

Chapter 1.

Introduction*

*Sections 1.1-1.7: Reproduced/adapted with permission from: Liang, J.C., Bloom, R.J. and Smolke, C.D. (2011) Engineering biological systems with synthetic RNA molecules. *Mol Cell*, **43**, 915-926.

Engineered biological systems have exciting potential in developing solutions to many global challenges, including environmental remediation, sustainability, scalable manufacturing, and health and medicine (1-4). Our ability to design and build synthetic biological systems is a key technology to improving the human condition. In addition, the redesign of biological systems can be used as an effective strategy to test, and thereby strengthen, our understanding of natural systems. Synthetic biology is an emerging research field with a primary goal of making the engineering of biology faster, less expensive, and more reliable. As such, core activities in synthetic biology have been focused on the development of foundational tools and technologies that assist in the design, construction, and characterization of biological systems (5,6). Recent advances in construction and fabrication technologies are supporting synthesis of large pieces of DNA including entire pathways and genomes (7). While progress has been made in the design of complex genetic circuits (8), current capabilities for constructing large genetic systems surpass our ability to design such systems. This growing ‘design gap’ has highlighted the need to develop methods that support the generation of new functional biological components and scalable design strategies for complex genetic circuits that will lay the foundation for integrated biological devices and systems.

The vast majority of genetic systems engineered to-date have utilized protein-based transcriptional control strategies (8). However, as the examples of functional RNA molecules playing key roles in the behavior of natural biological systems have grown over the past decade, there has been growing interest in the design and implementation of synthetic counterparts. Researchers have taken advantage of the relative ease with which RNA molecules can be modeled and designed to engineer functional RNA molecules that

act as diverse components including sensors, regulators, controllers (ligand-responsive RNA regulators), and scaffolds. More recently, researchers have begun to move beyond molecular design and integrate these synthetic RNA molecules as key elements in genetic circuits to program cellular behavior, highlighting the relevance and advantages of RNA-based control strategies.

1.1 RNA as a natural regulatory molecule

The growing interest in using RNA to build synthetic controllers is due in large part to the steadily increasing examples of natural RNA regulators that control gene expression through diverse mechanisms in different organisms. One of the earliest examples is the regulation of gene expression through RNA secondary structure. The study of differential expression of genes in phage genomes led to the discovery that secondary structure of a mRNA transcript can restrict access to the ribosome binding site (RBS), thereby inhibiting translation (9). Similarly, bacteria utilize the formation of tight hairpins in mRNA transcripts to stall and attenuate translation in the regulation of amino acid biosynthesis (10). In addition, RNA structure is highly temperature sensitive, such that in certain cases hairpin structures that inhibit translation and can be modulated by temperature have been found to have functional roles in the heat and cold shock responses of several bacteria (9). Finally, it has been shown that strong secondary structures on the 5' and 3' ends of a mRNA strand can protect the transcript from degradation by exoribonucleases and endoribonucleases (11). The resulting extended half-lives of the transcripts can significantly increase protein production and have functional roles in processes such as photosynthesis and bacterial cell adhesion (11,12).

In addition to structural mechanisms, the discovery that RNA can exhibit catalytic activity opened the door to a wider array of regulatory functions (13). These catalytic RNAs, or ribozymes, typically catalyze cleavage or ligation of the RNA backbone through a reversible phosphodiester cleavage reaction (14). Ribozymes have functional roles in alternative splicing, RNA replication, translation, and transcript stability and function in both prokaryotes and eukaryotes (14). Furthermore, the discovery that ribozyme cleavage of the *glmS* transcript in bacteria is inhibited by binding of the metabolite GlcN6P has led to several discoveries of ribozymes acting as key components in riboswitches, a class of RNA regulators that respond to cellular metabolites and cofactors to modulate enzyme levels in related biosynthesis (15). Finally, RNase P is a catalytic RNA that functions in trans and can carry out multiple turnover cleavage events in the processing of 5' leader sequences from tRNA (14). The discovery of ribozymes with natural gene regulatory activity in trans presents an intriguing proof of principle that a single catalytic RNA can be used to regulate several different genes in a biological system.

The last major mechanism that RNA uses to regulate protein synthesis is through antisense-mediated regulation of translation. Trans-acting small RNAs in bacteria are generally transcribed from their own independent transcripts and can range from 50-514 nucleotides (nts) in length (16). These RNAs often act in concert with the RNA binding protein Hfq and can promote or inhibit translation of their target mRNAs by relieving secondary structural elements or inhibiting ribosome initiation or processivity (17). Similarly, in higher eukaryotes, RNA interference (RNAi) pathways use small RNAs, siRNA and microRNAs, to guide protein complexes to complementary mRNAs, leading

to silencing of those targets (18). MicroRNAs have diverse roles in almost every cellular process and they are currently thought to regulate up to a third of human genes (19). RNAi, particularly the siRNA pathway, is widely used as a tool in biological research for genetic loss of function studies and is currently being explored for therapeutic and biotechnological uses (20).

Finally, RNA is involved in gene regulation in several other less widespread or well-studied mechanisms. For example, noncoding RNAs have been characterized that have roles in quality control of translation (16), binding to and inhibiting proteins involved in protein synthesis, and epigenetic DNA modification (21). Furthermore, the recent discovery of the widespread regulatory activity and conservation of long noncoding RNAs represents an exciting area for further research in RNA-based gene regulation (22).

1.2 RNA as a programmable and efficient substrate for engineering biological controllers

In addition to the diversity of mechanisms by which RNA can act as a regulatory molecule in nature, RNA exhibits several properties that make it an attractive design substrate in synthetic systems. RNA is composed of four building blocks that interact through well-characterized hydrogen-bond, base-stacking, and electrostatic interactions. The folding of RNA is primarily dictated by its secondary structure, in contrast to the folding of proteins, which involves a large degree of tertiary interactions. Models that predict RNA secondary structure have been developed based on the optimization of energies contributed by the Watson-Crick AU and GC base pairs as well as the GU

wobble pair (23-25). Progress in RNA three-dimensional structural studies have revealed a set of non-canonical base pairing interactions, or Hoogsteen base pairs, that are key in the formation of the RNA tertiary structure. Early modeling frameworks that predict RNA tertiary structure based on a primary sequence have been developed and demonstrated to generate native-like structure predictions (26,27). In addition to structure prediction based on thermodynamics, RNA kinetic folding program has also been developed to capture the stochastic nature of the RNA folding process (28). These computational tools can be utilized to access structural information encoded in the primary RNA sequence, thereby aiding the rational design of genetic controllers based on hybridization schemes or structural elements.

RNA-based controllers exhibit additional advantages as functional elements in synthetic biological systems as the field moves toward more complex genetic systems design. RNA controllers generally exhibit more compact genetic footprints than their protein counterparts. In addition, RNA controllers generally place less of an energetic and resource load on the host cell as functional RNA molecules do not require the translation process to synthesize the functional elements. Another important consideration in genetic system design is timing of coupled control processes, where RNA-based posttranscriptional control strategies will generally act at faster time scales than transcription-based control strategies. The energetic, resource, and space efficiencies of RNA-based control strategies present important features supporting scaling to large-scale genetic system designs.

1.3 Approaches to generate functional RNA components

Functional RNA components are the basic building blocks for constructing genetic devices that encode human-defined functions. A fundamental challenge in the field then is the efficient generation of new component functions, such as sensing, information transmission, and actuation functions, that can be assembled into more complex devices. Three general approaches are taken to engineer new functional RNA components: harvesting from nature, computational design, and molecular evolution (Figure 1.1).

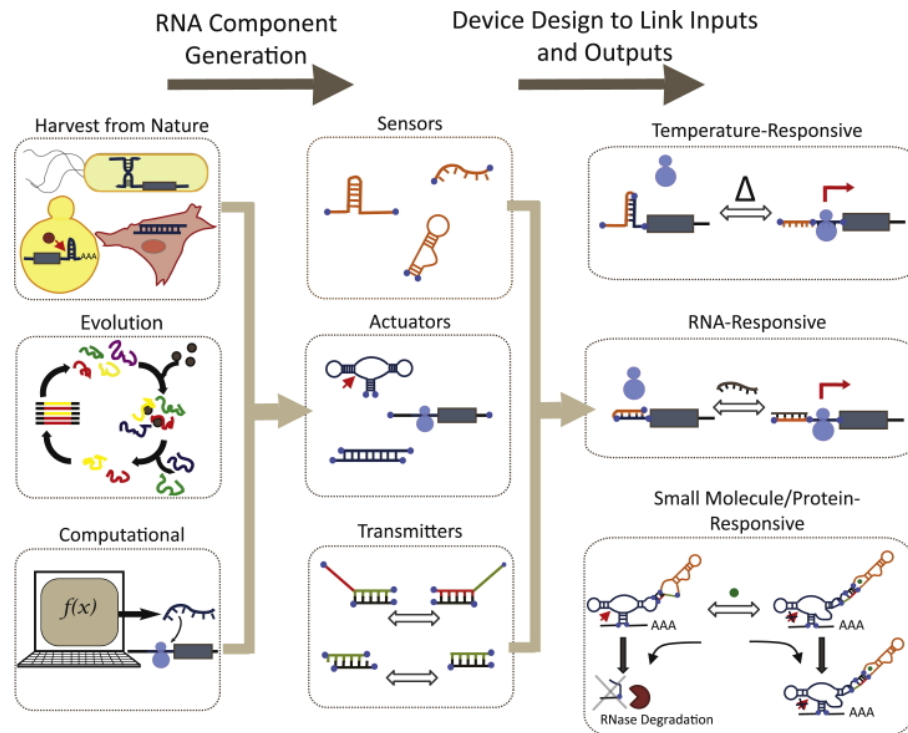


Figure 1.1 RNA components used to engineer synthetic control functions can be harvested from natural systems or generated using molecular evolution and computational approaches. These components, encoding sensing, actuation, and information transmission activities, can then be assembled into RNA devices using various molecular engineering strategies to link one or more inputs of interest, such as temperature, RNA, small molecules, or proteins, to desired regulatory activities.

1.3.1 Harvesting and refining RNA components from nature

Many functional RNA components are derived from naturally-occurring elements, as advances in RNA biology have led to discoveries of natural RNA molecules that exhibit diverse functional activities. In certain instances the RNA component can be ‘harvested’ from its native context and used in a synthetic genetic context where it will exhibit the desired activity. In one example, researchers isolated a ligand-binding RNA sequence to thiamine pyrophosphate (TPP) from a natural riboswitch and linked this natural RNA ligand-binding component to a hammerhead ribozyme to construct a synthetic RNA-based sensing-actuation element that responded to TPP in *E. coli* (29). The native function of a naturally-occurring RNA element can also be altered to generate new functions or ‘refined’ to make the element compatible with modular integration into broader genetic device or system platforms. Combined rational and evolutionary strategies have been applied to alter the native function of natural RNA elements. In one example, the native activity of an RNase III hairpin substrate was altered in yeast by modifying key sequences that are essential to protein-binding and cleavage activities within the regulatory element (30,31). The resulting hairpin libraries were inserted in the 3’ UTR of the target transcript, and an *in vivo* screen was performed to identify a set of new hairpin sequences exhibiting a wide-range of regulatory activities. Similar approaches have been applied to alter the native activities of diverse RNA regulatory elements, including ribozymes (32), RNase cleavage sites (33), stabilizing elements (34), RBS sequences (35), and riboswitches (36).

1.3.2 Computational tools for the design of RNA components

Computational tools have been developed to aid researchers in designing functional RNA components. For example, a computational method for designing synthetic RBS sequences was recently described based on predicting the energies of secondary structures around and including the RBS and the energies associated with the RBS:rRNA interaction (37). A two state thermodynamic model was devised, the initial state being the free 30S ribosomal subunit and folded mRNA transcript and the final state being the 30S complex bound to the transcript. The difference in free energy between these two states was used to predict a translation initiation rate for a given mRNA sequence. The computational method was applied to generate RBS sequences in *E. coli* that resulted in fluorescent reporter levels that spanned several orders of magnitude. This tool was then applied to design and optimize the expression of an input to a genetic AND gate. As another example, design tools for microRNA and shRNA elements have been developed by companies such as Life Technologies and Thermo Scientific. Given the sequence of a target gene of interest, these RNAi design tools can predict targeting sequences that will allow for efficient RNAi-mediated knockdown of that gene. These programs will design a shRNA or miRNA that incorporate the predicted targeting sequences, including the structural elements necessary for correct biogenesis and silencing efficiency, and primer sequences for construction and cloning of the regulatory element.

Forward engineering with current computational tools is not precise, and strategies for most effectively using these tools require some level of generating multiple functional RNA elements and then screening for those sequences that exhibit desired

activities. For example, the RBS design tool has a probability of 0.47 of achieving a target protein expression level within two-fold. Further development of design tools that more accurately predict the precise DNA sequence needed to achieve a quantitative functional activity for different RNA components will allow for the efficient design, construction, and implementation of functional components tailored for various networks and systems.

1.3.3 Evolving new RNA component functions

While many functional RNA components can be harvested from natural biological systems and then refined and enhanced through rational design and evolutionary methods, such strategies can be limited to generating functions that are closely-related to the native activities. To generate new RNA component functions *de novo*, researchers have turned to *in vitro* selection strategies. The ability to readily interconvert between genetic information (DNA) and functional molecule (RNA) through transcription and reverse transcription (RT) processes and the ability to readily amplify DNA through polymerase chain reaction (PCR) processes allows efficient sampling of large RNA sequence space *in vitro*. Thus, large RNA libraries can be searched for rare functional sequences, which can subsequently be recovered, amplified, and searched again.

RNA presents a unique advantage over protein as a design substrate for control devices, as new sensing functions can be generated *de novo* through an *in vitro* selection strategy, systematic evolution of ligands by exponential enrichment (SELEX) (38,39). SELEX typically starts with an initial RNA library of $\sim 10^{14}$ - 10^{15} molecules, each

composed of a randomized region spanning ~30-70 nts flanked by constant sequences. Ligand-binding sequences are isolated by partitioning the RNA library through any of a number of different strategies, although affinity chromatography-based methods are most commonly used. The recovered sequences are subsequently amplified through RT-PCR and used as the starting library for the next round of selection. Iterative rounds of selection are performed, and the selection stringency and counterselections can be tailored to enrich for RNA sequences with high affinities and specificities to the target ligand. As one example, by incorporating counterselections against caffeine in the later selection rounds, an aptamer was selected to theophylline that exhibits a 10,000-fold lower affinity for caffeine, which differs from theophylline by a single methyl group (40). *In vitro* selection strategies have also been applied to the generation of novel ribozymes capable of RNA ligation (41).

When utilizing RNA components generated through *in vitro* selection strategies for cellular applications, the activities of the *in vitro* optimized components may not translate directly to the complex cellular environment. Cell-based selection and screening strategies have been used to perform a secondary screen on *in vitro* enriched RNA component libraries by the component function to a measurable gene expression output. In one example, an *in vitro* enriched aptamer pool to a small molecule ligand, atrazine, was screened in *E. coli* through a cell-based motility assay (2). The RNA aptamer library was coupled to an RBS through a randomized linker region, where the resulting device library was linked to a gene that controls cell motility, such that functional aptamer sequences could be recovered from cells exhibiting greatest mobility in the presence of atrazine on solid medium. A similar approach has also been taken by directly inserting an

in vitro enriched aptamer pool to neomycin in the 5' UTR of a GFP reporter gene in *E. coli* and screening for *in vivo* functional neomycin aptamers (42). The combined *in vitro* and *in vivo* approaches provide a powerful strategy for generating and tailoring new component functions for the cellular environment.

1.4 Design strategies for small molecule-responsive synthetic RNA devices

Construction of RNA control devices generally starts by functionally linking sensor and actuator components to support transmission of information detected by the sensor into regulated activity of the actuator. Two strategies are generally taken in the design of RNA control devices: (i) direct linkage of sensor and actuator components; (ii) linkage of sensor and actuator components through a distinct information transmission component. In the first strategy, the regulatory effect is imparted by the resulting conformational change in the sensor component in the presence of the input signal, which directly affects the activity of the actuator component. In the second strategy, the transmitter component guides secondary structure changes in the sensor and actuator components, which direct these components between active and inactive conformations. Both modular and non-modular design strategies have been adopted in the construction of RNA controllers. Modular device design strategies introduce standardized communication interfaces between the sensor and actuator components to insulate the specific sequences of these components from one another, resulting in design platforms that support the interchange of functional components (i.e., sensor, actuator) without significant device redesign. Utilizing these design approaches researchers have constructed small molecule-responsive RNA control devices by coupling RNA aptamers

(sensors) to various gene-regulatory (actuator) components, whose mechanism of action is generally specific to the gene expression machinery associated with the host cell.

1.4.1 Prokaryotes

Small molecule-responsive control devices have been engineered in bacteria by modifying or adding component functions in natural riboswitch or regulatory elements. In one example, researchers modified the information transmission function encoded in a natural TPP-responsive riboswitch, which exhibits TPP-dependent gene repression, by randomizing a sequence within the riboswitch and screening the resulting library for new synthetic devices that exhibit TPP-dependent gene activation (43). In another example, a theophylline aptamer was coupled to a natural riboswitch through a randomized linker sequence and the resulting library was screened for devices that exhibit logic operations (AND, NAND) by modulating ribosome access to the RBS (44). RNA control devices have also been built by direct integration of RNA sensors into natural regulatory elements. A theophylline aptamer was integrated into a stem in the group I self-splicing intron from the bacteriophage T4 thymidylate synthase gene to construct a theophylline-dependent RNA splicing device that regulated *E. coli* growth (45).

The majority of small molecule-responsive RNA control devices in bacteria have been designed to control translation initiation, due to the relative ease with which ribosome loading can be modulated through structural rearrangement of the RBS. For example, RBS-based devices were built in *E. coli* by coupling a theophylline aptamer to the RBS through a linker sequence capable of structural rearrangement through strand-displacement (46) or helix-slipping (47) mechanisms. The accessibility of the RBS

(related to its single-stranded state) is altered by ligand-dependent changes in base pairing interactions between the sensor and linker (strand-displacement) or local nucleotide shifts within the linker (helix-slipping). In the device design strategy based on strand displacement, part of the aptamer sequence directly interacts with the RBS sequence, such that independent modification of either component may disrupt these interactions and thus require redesign of the linker sequence to maintain device function. Therefore, cell-based screening strategies have been developed to support the identification of new linker sequences through colorimetric (48), flow cytometry (49), and motility assays (50). Other examples of RBS-based devices were designed by coupling a theophylline aptamer to an indirect actuator, a hammerhead ribozyme, which in turn was coupled to a direct actuator, a RBS (51,52). In this device design, the RBS is sequestered within the ribozyme stem, such that ligand-induced ribozyme cleavage results in unwinding of the ribozyme stem, thereby increasing ribosomal access to the RBS and thus gene expression levels.

1.4.2 Yeast

Different design approaches have been taken to engineer small molecule-responsive RNA control devices in eukaryotes, as fewer examples of natural riboswitch elements are available to build from and more sophisticated gene-regulatory mechanisms are available than in prokaryotes. RNA control devices have been built by integrating the RNA sensor directly into part of the actuator component, such that the conformational change associated with ligand binding to the aptamer affects the activity of the actuator. In one example, a tetramethylrosamine (TMR) aptamer and a synthetic RNA

transcriptional activator were joined through the stem of the transcription activator, where part of the stem was randomized to allow screening for TMR-responsive gene-regulatory activity in *Saccharomyces cerevisiae* (53). In another example, a theophylline aptamer was integrated into the core cleavage region of a synthetic RNase III hairpin, which when placed in the 3' UTR of a target transcript, directed cleavage and subsequent inactivation of that transcript in yeast (31). However, the binding of theophylline to the integrated aptamer restricts cleavage activity of the RNase III enzyme, thereby activating gene expression. Other small molecule-responsive RNA control devices have been built by integrating an aptamer in proximity to a 5' splice site (54) or a branch point (55). As an example, a tetracycline aptamer in close proximity to a 5' splice site, where the consensus sequence of the splice site is integrated into the aptamer stem (54). The binding of tetracycline to the aptamer results in a conformation that potentially prevents access to the splice site, thereby achieving ligand-dependent regulation of splicing.

As a second design approach, composition frameworks have been developed to support the modular assembly of RNA control devices from underlying functional components in yeast (56). In the proposed framework, ribozyme-based devices were constructed by linking an aptamer to a hammerhead ribozyme through a distinct transmitter sequence designed to both encode the information processing function of the device and insulate the sensor and actuator components. This design strategy was demonstrated to support independent swapping of the sensor and transmitter components and thus tailoring of the encoded sensing and information processing functions without significant device redesign. Rational design strategies have been developed that utilize RNA structure prediction programs to guide sequence changes in the transmitter

component to alter the energetic values associated with alternate device conformations, thereby changing the energetic partitioning between functional states of the device and the resulting device regulatory activities (56,57). The composition framework also supported extension to the modular assembly of multiple sensor, transmitter, and actuator domains to construct devices that exhibit higher-order information processing functions, including logic operations, signal and bandpass filters, and programmed cooperativity (57). The utilization of ribozymes as actuators in this RNA device platform also allows for transportability of the device across different organisms, because the ribozyme activity is independent of cell-specific gene-regulatory machinery (3).

1.4.3 Mammalian cells

RNA control devices in mammalian cells have been designed through numerous strategies ranging from directly adapting regulatory mechanisms that were developed in simpler organisms to taking advantage of the rich complexity of RNA processing pathways unique to higher organisms. Several RNA control devices have been developed through design strategies that directly integrate an RNA aptamer to a gene-regulatory element. These approaches have been applied to the design of small molecule-responsive RNA control devices that conditionally silence target genes through RNAi-mediated silencing mechanisms (58,59). These devices have been designed by integrating an RNA aptamer directly within the basal segments of a miRNA or in the loop region of a shRNA, such that binding of the input molecule prevents correct biogenesis of the RNAi substrate and thus results in increased target gene expression levels. Conditional RNAi silencing in response to a wide range of ligands, including theophylline, tetracycline, and xanthine,

have been demonstrated with these RNAi-based devices. While direct integration strategies generally limit the tuning of device regulatory activity that one can achieve through molecular alteration strategies, researchers demonstrated a tuning strategy for the miRNA-based device based on rational design of device clusters (59).

Other mammalian RNA control devices have been developed through design strategies that couple the sensor and actuator components through a distinct transmitter component. In one example, a transmitter and RNA aptamer were integrated into the loop region of a shRNA (60). In the active ligand-unbound conformation, the aptamer was incorrectly formed and the RNAi machinery correctly processed the device, resulting in silencing of the target gene. The inactive conformation of the device coincided with correct formation of the aptamer, such that binding of the small molecule input to the aptamer stabilized this conformation. The RNAi machinery did not correctly process this inactive conformation, thus resulting in increased target protein levels in the presence of input. As another example, the previously-described ribozyme-based devices, originally demonstrated in yeast, were shown to retain small molecule-responsive gene-regulatory activity in mammalian cell culture and *in vivo* (3). RNA devices that incorporate a distinct transmitter component can generally be tuned by altering the sequence of the transmitter component to optimize the thermodynamics of the competing hybridization reactions underlying the strand-displacement mechanism. For both the shRNA- and ribozyme-based devices, sequence alterations within the transmitter components were used to optimize the regulatory performance of the resulting devices in mammalian cells (3,60).

1.5 Broad applicability of the small molecule-responsive synthetic RNA devices

Synthetic RNA control devices hold much promise for leading to transformative advances in how we interact with and program biology, providing access to otherwise inaccessible information on cellular state, and allowing sophisticated exogenous and embedded control over cellular functions. Small molecule-responsive synthetic RNA devices have been integrated into engineered biological systems for applications spanning biosynthesis, bioremediation, to health and medicine (Figure 1.2).

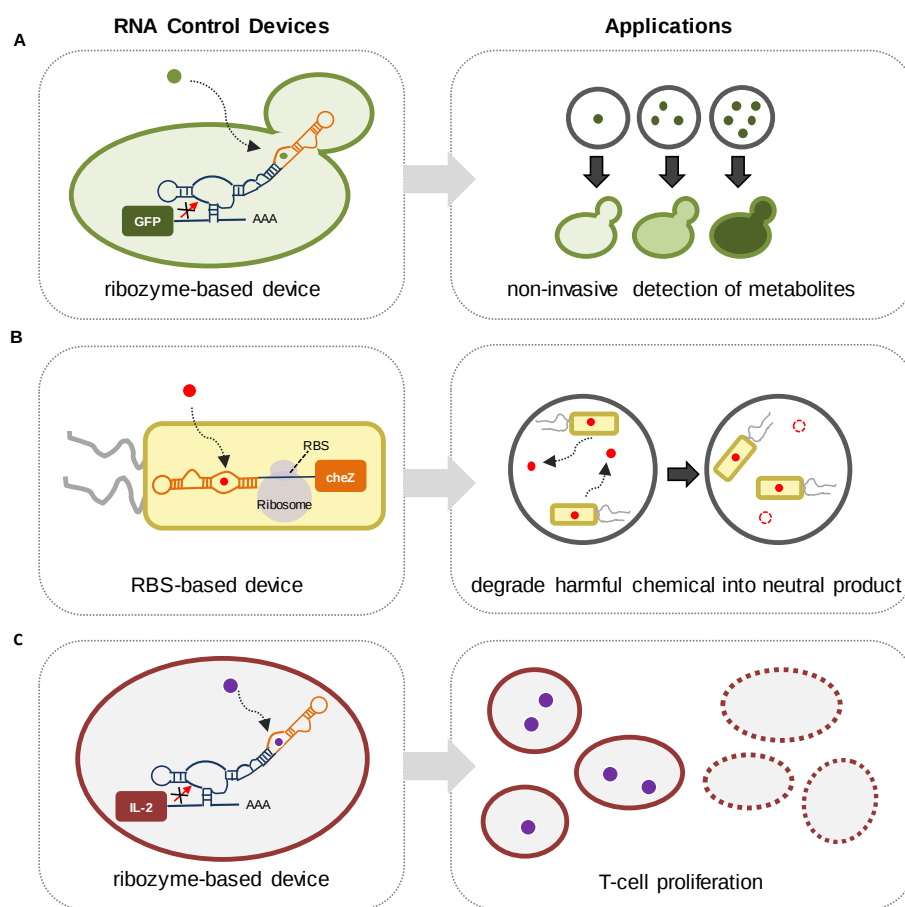


Figure 1.2 RNA-based controllers have been integrated into engineered biological systems for applications spanning biosynthesis, bioremediation, to health and medicine. (A) A metabolite-responsive ribozyme-based device linked to a fluorescent reporter

output was demonstrated in yeast as a noninvasive sensor of metabolite concentration. (B) A pollutant-responsive RBS-based device linked to a motility gene output was demonstrated in bacteria to program the cells to move along a gradient of the pollutant. (C) A small molecule-responsive ribozyme-based device linked to a proliferative cytokine gene output was demonstrated in T cells to control T-cell survival.

1.5.1 Implementation in genetic networks directed to biomanufacturing/biosynthesis applications

Small molecule-responsive synthetic RNA devices have important applications in biosynthesis processes, where they can be implemented as noninvasive sensors of metabolite accumulation and controllers for optimizing flux and product yield. In one example, an RNA control device was utilized for noninvasive detection of metabolite accumulation in yeast (56) (Figure 1.2A). Cells were engineered to express a construct harboring a xanthine-responsive ribozyme-based device regulating a GFP reporter gene. Xanthosine was fed to the yeast cells, which converted this fed substrate to xanthine through an endogenous enzyme activity. The conversion of xanthosine to xanthine was monitored indirectly as an increase in GFP levels, where increases in fluorescence correlated with increased product accumulation as measured by LC-MS.

1.5.2 Implementation in genetic networks directed to environmental/agricultural applications

Small molecule-responsive synthetic RNA devices also have important applications in agricultural biotechnology and environmental remediation. In the latter case, engineered organisms that may be released into the environment will need precise control over designed functions and safety mechanisms in place to prevent uncontrolled release. In one example, an RBS-based RNA control device was implemented to detect a

toxic environmental pollutant, atrazine (2) (Figure 1.2B). The atrazine-responsive device was coupled to the *cheZ* gene to control the motility of an *E. coli* strain engineered to express an atrazine-catabolizing enzyme activity. The presence of atrazine activates the expression the *cheZ* gene, which allows cells to move along the source of the pollutant and convert atrazine into less harmful product, hydroxyatrazine.

1.5.3 Implementation in genetic networks directed to health and medicine application

Finally, small molecule-responsive synthetic RNA devices can be applied to improve the safety or efficacy of cell-based therapeutics. Recent progress has been made in the application of RNA devices to adoptive T-cell therapy, an approach currently in clinical trials, in which a patient's own T cells are harvested and trained against tumor-specific antigens or engineered to express chimeric antigen receptors before being injected back into the body. A major challenge associated with this strategy is ensuring that the T cells can survive and proliferate sufficiently when engrafted into the host to eradicate the diseased cells. To overcome this problem, small molecule-responsive RNA devices were applied to control T-cell proliferation in response to the application of drug molecules (3) (Figure 1.2C). Researchers placed ribozyme-based devices in the 3' UTR of transcripts encoding the proliferative cytokines, IL-15 and IL-2, such that in the absence of the small molecule input the transcript was destabilized through ribozyme cleavage, resulting in reduced T-cell survival. In the presence of the drug input, IL-2 or IL-15 was produced, resulting in increased T-cell proliferation and survival. As a proof of

concept, this platform was shown to be modular in that aptamers to different small molecules (theophylline, tetracycline) were used to control the expression of different proliferative cytokines and hence proliferation of the T cells. The regulatory performance of the RNA-based system was tailored rationally by integrating multiple devices in series to improve the regulatory stringency. Currently, RNA sensors to clinically relevant molecules are needed to advance this strategy forward for therapeutic applications.

1.6 Challenges for the scalable construction and programming of RNA devices

The broad implementation of synthetic RNA control devices in genetic networks will require enabling technologies that support scalable generation of new functional RNA components for the construction of devices and quantitative tailoring of device regulatory performance. Currently, the scalable construction of RNA devices is limited by a small number of existing functional RNA components, in particular components that exhibit sensing functions to desired molecular inputs for cellular engineering applications. While advances have been made in the generation of protein aptamers by utilizing high-throughput, high-efficiency, solution-based partitioning methods, such as capillary electrophoresis (61), the generation of small molecule aptamers still relies on low-throughput, inefficient, matrix-based partitioning methods (38,39), thus limiting our ability to generate new aptamers to diverse molecular ligands.

Another challenge in the field is the reliable construction of devices from components to achieve desired regulatory performance for specific cellular engineering applications. The *in vitro*-generated components need to be linked to other functional component in order to construct synthetic control devices. The regulatory performance of

the resultant devices depends both on the component linkage strategy and the activities of individual components within the host cell. Therefore, the component activities may require further optimization in the context of the device and cellular environment to achieve desired gene-regulatory activities for the targeted applications. Cell-based selection and screening strategies have been developed by linking the gene-regulatory activities of the control devices to a single measurable gene expression output (48-50,62). Although these strategies have been demonstrated for tailoring device regulatory performance, the efficiency of these single-output methods can be negatively impacted by the stochastic nature of gene expression, thereby making these methods less quantitative and efficient.

1.7 Interrelationship among the thesis projects

To allow scalable construction and programming of RNA control devices, this thesis focuses on developing a scalable framework for generating new RNA component functions and quantitative tailoring of the resultant device functions for broad applications within cellular environments. Chapter 1 provides an overview of strategies in generating and optimizing RNA functional components and diverse approaches to the construction of small molecule-responsive synthetic RNA control devices.

We focus primarily on the construction and programming of ribozyme-based devices, as devices utilizing ribozymes as actuator component have the advantage of being portable across different environments (i.e., *in vitro*, *in vivo*). Chapter 2 describes the development of a high-throughput and quantitative two-color FACS-based screen for efficient tailoring of RNA device activity. This screening strategy enables direct

generation and optimization of RNA component functions within the device platform in the cellular environment, which aids in the broader integration of these devices within biological networks. Chapter 3 describes work on an efficient cis-blocking strategy for *in vitro* ribozyme characterization, as our work demonstrated that the *in vitro* cleavage rates associated with these ribozyme-based devices are closely correlated to the corresponding gene-regulatory activities. Chapter 4 describes progress towards the development of a high-throughput, solution-based *in vitro* aptamer selection strategy based on the ribozyme device platform. The successful development of this method will enable the generation of new RNA aptamers to diverse small molecule ligands without the need of conjugation to a solid matrix. Ultimately, this method can be coupled to the two-color FACS-based screen, such that an *in vitro* enriched library can be subjected to a secondary screen for *in vivo* functional activity. Chapter 5 describes the development of a label-free and quantitative method supporting rapid characterization of RNA aptamers to small molecules. Chapter 6 describes the immediate efforts that need to be directed to integrating these enabling technologies into a scalable framework for the engineering of new synthetic devices. Collectively, this thesis describes powerful strategies for tailoring RNA control devices to respond to application-specific small molecule inputs with tailored regulatory properties, thereby enabling their effective implementation within biological networks.

References

1. Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J. *et al.* (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, **440**, 940-943.
2. Sinha, J., Reyes, S.J. and Gallivan, J.P. (2010) Reprogramming bacteria to seek and destroy an herbicide. *Nat Chem Biol*, **6**, 464-470.
3. Chen, Y.Y., Jensen, M.C. and Smolke, C.D. (2010) Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc Natl Acad Sci U S A*, **107**, 8531-8536.
4. Culler, S.J., Hoff, K.G. and Smolke, C.D. (2010) Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science*, **330**, 1251-1255.
5. Endy, D. (2005) Foundations for engineering biology. *Nature*, **438**, 449-453.
6. Smolke, C.D. and Silver, P.A. (2011) Informing biological design by integration of systems and synthetic biology. *Cell*, **144**, 855-859.
7. Carr, P.A. and Church, G.M. (2009) Genome engineering. *Nat Biotechnol*, **27**, 1151-1162.
8. Purnick, P.E. and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol*, **10**, 410-422.
9. Kozak, M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, **361**, 13-37.

10. Yanofsky, C. (1981) Attenuation in the control of expression of bacterial operons. *Nature*, **289**, 751-758.
11. Alifano, P., Bruni, C.B. and Carlomagno, M.S. (1994) Control of mRNA processing and decay in prokaryotes. *Genetica*, **94**, 157-172.
12. Ehretsmann, C.P., Carpousis, A.J. and Krisch, H.M. (1992) mRNA degradation in prokaryotes. *FASEB J*, **6**, 3186-3192.
13. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell*, **31**, 147-157.
14. Serganov, A. and Patel, D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet*, **8**, 776-790.
15. Mandal, M. and Breaker, R.R. (2004) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol*, **5**, 451-463.
16. Gottesman, S. (2004) The small RNA regulators of *Escherichia coli*: roles and mechanisms*. *Annu Rev Microbiol*, **58**, 303-328.
17. Schluter, J.P., Reinkensmeier, J., Daschkey, S., Evguenieva-Hackenberg, E., Janssen, S., Janicke, S., Becker, J.D., Giegerich, R. and Becker, A. (2010) A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. *BMC Genomics*, **11**, 245.
18. Carthew, R.W. and Sontheimer, E.J. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell*, **136**, 642-655.
19. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215-233.

20. Nicoloso, M.S., Spizzo, R., Shimizu, M., Rossi, S. and Calin, G.A. (2009) MicroRNAs--the micro steering wheel of tumour metastases. *Nat Rev Cancer*, **9**, 293-302.
21. Mattick, J.S., Amaral, P.P., Dinger, M.E., Mercer, T.R. and Mehler, M.F. (2009) RNA regulation of epigenetic processes. *Bioessays*, **31**, 51-59.
22. Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet*, **10**, 155-159.
23. Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res*, **31**, 3429-3431.
24. Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M. and Pierce, N.A. (2011) NUPACK: Analysis and design of nucleic acid systems. *Journal of computational chemistry*, **32**, 170-173.
25. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A*, **101**, 7287-7292.
26. Das, R. and Baker, D. (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci U S A*, **104**, 14664-14669.
27. Parisien, M. and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51-55.
28. Flamm, C., Fontana, W., Hofacker, I.L. and Schuster, P. (2000) RNA folding at elementary step resolution. *RNA*, **6**, 325-338.

29. Wieland, M., Benz, A., Klauser, B. and Hartig, J.S. (2009) Artificial ribozyme switches containing natural riboswitch aptamer domains. *Angew Chem Int Ed Engl*, **48**, 2715-2718.
30. Babiskin, A.H. and Smolke, C.D. (2011) A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. *Mol Syst Biol*, **7**, 471.
31. Babiskin, A.H. and Smolke, C.D. (2011) Engineering ligand-responsive RNA controllers in yeast through the assembly of RNase III tuning modules. *Nucleic Acids Res*, **39**, 5299-5311.
32. Chen, X., Denison, L., Levy, M. and Ellington, A.D. (2009) Direct selection for ribozyme cleavage activity in cells. *RNA*, **15**, 2035-2045.
33. Pflieger, B.F., Pitera, D.J., Smolke, C.D. and Keasling, J.D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol*, **24**, 1027-1032.
34. Smolke, C.D., Carrier, T.A. and Keasling, J.D. (2000) Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. *Appl Environ Microbiol*, **66**, 5399-5405.
35. Anderson, J.C., Clarke, E.J., Arkin, A.P. and Voigt, C.A. (2006) Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol*, **355**, 619-627.
36. Dixon, N., Duncan, J.N., Geerlings, T., Dunstan, M.S., McCarthy, J.E., Leys, D. and Micklefield, J. (2010) Reengineering orthogonally selective riboswitches. *Proc Natl Acad Sci U S A*, **107**, 2830-2835.

37. Salis, H.M., Mirsky, E.A. and Voigt, C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol*, **27**, 946-950.
38. Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818-822.
39. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.
40. Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425-1429.
41. Bartel, D.P. and Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science*, **261**, 1411-1418.
42. Weigand, J.E., Sanchez, M., Gunnesch, E.B., Zeiher, S., Schroeder, R. and Suess, B. (2008) Screening for engineered neomycin riboswitches that control translation initiation. *RNA*, **14**, 89-97.
43. Nomura, Y. and Yokobayashi, Y. (2007) Reengineering a natural riboswitch by dual genetic selection. *J Am Chem Soc*, **129**, 13814-13815.
44. Sharma, V., Nomura, Y. and Yokobayashi, Y. (2008) Engineering complex riboswitch regulation by dual genetic selection. *J Am Chem Soc*, **130**, 16310-16315.
45. Thompson, K.M., Syrett, H.A., Knudsen, S.M. and Ellington, A.D. (2002) Group I aptazymes as genetic regulatory switches. *BMC Biotechnol*, **2**, 21.

46. Desai, S.K. and Gallivan, J.P. (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J Am Chem Soc*, **126**, 13247-13254.
47. Suess, B., Fink, B., Berens, C., Stentz, R. and Hillen, W. (2004) A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res*, **32**, 1610-1614.
48. Lynch, S.A., Desai, S.K., Sajja, H.K. and Gallivan, J.P. (2007) A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chemistry & biology*, **14**, 173-184.
49. Lynch, S.A. and Gallivan, J.P. (2009) A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res*, **37**, 184-192.
50. Topp, S. and Gallivan, J.P. (2008) Random walks to synthetic riboswitches--a high-throughput selection based on cell motility. *Chembiochem : a European journal of chemical biology*, **9**, 210-213.
51. Wieland, M. and Hartig, J.S. (2008) Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angew Chem Int Ed Engl*, **47**, 2604-2607.
52. Ogawa, A. and Maeda, M. (2008) An artificial aptazyme-based riboswitch and its cascading system in E. coli. *Chembiochem : a European journal of chemical biology*, **9**, 206-209.
53. Buskirk, A.R., Landrigan, A. and Liu, D.R. (2004) Engineering a ligand-dependent RNA transcriptional activator. *Chemistry & biology*, **11**, 1157-1163.

54. Weigand, J.E. and Suess, B. (2007) Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. *Nucleic Acids Res*, **35**, 4179-4185.
55. Kim, D.S., Gusti, V., Dery, K.J. and Gaur, R.K. (2008) Ligand-induced sequestering of branchpoint sequence allows conditional control of splicing. *BMC Mol Biol*, **9**, 23.
56. Win, M.N. and Smolke, C.D. (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A*, **104**, 14283-14288.
57. Win, M.N. and Smolke, C.D. (2008) Higher-order cellular information processing with synthetic RNA devices. *Science*, **322**, 456-460.
58. An, C.I., Trinh, V.B. and Yokobayashi, Y. (2006) Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. *RNA*, **12**, 710-716.
59. Beisel, C.L., Chen, Y.Y., Culler, S.J., Hoff, K.G. and Smolke, C.D. (2011) Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. *Nucleic Acids Res*, **39**, 2981-2994.
60. Beisel, C.L., Bayer, T.S., Hoff, K.G. and Smolke, C.D. (2008) Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol Syst Biol*, **4**, 224.
61. Berezovski, M.V., Musheev, M.U., Drabovich, A.P., Jitkova, J.V. and Krylov, S.N. (2006) Non-SELEX: selection of aptamers without intermediate amplification of candidate oligonucleotides. *Nat Protoc*, **1**, 1359-1369.

62. Fowler, C.C., Brown, E.D. and Li, Y. (2008) A FACS-based approach to engineering artificial riboswitches. *Chembiochem : a European journal of chemical biology*, **9**, 1906-1911.