. Many such models use networks in a population explored by means of mutations, crossover and duplication to be selected for a defined goal. The evolved systems typically result in intricately wired nonmodular solutions because these are more optimized than their human-engineered counterparts. A lack of modularity has been cited as one of the reasons why computational evolution can generate design patterns for simple tasks, but cannot be scaled-up to more complex tasks. If the network evolution is constrained to fulfill modularly varying goals, then the achieved architecture is built of more computational units solving subproblems; this framework has an increased modularity but is suboptimal. Modularity decreases quickly when the network is trained on a single goal or nonmodularly varying goals. Kashtan and Alon [10] have suggested that modularity allows a higher adaptability to be achieved, and is therefore a characteristic that a biological network must have in order to evolve in a constantly changing environment that requires a certain set of basic functions in different combinations.

10.2.2 Network Architecture

An alternative approach is to abstract features from the overall architecture of cellular networks. The architecture of a network places boundaries on its performance capabilities,

and also explains its possible evolutionary path [3]. Clearly, cellular networks differ from regular networks, where nearest neighbors are linked in a regular fashion, or from random networks, where randomly selected nodes are joined together. In cellular networks, a few nodes have a large number of connections, while most of the nodes have relatively few connections – this is a feature of a 'scale-free' network. 'Scale-free' means that the number of molecules (N) with a given number of connections (k) falls off as a power law, $N(k) \sim k^{-g}$, where no characteristic peak value can be found. In a scale-free network, the average distance between any two nodes is almost as small as the random network, while the extent to which neighbors of a node are themselves connected (known as its clustering coefficient) is almost as large as in a regular network. Protein–protein interaction maps have the features of a scale-free network, with their degree sequences (number of edges per node) often following a long-tailed distribution [11]. However, the fact that a network has scale-free properties is of limited use, since power laws occur widely in nature, possibly with different mechanistic origins. Thus, a much closer examination of small-scale networks, such as subnetworks or molecular complexes, should complement the top-down network description [3].

It has been suggested that biological networks have additional constraints that are beyond simple scale-free networks [12]. Networks that are simple connection networks, such as the Internet, are able to grow in an unconstrained way, whereas regulatory networks – such as genetic regulatory networks in biology – must be able to operate in a globally responsive way. In order to maintain global connectivity, the number of connections must be scaled quadratically with the network size. As a consequence, the need for an increased number of connections at the regulatory level naturally imposes a limit on the size of the network and its complexity [12]. Although dedicated hierarchies could solve such a scalability problem, each level of regulatory hierarchy will introduce time delays and increase stochastic noise [13]. Regulatory proteins scale almost quadratically with genome size in prokaryotes [14], and the extrapolation of this relationship suggests that prokaryotes have reached their complexity limit by their reliance on a protein-based regulatory architecture. Eukaryotes have a far more developed RNA processing and signaling system than prokaryotes, which appears to be linked to a more sophisticated pathway of gene regulation. Recently it was suggested that, in addition to being a digital storage medium, noncoding RNA themselves are actually transmitting digital signals [15]. In contrast, regulatory proteins act mainly as analogue components because their signals are transmitted as their concentrations. Following the comparison with electronic circuits, it is possible that the cellular network complexity limit was lifted by the use of both digital and analogue signals.

In summary, biological networks present different features at different scales, behaving like scale-free networks on a large scale, and consisting of recurring network motifs and basic functionalities on a smaller scale. Network motifs found in transcriptional networks illustrate that the network design has functional consequences. Other modalities of regulatory strategies such as RNA processing and post-translational modifications, although sophisticated regulatory examples are known, have not been discussed here. Investigating the cellular networks at different levels of complexity starting from basic network motifs merits future research efforts that would lead to an understanding of the complexity of regulation strategies and provide useful insights.

10.3 Synthetic Approaches to Understand Cellular Networks

A network description in an abstract sense is not enough to understand cellular networks with quantitative details and to construct predictive models. Rather, the investigation of detailed kinetics and reaction mechanisms among the constituent macromolecules is required. The reductionist approach attempts to explain the behavior of cellular networks in terms of the behavior of the components. Despite many molecular components of biological organisms being identified and characterized using genetic and biochemical techniques, it is still not possible to predict system behavior, except in the simplest systems. This indicates that the great complexity of cellular network hinders the prediction of system behavior from characterized components, and that alternative approaches for understanding cellular network behavior and design principles may be necessary.

Synthetic biology provides an alternative to the study of cellular networks, by constructing increasingly complex analogues of natural circuits. This is a 'bottom-up' approach that attempts to test the sufficiency of mechanistic models by actively synthesizing them: this allows insights to be gained that observation and analysis alone do not provide [16]. A synthetic biology approach shares the spirit of engineering community in that a successful model system should operate upon synthesis. For engineering purposes, parts are most suitable when they contribute independently to the whole. This 'independence property' allows one to predict the behavior of an assembly by characterizing parts. In terms of satisfying independence property, the DNA molecules described by the Watson–Crick model stand out because each nucleotide pair contributes independently to the stability of a duplex, to a good approximation [17]. However, the DNA molecule is rather an exception than the rule; for instance, the behavior of a protein is generally not a function of the behavior of its constituent amino acids.

Although amino acids may be a poor unit for the application of independence property, natural folded proteins can be treated as interchangeable parts. Several synthetic networks constructed by rearranging the regulatory components in a cell have been characterized, including autoregulators [18,19], feedforward cascades [13,20], bistable memory element [21] and oscillators [22,23]. In order for this type of network design to lead to an improved understanding of naturally occurring networks, detailed studies of the synthetic systems are needed [16], for example, through a systematic examination of the effects of parameter variations with quantitative modeling and analysis [24]. Some example networks and their design principles will be discussed.

10.3.1 Synthetic Networks *In Vivo*

A bistable memory was constructed by Gardner *et al.* [21] by employing a mutual repression system which used two genes that each coded for a transcriptional repressor of the other gene. These authors used combinations of the lac repressor (LacI), tetracycline repressor (TetR) and the temperature-sensitive lambda repressor (cI). An external stimulus inhibits the activity of a specific repressor and pushes the system to one steady state. For the mutual repression system shown in Figure 10.2a, isopropyl-β-D-thiogalactopyranoside (IPTG) inhibits the lac repressor, while a high temperature inhibits the cI repressor. Thus, the addition of IPTG pushed the system to a lac-off/lambda-on state and a concomitant increase in the green fluorescent protein (GFP) signal. This system demonstrated hysteresis,

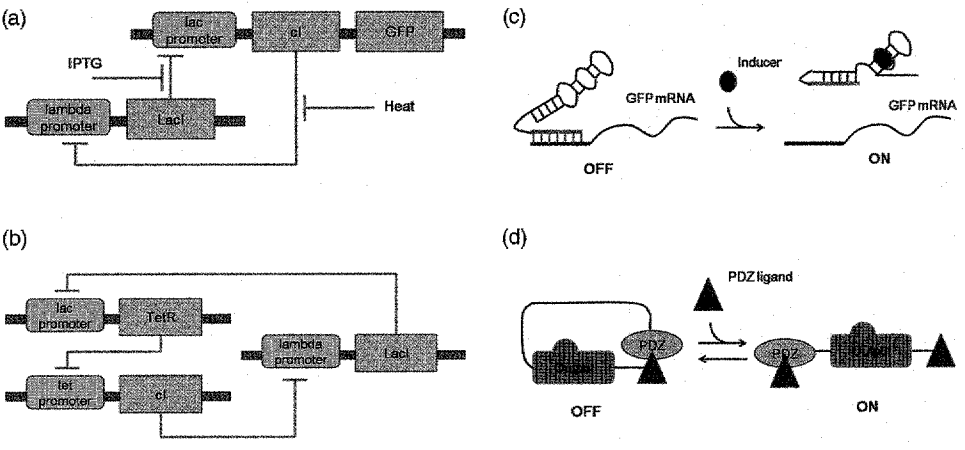

Figure 10.2 *Engineered* in vivo *networks. (a) A genetic toggle switch uses a mutual repression motif. Two genes, lac repressor and lambda repressor, repress the expression of the other gene. Transient exposure to either heat or IPTG will shift the steady-state of the system to the expression of only one repressor; (b) A circular arrangement of repressors comprises the genetic ring oscillator. Oscillatory output was observed via GFP expression regulated by the tetracycline repressor; (c) The RNA 'anti-switch' relies on ligand-binding regions of RNA that, when bound to ligand, induce changes in RNA structure. When bound to an inducer ligand, the anti-switch hides an antisense region of RNA that hybridizes to the 5'-UTR of target mRNA, encompassing the translational start site; (d) An allosteric switch based on the natural N-WASP allosteric switch. A PDZ-binding domain is used with a C-terminal PDZ ligand, resulting in autoinhibition of N-WASP output domain. When an exogenous PDZ ligand is added, the intramolecular PDZ interaction is disrupted, and the output domain stimulates actin polymerization. (Reprinted by permission from Macmillan Publishers Ltd. Refs. [21, 22, 26, 27])*

such that once the switch was flipped toward one steady state it remained there, even in the absence of the original stimulus. Several plasmid constructs with different promoters and ribosome-binding sequences were shown to be bistable, except for one construct. Thus, bistability can be achieved for a wide range of parameter space, if two repressor strengths are balanced. Furthermore, a toggle switch design can be embedded in a larger system. Kobayashi *et al.* [25] used a lac repressor/lambda repressor toggle switch as a memory subsystem within the DNA damage sensor. The lambda repressor is naturally cleaved upon DNA damage and induction of the SOS response, leading to a lac-on/lambda-off state. The engineered cells also contained the *traA* gene, which activates biofilm formation under the control of lambda repressor. Consequently, exposure of the cells to DNA-damaging agents resulted in biofilm formation.

The first synthetic oscillator was a ring oscillator constructed by Elowitz and Leibler [22], where three repressors (the lac, lambda and tetracycline repressors) regulated the expression of the next repressor in the cycle (Figure 10.2b). A GFP reporter protein under the control of tetracycline repressor was used to monitor periodic changes of output. An important part of the design process was a rough quantitative model of the system to explore parameter spaces. A tightly regulated promoter and a shorter protein half-life improved the performance in the mathematical analysis, which was implemented in the experimental

design. The authors described a single plasmid construct, which suggests that the approximate calculation used to design the ring oscillator was enough to achieve oscillatory behavior in engineered *E. coli* cells. Interestingly, the oscillation period showed much more variability than did natural oscillators, with only 40% of the cells exhibiting oscillation. These findings suggested that the stability properties observed in wild-type circadian oscillators might result from the coupling of these clocks to other cellular processes. Alternatively, the architecture of the oscillator itself may dictate the stability of oscillation. In fact, the models of circadian oscillators fall in the class of relaxation oscillators [28], where a positive feedback loop and a negative feedback loop operate with slow and fast time scales. The synthetic oscillator design of Elowitz and Leibler does not fall into this category, but is a phase oscillator [29]. The oscillator design of Atkinson *et al.* [23] involved a positive autoregulatory circuit linked to a repressor module, analogous to the relaxation oscillator model of Barkai and Leibler [28]. Atkinson and colleagues used the components of a nitrogen-regulated response system for the activation signal and LacI for the inhibitory signal. This design did not involve a degradation sequence, as was used by Elowitz and Leibler [22], to shorten the protein lifetime, and the experiments were performed in a continuous bioreactor under constant cell density. Surprisingly, this oscillator displayed oscillation dynamics at population level, despite the oscillation being damped. Through mathematical analysis, the authors suggested a variety of parameter changes, such as messenger RNA stability and protein stability, to achieve sustained oscillation. Yet, an experimental exploration of such parameter change was not achieved and the mechanism for synchronization was unclear [30].

RNA molecules play important and diverse regulatory roles in the cell by virtue of their interaction with other nucleic acids, proteins and small molecules. For instance, diverse *cis* and *trans* gene regulation by noncoding RNA molecules such as microRNAs [31] and antisense RNAs [32] have been characterized in natural organisms. Researchers have engineered RNA molecules with new biological functions realized in bacteria and yeast [26, 33]. Isaacs *et al.* [33] achieved the repression of a target gene by forming a hairpin structure in the 5′ untranslated region (UTR) of the mRNA (*cis*-regulator), sequestering the ribosome-binding sequence. The expression of a targeted *trans*-RNA activator allowed translation from modified mRNA by exposing the ribosome-binding sequence. Bayer and Smolke [26] developed RNA regulatory molecules that have an aptamer domain to recognize specific effector molecules and an antisense domain to control gene expression, analogous to naturally found riboswitches (Figure 10.2c). The specific and dose-dependent switching responses of these regulatory RNA molecules have been demonstrated; for example, theophylline and tetracycline were each used to control the expression of GFP and yellow fluorescent protein (YFP) reporter proteins, without significant crosstalk. The stem stability of the designed RNA regulators turned out to be an important parameter that shifted switching thresholds. These results point to an intriguing possibility where designed RNA switches can be employed as cellular sensors and effectors to create programmable cells [34]. However, the engineered synthetic RNA regulation systems mainly demonstrated switching behavior rather than general network construction; consequently, quantitative models for the dynamics of RNA regulators need to be developed.

The signal transduction cascades composed of multiple proteins with enzymatic and structural interactions mediate many cellular functions and interactions with the environment. The interaction domains within signaling proteins can be rearranged to create

novel interactions. For example, when Dueber *et al.* [27] described the modular reprogramming of an allosteric protein signaling switch in yeast, their hybrid protein was constructed with an N-WASP-regulated actin polymerization output domain, a PDZ domain and a PDZ ligand (Figure 10.2d). This synthetic design has autoinhibitory architecture because the binding of a PDZ domain and a PDZ ligand blocks actin polymerization output, analogous to its natural counterpart GTPase-binding domain that represses actin polymerization. An external supply of PDZ ligand releases this autoinhibition in a dose-dependent manner. Furthermore, a library of hybrid proteins was created using PDZ- and SH3- binding domains with a variety of ligand affinities. Exploiting novel protein–protein interactions in addition to transcriptional regulation will enlarge the design space of synthetic networks.

10.3.2 Synthetic Networks *In Vitro*

An *in vitro* reconstruction with known components offers a unique opportunity to investigate how system behavior derives from reaction mechanisms. The first nontrivial system behavior created by an *in vitro* chemical system was the Belousov–Zhabotinsky oscillator [35], although it was difficult to see how these reaction mechanisms could support a wide variety of chemical logic, as is found in biochemistry. An excellent example of *in vitro* reconstruction using biochemical components is the cyanobacterial circadian clock, the operation of which has been shown to be independent of transcription and translation [36]. Operating and characterizing biochemical circuits in a cell-free system present some challenges, partly due to the complexity of synthesis machinery. A reconstituted cell-free transcription–translation system requires almost 100 purified components [37] or poorly characterized cell extracts [38]. Yet, several research groups were able to successfully construct a variety of interesting circuits within cell-free transcription–translation systems. For instance, Noireaux *et al.* [38] constructed transcriptional activation and repression cascades, where the protein product of each stage activated or inhibited the following stage. Isalan *et al.* [39] constructed a transcription–translation network that emulated *Drosophila* embryonic patterns and, by utilizing regulatory interaction mediated by previously characterized zinc-finger proteins, different network connections were tested. The patterning behavior was qualitatively correct and more mutual repression led to an overall lower activity, but with sharper patterns. Moreover, the addition of a protease stabilized the pattern over time. Thus, these bare-bone *in vitro* systems can be used to illustrate design principles, although further refinement of model systems and quantitative characterization would be required.

Nucleic acid-based networks greatly reduce the complexity of the production machinery. For example, feedback circuits modeled after predator–prey dynamics have been constructed as a much simpler *in vitro* system containing only three enzymes – T7 RNA polymerase, M-MLV reverse transcriptase and *E. coli* RNase H [40, 41]. The reaction scheme is based on self-sustained sequence replication, an isothermal amplification scheme for the coupled amplification of both DNA and RNA oligomers [42]. Mathematical modeling suggests that coupling prey and predator cycles (where the prey cycle provides a primer for the predator cycle) with an appropriate flow rate in a chemostat can lead to oscillation. Yet, a quantitative agreement of models and experiments was not achieved, possibly because of unmodeled dead-end side reactions and further couplings of reaction

rates by the common use of enzymes. Kim *et al.* [43] presented an alternative approach which relied on the transcription and degradation of RNA signals rather than replication and dilution. These authors constructed and analyzed feedforward circuits and a bistable mutual repression circuit with reasonable agreement to a mathematical model. However, it remains to be seen whether such nucleic acid-based networks can be utilized for regulating cellular behavior.

10.4 Challenges in Synthetic Networks

10.4.1 Saturation of Degradation Machinery *In Vivo*

Predictions about network behaviors through computational modeling and analytical theory is central to computational and systems biology. Many models of biological systems use simplifying assumptions [22, 44, 45] such as no spatial dependency of molecular species and no crosstalk between promoters. It is a widely accepted abstraction to view translation, transcription and degradation as composite processes, neglecting the detailed underlying reactions; however, these simplifying assumptions turned out to be inappropriate in some cases.

The transcriptional regulatory networks of Guet *et al.* [46] used three repressors – the lac repressor, lambda repressor and tetracycline repressor – with combinatorially assigned promoters; this allowed for a total of 27 different network topologies. The output of the network was monitored using GFP under control of the lambda repressor. Experimentally, GFP outputs were measured under four conditions: (1) without effector; (2) with IPTG, which inhibits LacI; (3) with anhydrotetracycline (aTc), which inhibits TetR; and (4) with both effectors. Kim and Tidor [47] studied the behavior of these combinatorial circuits by assuming a monotonic dependency of transcription, translation and degradation reactions to substrates and effectors, without detailed functional description or parameterization. Thus, without any detailed measurements of regulatory functions, it was possible to predict – for certain network topologies – the network output as upregulation, downregulation, or no change. Interestingly, two networks of equivalent topology (but with interchanged regulatory elements) showed different behavior in the study conducted by Guet and coworkers. According to the model, the addition of IPTG to the first network led to an increased production of both LacI and TetR, as the effect of LacI autorepression was decreased. Consequently, the model predicted that the cI level would decrease and the GFP output level would increase, in contrast to the experimental observations (Figure 10.3a). However, the addition of aTc in network 2 showed an increase of GFP output level, as predicted by the model (Figure 10.3b).

After ruling out some of the potential weakness of their model, such as not accounting for cell growth and stochastic noise, Kim and Tidor proposed that the saturation of degradation machinery could be one possible mechanism to reconcile the experimental results and model predictions. As all three repressors of the synthetic network were known to carry ssrA tags, they would be degraded by a special cellular machinery, the Clp system [48]. Because the components of Clp system are at fairly low cellular concentrations, this degradation machinery could be saturated. In network 1, IPTG released the LacI repression on both LacI and TetR production, which in turn reached high cellular concentrations and

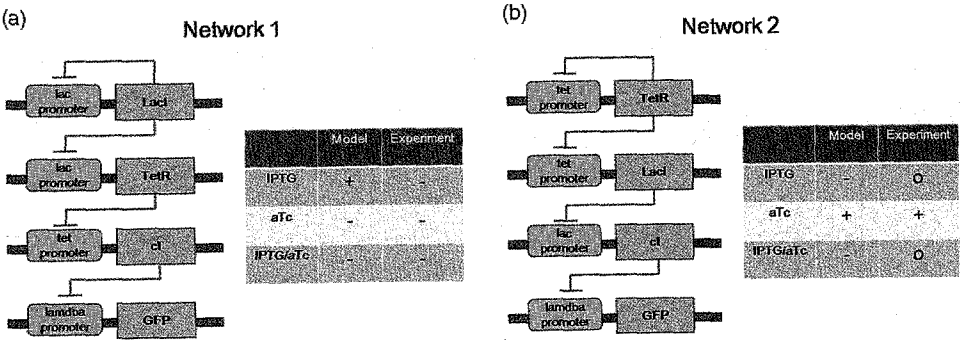

Figure 10.3 *Experimental and predicted behavior of synthetic repressor networks consisting of lac, lambda and tetracycline repressors. The two networks shown in (a) and (b) have identical topologies but with lac repressor and tetracycline repressor interchanged. In the rows titled IPTG, aTc and IPTG/aTc, the GFP level changes are shown as + (increase), − (decrease), 0 (no change). (Reproduced by permission of Cold Spring Laboratory Press. Ref. [47])*

outcompeted cI for degradation. Consequently, under saturating degradation conditions cI could potentially accumulate, even with basal expression rates. A direct measurement of cellular repressor levels could answer the question of the validity of this scenario. The results of this study show that care must be taken for a seemingly general assumption such as the monotonic dependency of production and degradation functions on substrates and effectors, particularly with synthetic networks that introduce new components and novel interactions among the cellular machinery.

10.4.2 Saturation of Production Machinery *In Vitro*

Noireaux and colleagues [38] characterized the cell-free genetic circuits constructed in a transcription–translation extract by engineering transcriptional activation and repression cascades in which the protein product of each stage was the input required to drive or block the following stage. The protein expression reactions were carried out in batch mode, without any continuous exchange of nutrients and byproducts. In order to boost protein production, 5′-polyguanylic acid was used to increase the mRNA lifetime [49] from 20–30 min to 2 h. At the same time, both the creatine phosphate concentration (for ATP regeneration) and the magnesium concentration were adjusted to optimal levels.

A single-level cascade was constructed as a T7-luc plasmid composed of T7 RNA polymerase promoter site and firefly luciferase gene. Upon the addition of T7 RNA polymerase, this single-level cascade began to accumulate luciferase protein after 15 min, reaching a maximum concentration of 500 nM after 6 h. A two-stage cascade, constructed with the plasmids T7-SP6RNAP and SP6-luc, used SP6 RNA polymerase produced from T7-SP6RNAP plasmid to drive the production of luciferase output from the luciferase gene downstream of SP6 polymerase promoter (Figure 10.4a). The two-stage cascade started to produce luciferase after a 1 h delay, such that the final luciferase level was 100 nM – fivefold less than for a single-stage cascade. A three-stage cascade constructed with the plasmids T7-SP6RNAP, SP6-rpoF and Ptar-luc using *E. coli* sigma factor F from the *rpoF*

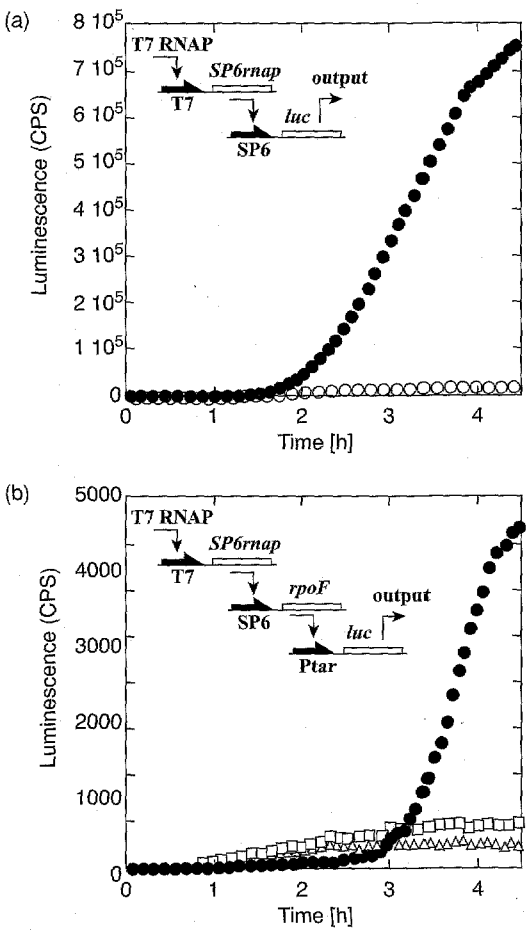

Figure 10.4 *Two- and three-stage cascades. (a) Kinetics of expression of the cascade with T7 RNA polymerase, and both T7-SP6rnap and SP6-luc plasmids (filled circles) or SP6-luc plasmid only (open circles); (b) Kinetics of expression of the cascade with T7 RNA polymerase and all three plasmids (filled circles) or two plasmids, SP6-rpoF and Ptar-luc (open squares) or Ptar-luc only (open triangles). (Copyright National Academy of Sciences, U.S.A. Ref. [38])*

gene as a new relay signal (Figure 10.4b) produced luciferase after about a 3 h delay, reaching a final concentration of only 1 n*M* after 6 h. Interestingly, substantial time delays and dramatic decreases in output were observed with each additional stage.

A detailed characterization of the two-stage cascade with various RNA polymerase and plasmid concentrations revealed that the translation machinery was saturated for the combinations of polymerase and plasmid concentrations which resulted in high transcript concentrations. Above the first-stage transcription rate, that maximized luciferase production, the overproduced first-stage mRNA occupied translation machinery and

inhibited luc mRNA translation. In contrast, luciferase production did not show saturation for similar RNA polymerase and plasmid concentrations if short-lifetime mRNAs without polyG modification were transcribed. The results of the study indicated that a conventional approach of maximizing single-protein synthesis in cell-free systems must be reconsidered for *in vitro* gene circuits. The authors suggested that a rapid turnover of mRNA might avoid saturation of the translation machinery and that implementing gene autoregulation would prevent overproduction.

In a follow-up study of the cell-free expression system, Noireaux and Libchaber [50] employed the phospholipid encapsulation of synthesis machinery to construct a vesicle bioreactor. Without access to nutrients outside, the vesicle bioreactor could not prolong the expression of reporter proteins by more than 5 h. In order to solve the material and energy limitation, the α-hemolysin pore protein from *Staphylococcus aureus* was expressed inside the vesicle to create a selective permeability for nutrients. Subsequently, the vesicle bioreactor thus created could take up nutrients from a feeding solution containing amino acids and nucleic acids, and maintained protein expression for up to four days. This study proved to be an important step towards the synthesis of a minimal, self-reproducing cell.

10.4.3 Saturation in a Mutual Repression Circuit

The saturation of production and degradation machinery has a significant impact on the network dynamics. Take an example of a mutual repression circuit where two repressors, X and Y, downregulate the synthesis rates of each other (Figure 10.5a). By assuming equivalence of the two repressors, the behavior of the circuit can be understood using the following dimensionless model (Equation 10.1):

$$\begin{aligned} \frac{dx}{dt} &= \frac{\alpha}{1+y^n} - x, \\ \frac{dy}{dt} &= \frac{\alpha}{1+x^n} - y, \end{aligned} \tag{10.1}$$

where x and y are the concentrations of the repressors, α is the effective synthesis rate of repressors, and n is the cooperativity of repressor binding. The repressor binding to promoter is fast compared to transcription, translation and degradation processes. Therefore, it is assumed that the promoter–repressor binding is already at steady state when considering repressor production and degradation dynamics. Thus, the fraction of active gene x with an unoccupied promoter region can be described by $1/(1+y^n)$, and similarly for the fraction of active gene y. With the repressor cooperativity >1 and for a large synthesis rate, the two nullclines ($dx/dt = 0$ and $dy/dt = 0$) were seen to intersect at three points, producing one unstable and two stable steady states [21]. The nullclines for the circuit with cooperativity of two and maximum production rate of five indicates such bistable behavior (Figure 10.5b).

Consider the case where the production machinery is saturated for the mutual repression circuit. Assuming that α is the maximum synthesis rate for the system, and that the sharing of synthesis machinery is strictly between two repressor genes with unoccupied promoters,

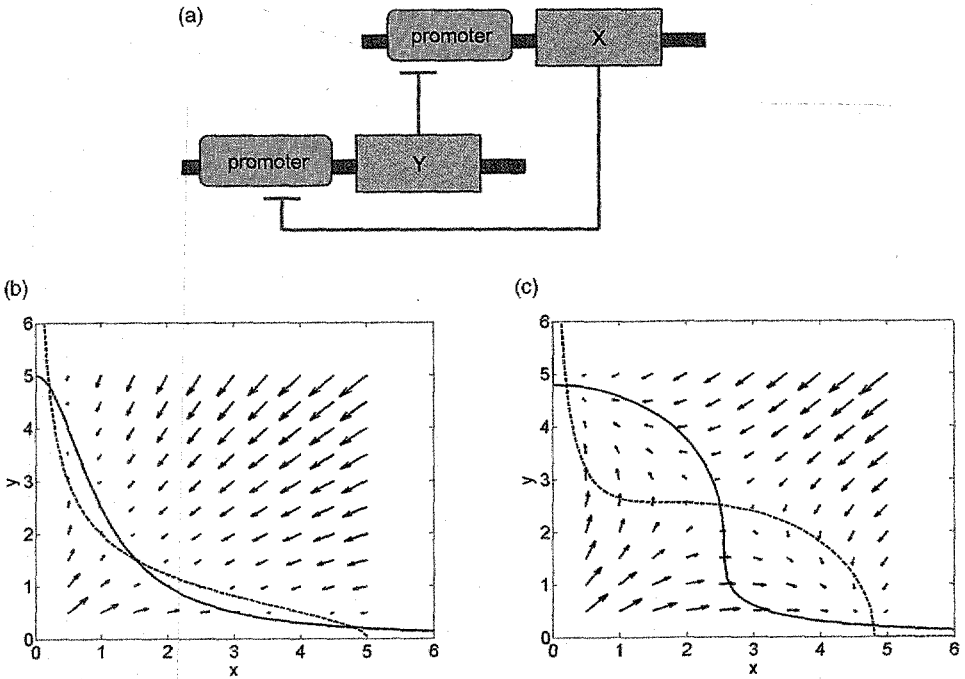

Figure 10.5 *Dynamics of a mutual repression system. (a) A mutual repression system constructed from two repressors, X and Y, that repress the expression of each other; (b, c) Dynamics of mutual repression system without saturation of synthesis machinery (b) or with saturated synthesis machinery (c). Nullclines are drawn for both $dx/dt = 0$ (dotted line) and $dy/dt = 0$ (solid line) with vector flow (arrows). The parameters are $\alpha = 5$ and $n = 2$*

the behavior of the circuit can be described using the following dimensionless model (Equation 10.2):

$$\begin{aligned}\frac{dx}{dt} &= \alpha \cdot \frac{\frac{1}{1+y^n}}{\frac{1}{1+y^n}+\frac{1}{1+x^n}} - x = \alpha \cdot \frac{1+x^n}{2+x^n+y^n} - x,\\ \frac{dy}{dt} &= \alpha \cdot \frac{\frac{1}{1+x^n}}{\frac{1}{1+x^n}+\frac{1}{1+y^n}} - y = \alpha \cdot \frac{1+y^n}{2+x^n+y^n} - y.\end{aligned} \tag{10.2}$$

The nullclines for the circuit intersect at three points with a cooperativity of two and a maximum synthesis rate of five, analogous to the previous example (Figure 10.5c). However, the circuit dynamics around the unstable steady state is different: the approach towards the unstable steady state is slower, while the exit from the unstable steady state is faster than the previous example. The production of repressor X, in effect, inhibits the production of repressor Y because the two promoters compete for the same synthesis machinery. Thus, it is expected that saturated production leads to bistability even when the repressor cooperativity is relatively low. For example, bistability is achieved for the repressor cooperativity of 1.4 and a maximum synthesis rate of five with saturated synthesis

(Equation 10.2), but bistability is not achieved for the same parameters in the other model (Equation 10.1).

On the other hand, saturation of the degradation machinery would be detrimental to the bistability of a mutual repression circuit because the accumulation of one repressor would allow an accumulation of the other repressor. In natural organisms, it is rarely the case that a few proteins dominantly occupy the synthesis and degradation machinery. However, for synthetic networks *in vivo* or *in vitro*, inducing the overproduction of network elements can lead to the saturation of such machinery. Hence, saturation effect must be carefully modeled, depending on the context, and can potentially be exploited for circuit operation.

10.4.4 Waste Product in an *In Vitro* Oscillator

Kim and colleagues developed an experimental analogue to a genetic regulatory circuit that uses only T7 RNA polymerase and *E. coli* RNase H in addition to synthetic DNA templates regulated by RNA transcripts [43]. A synthetic template – a gene analogue – consists of a regulatory domain, a promoter and an output domain. Each synthetic template requires a DNA oligonucleotide activating signal that complements the promoter region for a strong transcription of its output. The addition of an RNA inhibitor complementary to the DNA-activating signal hybridizes to – and consequently eliminates – the DNA-activating signal from the target synthetic template and greatly reduces transcription rates. At the same time, the degradation of RNA signals by RNase H releases the DNA signals from a functionally inert DNA–RNA hybrid state. Thus, the difference of activating and inhibitory signals determines the transcription speed of outputs. Consequently, a sigmoidal response curve with adjustable thresholds is achieved through a competitive binding of nucleic acid species.

A two-node oscillator was constructed as follows. An RNA activator (rA) activates the production of an RNA inhibitor (rI) by regulating a synthetic template (gene I), while the RNA inhibitor, in turn, inhibits the production of RNA activator by controlling gene A (Figure 10.6a). These two genes form a negative feedback loop and can potentially show oscillatory behavior. By measuring RNA signals, up to six oscillation cycles were observed before the production rate could no longer be sustained due to exhaustion of the NTP fuel (Figure 10.6c). Interestingly, the concentration of rI was seen to build up after each cycle, although it was expected that the RNA inhibitor signal would oscillate around a fixed threshold, the concentration of DNA-activating signal. One hypothesis was that the short fragments of rI generated by degradation process might interfere with the correct hybridization reaction of rI signals to its regulatory target, gene A, and therefore, more signals would be needed to overcome the interference. The short fragments of rI produced by RNase H processing would encompass the toehold binding sequence of rI because RNase H cannot process several bases on the 5′ side of the RNA strand on an RNA/DNA hybrid substrate [51]. Thus, the short fragment of rI could block the (otherwise freely available) toehold region that was essential for providing a fast kinetic pathway [52]. The concentration of short degradation products estimated from the gel showed a linear build-up over time (Figure 10.6b). Intriguingly, subtracting a fraction of short products from rI signal resulted in an oscillation around a fixed threshold (Figure 10.6c). A mathematical model taking account of the interference from short products was able to reproduce these experimental observations qualitatively. Taken together, the *in vitro* oscillator

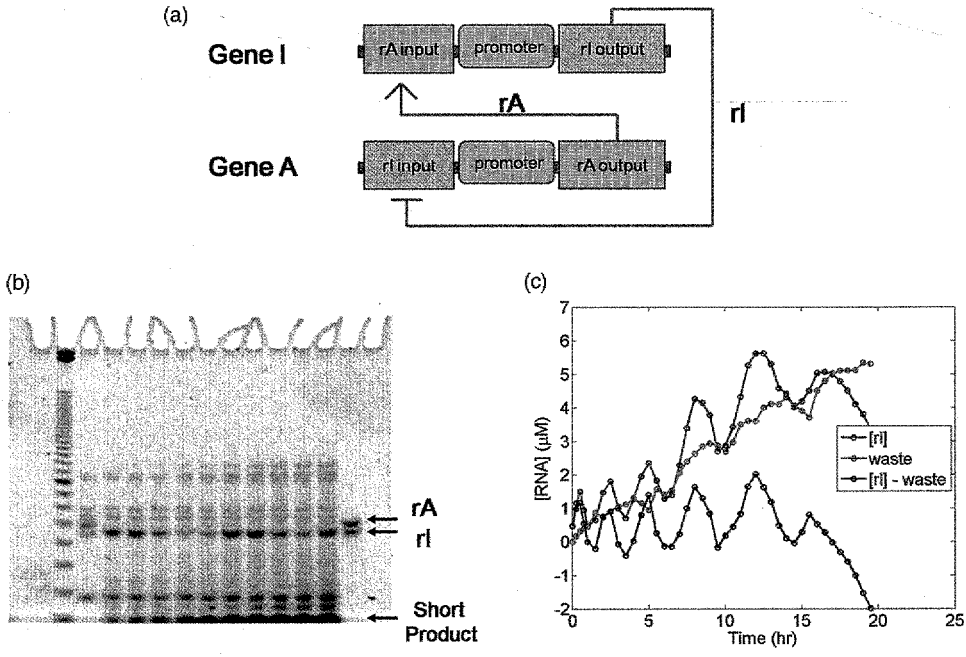

Figure 10.6 *A synthetic* in vitro *oscillator. (a) The synthetic oscillator is composed of two gene analogues, an activator and a repressor; (b) Gel measurement of oscillator outputs up to 4 h. The leftmost lane contains size markers, while the rightmost lane contains purified rA and rI products. It is assumed that the band of ~35 nucleotides in the gel is representative of accumulating short products; (c) The rI signal, the short product level normalized to be of similar scale to rI, and rI signal minus the normalized short product level are shown*

demonstrated sustained oscillations and was robust to the build-up of interfering signals to some extent. However, for a sustained and reliable operation of oscillators the incomplete degradation products need to be further processed, ultimately to the mononucleotide level.

Lessons can be learned from the degradation machinery of natural organisms. For example, *E. coli* has a high-molecular-weight complex called the degradosome which consists of RNase E, polynucleotide phosphorylase (PNPase), an ATP-dependent helicase, RhlB and enolase, a glycolytic enzyme [53]. When the decay of mRNA is initiated via endonucleolytic cleavage by RNase E, the newly formed 3′ end can be attacked by PNPase, which performs processive exonucleolytic digestion. The ATP-dependent RNA helicase in the degradosome presumably helps the degradation by unwinding RNA structures that impede the cleavage by RNase E and PNPase. The concerted action of these enzymes would explain the observation that, once initiated, the decay of mRNA proceeds without any accumulation of the decay intermediates. Although many mRNAs are subject to alternative decay processes, the existence of a highly orchestrated multienzyme complex such as the degradosome indicates that a complete degradation of messages without byproducts is an essential regulatory step.

10.5 The Minimal Cell

On a larger scale of synthetic efforts, the assembly of a type of cell – that is, a self-replicating, membrane-encapsulated collection of biomolecules – would be the next major challenge [54]. However small, a cellular gene set must be self-sufficient in the sense that cells generally import metabolites, but not functional macromolecules. *Mycoplasma genitalium*, a parasitic bacterium with a small genome size, is recognized as an attractive model in the search for the minimal genome. After comparing the 468 predicted *M. genitalium* protein sequences with the 1703 *Haemophilus influenzae* protein sequences, Mushegian and Koonin [55] suggested 256 genes as a minimal genome set, including 234 *M. genetalium* genes. Most of the proteins encoded by genes from the minimal set suggested by these authors had eukaryotic or archaeal homologues, whereas the key proteins of DNA replication did not, which led these authors to speculate that the last common ancestor had an RNA genome. The estimated gene number could be further reduced by eliminating cofactors and regulatory genes, and by applying the parsimony principle [56].

A recent estimate suggested that the minimal genome would comprise 151 genes, 38 RNAs and 113 proteins [54]. Lipids alone have been shown to be sufficient for the formation of rudimentary membranous compartments capable of both the transmembrane transport of small molecules and autocatalytic fission [57]. A bare-bones genome would perform basic DNA replication, transcription and translation processes, in which alternative approaches for essential mechanisms such as the adaptation of rolling circle amplification for DNA replication were employed to reduce the number of genes. A surprisingly large fraction (96%) of the minimal gene set is devoted to translation mechanisms, including ribosome components, a set of transfer RNAs (tRNAs), a set of translational initiation, elongation and release factors, and a few chaperones. In light of this, the simplest approach for creating a minimal cell may be to evolve an RNA polymerase made exclusively from RNA that would replace all of the protein components of the *in vitro* replicating and evolving systems [57]. An exciting development in this direction is the templated assembly of RNA products catalyzed by ribozymes [58]; these ribozymes used nucleoside triphosphates and the coding information of an RNA template to extend an RNA primer by the successive addition of up to 14 nucleotides, with high accuracy. These findings support the 'RNA-world' hypothesis regarding the early evolution of life – the main tenet of which is that ribozymes would have been far easier to duplicate than proteinaceous enzymes. Given that most of the minimal gene set is devoted to translation, a nucleic acid-based artificial cell would certainly be attractive, justifying a search for different sets of ribozymes through *in vitro* evolution approaches.

Estimates of the minimal genome typically do not include catabolism (nucleases and proteases), the active conversion or removal of waste products (energy-regenerating enzymes and membrane transporters) and regulatory feedback. It is unclear whether a minimal cell could sustain growth and replication without such regulatory mechanisms. At any rate, a much simpler purified system based on a real cell would be easier to model and understand, and it could certainly answer questions that cannot be answered *in vivo*, such as which set of macromolecules would be sufficient for a functional cellular subsystem [54]. The iterative synthetic process in which the performance of an *in vitro* model system is continuously improved may, in time, culminate in viable minimal cells as complex analogues of cells.

10.6 Conclusions

Today, synthetic biology provides the ability to study cellular regulation and behavior using *de novo* networks, with future applications of synthetic systems extending also to the fields of medicine and biotechnology. Yet, challenges remain that call for novel approaches and creative solutions. Synthetic networks *in vivo* have recycled previously used parts because a single point mutation may alter the *in vivo* activity of the network, and it is difficult to predict how redesigned molecules such as synthetic promoters would behave [59]. Mutations and the loss of synthetic network control can be a serious problem, especially when a large population of cells is considered. A 'population control' circuit [60] has been described which utilized a bacterial quorum-sensing system linked to a cell death signal to regulate the cell density of an *E. coli* population. Here, the steady-state cell density in the regulated cell culture was about tenfold lower than that of the control culture. However, due to the disadvantage in growth rate, cells that acquired mutations to disrupt the synthetic circuit control easily outgrew the regulated cells. A microfluidic microreactor was used to alleviate this problem by greatly reducing the population size [61], and allowed the synthetic circuit behavior to be monitored over hundreds of hours. Engineered cells would retain the synthetic network design that conferred a selective advantage in cellular growth rate, allowing further observation and analysis. For *in vitro* networks, the lack of any complex feedback regulation for the production and degradation machinery can lead to a high variability and a lack of robustness in their performances. As observed previously, dead-end side reactions, the saturation of the enzyme machinery and interference from incomplete products must be correctly addressed for successful *in vitro* network construction. Further developments of *in vitro* networks, accompanied by effective encapsulation in membranous compartments and ensuing growth and fission, will provide a good starting point for a minimal cell.

These synthetic approaches have successfully demonstrated several interesting networks, and have provided valuable engineering tools to study motifs, modularity and the robustness of cellular networks. Nonetheless, the development of new frameworks for regulatory costs, trade-offs and energy consumption of network structures remains a major problem, the solution of which could eventually lead to the construction of viable minimal cells.

References

1. Hartwell, L.H. Hopfield, J.J., Leibler, S. and Murray, A.W. (1999) From molecular to modular cell biology. *Nature*, **402**, C47–52.
2. Alves, R., Antunes, F. and Salvador, A. (2006) Tools for kinetic modeling of biochemical networks. *Nature Biotechnology*, **24**, 667–72.
3. Bray, D. (2003) Molecular networks: The top-down view. *Science*, **301**, 1864–5.
4. Alon, U. (2003) Biological networks: The tinkerer as an engineer. *Science*, **301**, 1866–7.
5. Savageau, M.A. (1971) Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature*, **229**, 542–4.
6. Shen-Orr, S.S., Milo, R., Mangan, S. and Alon, U. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics*, **31**, 64–8.

7. Milo, R. *et al.* (2002) Superfamilies of evolved and designed networks. *Science*, **303**, 1538–42.
8. Dekel, E., Mangan, S. and Alon, U. (2005) Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Physical Biology*, **2**, 81–8.
9. Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A. *et al.* (2002) A genomic network for development. *Science*, **295**, 1669–79.
10. Kashtan, N. and Alon, U. (2005) Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 13773–8.
11. Strogatz, S.H. (2001) Exploring complex networks. *Nature*, **410**, 268–76.
12. Mattick, J.S. and Gagen, M.J. (2005) Accelerating networks. *Science*, **307**, 856–8.
13. Hooshangi, S., Thiberge, S. and Weiss, R. (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 3581–6.
14. van Nimwegen, E. (2003) Scaling laws in the functional content of genomes. *Trends in Genetics*, **19**, 479–84.
15. Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. *Human Molecular Genetics*, **15**, R17–29.
16. Benner, S.A. and Sismour, A.M. (2005) Synthetic biology. *Nature Reviews in Genetics*, **6**, 533–43.
17. SantaLucia, J. Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 1460–5.
18. Becskei, A. and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. *Nature*, **405**, 590–3.
19. Becskei, A., Seraphin, B. and Serrano, L. (2001) Positive feedback in eukaryotic gene networks: Cell differentiation by graded to binary response conversion. *The EMBO Journal*, **20**, 2528–35.
20. Basu, S., Mehreja, R., Thiberge, S., Chen, M.T. *et al.* (2004) Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 6355–60.
21. Gardner, T.S., Cantor, C.R. and Collins, J.J. (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, **403**, 339–42.
22. Elowitz, M.B. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**, 335–8.
23. Atkinson, M.R., Savageau, M.A., Myers, J.T. and Ninfa, A.J. (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell*, **113**, 597–607.
24. Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I. *et al.* (2004) Multistability in the lactose utilization network of *Escherichia coli*. *Nature*, **427**, 737–40.
25. Kobayashi, H., Kaern, M., Araki, M. *et al.* (2004) Programmable cells: Interfacing natural and engineered gene networks. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 8414–19.
26. Bayer, T.S. and Smolke, C.D. (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature Biotechnology*, **23**, 337–43.
27. Dueber, J.E., Yeh, B.J., Chak, K. and Lim, W.A. (2003) Reprogramming control of an allosteric signaling switch through modular recombination. *Science*, **301**, 1904–8.
28. Barkai, N. and Leibler, S. (2000) Circadian clocks limited by noise. *Nature*, **403**, 267–8.
29. Schibler, U. and Naef, F. (2005) Cellular oscillators: rhythmic gene expression and metabolism. *Current Opinion in Cell Biology*, **17**, 223–9.
30. Wong, W.W. and Liao, J.C. (2006) The design of intracellular oscillators that interact with metabolism. *Cellular and Molecular Life Sciences*, **63**, 1215–20.
31. Carrington, J.C. and Ambros, V. (2003) Role of microRNAs in plant and animal development. *Science*, **301**, 336–8.

32. Kramer, C., Loros, J.J., Dunlap, J.C. and Crosthwaite, S.K. (2003) Role for antisense RNA in regulating circadian clock function in *Neurospora crassa*. *Nature*, **421**, 948–52.
33. Isaacs, F.J., Dwyer, D.J., Ding, C. *et al.* (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, **22**, 841–7.
34. Isaacs, F.J., Dwyer, D.J. and Collins, J.J. (2006) RNA synthetic biology. *Nature Biotechnology*, **24**, 545–54.
35. Zaikin, A.N. and Zhabotinsky, A.M. (1970) Concentration wave propagation in 2-dimensional liquid-phase self-oscillating system. *Nature*, **225**, 535–7.
36. Nakajima, M., Imai, K., Ito, H. *et al.* (2005) Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science*, **308**, 414–15.
37. Shimizu, Y., Inoue, A., Tomari, Y. *et al.* (2001) Cell-free translation reconstituted with purified components. *Nature Biotechnology*, **19**, 751–5.
38. Noireaux, V., Bar-Ziv, R. and Libchaber, A. (2003) Principles of cell-free genetic circuit assembly. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 12672–7.
39. Isalan, M., Lemerle, C. and Serrano, L. (2005) Engineering gene networks to emulate *Drosophila* embryonic pattern formation. *PLoS Biology*, **3**, e64.
40. Wlotzka, B. and McCaskill, J.S. (1997) A molecular predatory and its prey: Coupled isothermal amplification of nucleic acids. *Chemistry and Biology*, **4**, 25–33.
41. Ackermann, J., Wlotzka, B. and McCaskill, J.S. (1998) In vitro DNA-based predator-prey system with oscillatory kinetics. *Bulletin of Mathematical Biology*, **60**, 329–53.
42. Guatelli, J.C., Whitfield, K.M., Kwoh, D.Y. *et al.* (1990) Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 1874–8.
43. Kim, J., White, K.S. and Winfree, E. (2006) Construction of an in vitro bistable circuit from synthetic transcriptional switches. *Molecular Systems Biology*, **2**, 68.
44. Arkin, A., Ross, J. and McAdams, H.H. (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *E. coli* cells. *Genetics*, **149**, 1633–48.
45. Thattai, M. and van Oudenaarden, A. (2001) Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 8614–19.
46. Guet, C.C., Elowitz, M.B., Hsing, W. and Leibler, S. (2002) Combinatorial synthesis of genetic networks. *Science*, **296**, 1466–70.
47. Kim, P.M. and Tidor, B. (2003) Limitations of quantitative gene regulation models: A case study. *Genome Research*, **13**, 2391–5.
48. Keiler, K.C., Waller, P.R.H. and Sauer, R.T. (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science*, **271**, 990–3.
49. Shen, X.-C. *et al.* (1999) Poly[G] improved protein productivity of cell-free translation by inhibiting mRNase in wheat germ extract. *Journal of Biotechnology*, **75**, 221–8.
50. Noireaux, V. and Libchaber, A. (2004) A vesicle bioreactor as a step toward an artificial cell assembly. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 17669–74.
51. Lima, W.F. and Crooke, S.T. (1997) Cleavage of single strand RNA adjacent to RNA-DNA duplex regions by *Escherichia coli* RNase H1. *Journal of Biological Chemistry*, **272**, 27513–16.
52. Yurke, B. and Mills A.P. Jr. (2003) Using DNA to power nanostructures. *Genetic Programming and Evolvable Machines*, **4**, 111–22.
53. Grunberg-Manago, M. (1999) Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annual Review of Genetics*, **33**, 193–227.
54. Forster, A.C. and Church, G.M. (2006) Towards synthesis of a minimal cell. *Molecular Systems Biology*, **2**, 45.

55. Mushegian, A.R. and Koonin, E.V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 10268–73.
56. Benner, S.A., Ellington, A.D. and Tauer, A. (1989) Modern metabolism as a palimpsest of the RNA world. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 7054–8.
57. Szostak, J.W., Bartel, D.P. and Luisi, P.L. (2001) Synthesizing life. *Nature*, **409**, 387–90.
58. Johnston, W.K., Unrau, P.J., Lawrence, M.S. *et al.* (2001) RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science*, **292**, 1319–25.
59. Ventura, B.D., Lemerle, C., Michalodimitrakis, K. and Serrano, L. (2006) From in vivo to in silico biology and back. *Nature*, **443**, 527–33.
60. You, L., Cox, R.S. 3rd, Weiss, R. and Arnold, F.H. (2004) Programmed population control by cell-cell communication and regulated killing. *Nature*, **428**, 868–71.
61. Balagadde, F.K., You, L., Hansen, C.L. *et al.* (2005) Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science*, **309**, 137–40.