Chapter 1

Introduction

RNA is a versatile regulatory biomolecule

Synthetic biology is a rapidly emerging field that promises to improve our ability to investigate and manipulate living organisms through the creation of novel biological tools and systems, with innovations supporting applications in health, energy, and biomanufacturing[1–3]. While advances in DNA synthesis have enabled the construction of large genetic systems[4], the capability to design and predictably regulate such systems lags behind. Synthetic RNA-based gene-regulatory devices are uniquely poised to address this need.

Once thought to be merely the intermediate between the genetic information stored in DNA and proteins that executed cellular function, RNA has been shown to perform a large diversity of functional activities, such as catalysis, metabolite binding, and gene regulation[5–8]. In addition, functional RNA molecules have been described that can modulate their activity in response to cellular and environmental inputs. For example, temperature-sensitive structural elements regulate gene expression in the heat and cold shock responses in bacteria[9], and metabolite-binding elements control the expression of enzymes in biosynthetic pathways[10–12]. To date most of these regulatory elements have been characterized in prokaryotes, but examples have been found in eukaryotes as well[13]. The many examples of naturally-occurring, ligand-responsive RNA-based gene-regulatory elements, or RNA switches, serve as the raw materials and inspiration for novel synthetic RNA-based regulatory devices[14].

As with proteins, the ability of RNA to perform functional activities arises from its three-dimensional folded structure. Unlike proteins, however, this structure is almost entirely determined by hydrogen-bonding, base-stacking, and electrostatic interactions

between the constituent monomers[15]. The relative simplicity of RNA intramolecular interactions has enabled the design of software models that computationally predict the secondary structures and associated free energies of a given RNA sequence with a high degree of accuracy[16–18]. Such software has greatly aided the design of engineered functional RNA molecules[19,20]. Facile protein structure prediction is not yet feasible due to the complexity of protein folding, and therefore protein-based devices such as allosteric transcription factors are currently far more challenging to engineer than their RNA-based counterparts.

Engineered RNA devices in eukaryotes enable dynamic modulation of gene expression in response to molecular and environmental signals

Synthetic RNA switches achieve gene regulation through a variety of mechanisms, but they generally contain two core components. The sensor component detects the input signal, such as a small molecule or protein, through a binding interaction, and the actuator component modulates gene expression through mechanisms such as transcription, post-transcriptional processing, translation, or messenger RNA (mRNA) stability. Many RNA-based devices utilize architectures that also incorporate a transmitter component, which links the sensor and actuator components and transmits information between them by modulating the activity of the actuator based on the ligand bound state of the sensor. The sensor component is typically an aptamer, an RNA sequence with high affinity and specificity for a small molecule or protein ligand. Many such binding elements can be found in nature[12,21], but new aptamers can be generated

with an *in vitro* selection method known as systematic evolution of ligands by exponential enrichment, or SELEX[22,23]. This method can be used to generate aptamer sequences to theoretically any small molecule or protein ligand of interest.

The earliest potential point of regulation of gene expression is transcription. In one example, an RNA regulator of transcription that responded to the small molecule tetramethylrosamine (TMR) was demonstrated in *Saccharomyces cerevisiae*. The TMR aptamer was linked to a transcriptional activator through a randomized transmitter component and functional devices were selected based on TMR responsiveness. Demonstrations of engineered ligand-responsive RNA-based regulators of transcription have not been reported to date in mammalian cells.

RNA-based devices that modulate gene expression through post-transcriptional processing, such as splicing, have been demonstrated in yeast and human cells[24,25]. Proper assembly of the spliceosome requires recognition of specific exonic and intronic sequence elements, and researchers have shown that the accessibility of these elements can be regulated by ligand binding to aptamer sequences. In one example, an aptamer for tetracycline was placed at the 5' splice site in the precursor mRNA (pre-mRNA) of a fluorescent reporter gene in yeast[20]. Binding of tetracycline altered the conformation of the region around the splice site, preventing splicing of the exons encoding the reporter gene and reducing expression by up to 32-fold. In another example, protein-responsive RNA-based devices were used to control alternative splicing of different transgenes in human embryonic kidney 293 (HEK293) cells[25]. Aptamers for three different proteins were placed in an intronic region such that protein binding to the aptamer sequences

prevented the exclusion of an exon containing a premature stop codon, thereby modulating the expression of the encoded transgene.

RNA interference (RNAi) is another post-transcriptional processing mechanism that has been utilized in RNA-based devices for controlling target gene expression. RNAi is a powerful platform for gene regulation in higher eukaryotes that is based on complementarity between the RNA regulator and the target gene, where the regulators can be encoded in diverse forms including microRNAs (miRNAs), short hairpin RNAs (shRNAs), and small interfering RNAs (siRNAs)[26]. These RNAi-based components must be processed by cellular protein machinery to silence gene expression, either through blocking translation initiation, interrupting polypeptide elongation, or degrading the transcript[14]. Many RNA switches that modulate processing in response to ligand input have been demonstrated[27,28]. In one example, a miRNA-based switch responsive to small molecules was demonstrated in HEK293 cells[29]. Ligand binding to an aptamer integrated into the base of the miRNA stem prevented processing of the primary miRNA (pri-miRNA) by Drosha, thereby increasing target gene expression levels as a function of increasing ligand concentrations. In another example, the aptamer for the archaeal ribosomal protein L7Ae was inserted in the loop region of an shRNA targeting an antiapoptotic gene[27]. By simultaneously regulating a proapoptotic gene with a separate device, the authors were able to control apoptosis in HeLa cells.

Regulation of translation initiation is a common mechanism employed by ligand-responsive RNA switches. Following the example of natural prokaryotic translation initiation riboswitches and their engineered counterparts[10–12], the aptamer is placed in the 5’ untranslated region (UTR) just upstream of the translation initiation codon, such that

ligand binding prevents the ribosome from binding and assembling properly. For switches responsive to small molecules[30], ligand binding can stabilize structures that discourage ribosome assembly, while in other cases protein binding prevents ribosome association through steric hindrance[31–33]. In one interesting study in HEK293T cells, protein binding to an aptamer in the 5' UTR of a bicistronic mRNA selectively repressed translation of the upstream gene while not affecting internal ribosome entry sequence (IRES)-dependent translation of the downstream gene[32].

Finally, effective regulation of gene expression can be accomplished by controlling the stability of mRNA, usually by modulating the susceptibility of mRNA to cellular ribonucleases (RNases). The ends of eukaryotic mRNAs are protected by the 5' 7-methyl-guanosine cap and the 3' poly(A) tail, which themselves are bound by various proteins that circularize the transcript. Directed cleavage in either of the UTRs or the coding region exposes the mRNA to rapid degradation by exoribonucleases. In one engineered switch exploiting this phenomenon, an aptamer that binds the caffeine analogue theophylline was integrated into a hairpin recognized by the RNase Rnt1p, such that ligand binding prevented Rnt1p-mediated cleavage in yeast[34]. Another type of device controlling mRNA stability is based on self-cleaving ribozymes, which will be described below.

Ligand-responsive ribozyme switches

Ribozymes are RNA enzymes that accelerate chemical reactions by adopting certain folded structures similar to peptide-based enzymes. Natural ribozymes were first

discovered in Group I introns[35], but have since been identified to be involved in many vital cellular processes from mRNA splicing[36] to peptide synthesis[37]. Many ribozymes catalyze the lysis of an RNA phosphodiester bond, either in its own strand (cis) or in a separate RNA molecule (trans), thereby cleaving it in two. Hammerhead ribozymes, first discovered in plant viroids[38] and shown to function in a variety of organisms[39], rapidly catalyze self-cleavage through a phosphodiester isomerization mechanism (Figure 1.1). The cleavage site is located in the ribozyme's catalytic core immediately downstream of the conserved NUX sequence, in which N is any nucleotide and X is either A, C, or U.

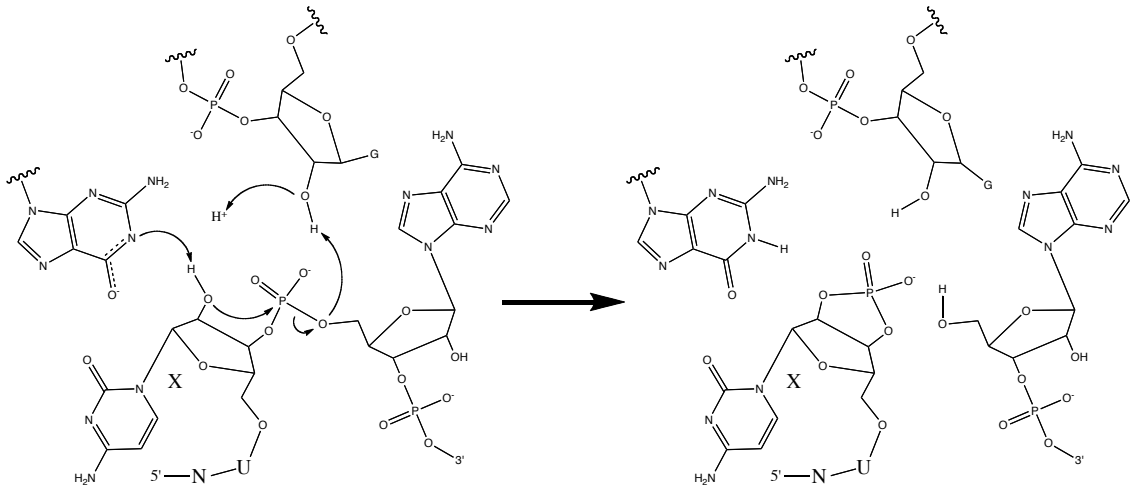

Figure 1.1. The phosphodiester isomerization mechanism of hammerhead ribozymes. Two nearby guanosines contribute to general base catalysis. In this example, 'X' is cytidine. Adapted from[40].

The Smolke laboratory has recently described a framework for constructing ribozyme-based gene-regulatory RNA devices[19]. The framework provides a modular assembly strategy for building these RNA devices from a sensor component, made of an

RNA aptamer, an actuator component, made of a satellite RNA of tobacco ringspot virus (sTRSV) ribozyme[41], and a transmitter component, made of a sequence that functionally couples the sensor and actuator components (Figure 1.2). The transmitter component is rationally designed based on competitive hybridization events that enable the device to distribute between two primary conformations: one in which the input cannot bind to the sensor and the other in which the input can bind to the sensor. Input binding shifts the distribution to favor the input-bound conformation as a function of increasing input concentration and is translated to a change in the activity of the actuator, where a 'ribozyme-active' state results in self-cleavage of the device. The RNA device is coupled to the 3' UTR of the target gene, where ribozyme self-cleavage inactivates the transcript and thereby lowers gene expression independent of cell-specific machinery.

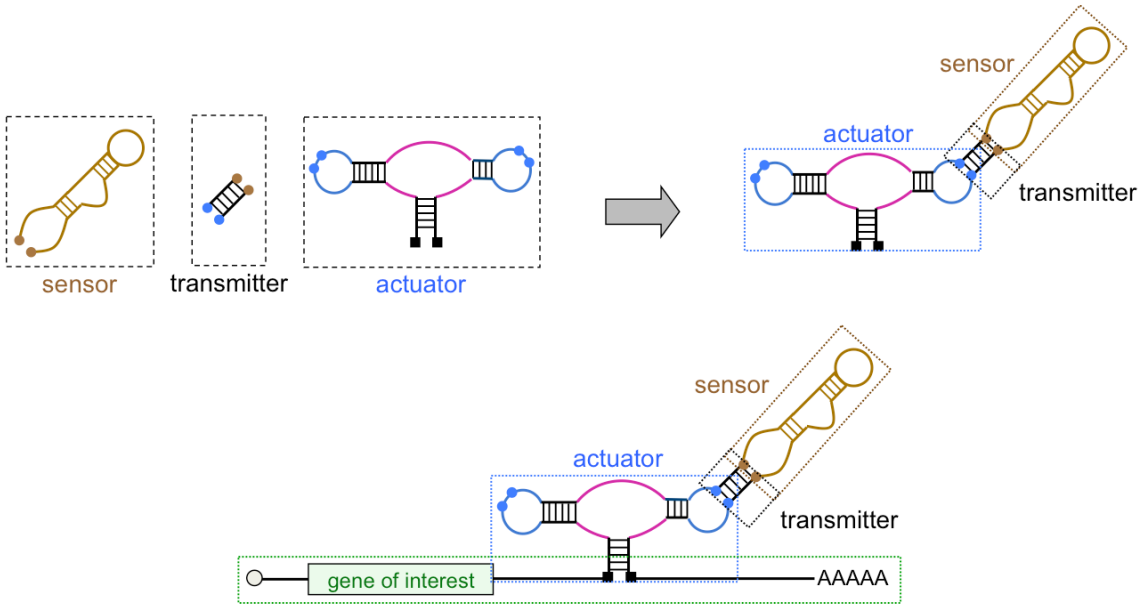

Figure 1.2. Assembly of a ribozyme switch from modular components. The aptamer is shown in light brown, the stems are shown in black, the catalytic core is shown in magenta, and loops and bulges are shown in blue. Adapted from[42].

The precise design of the transmitter component determines whether the ribozyme switch will repress or enhance gene expression, unlike many of the switches described above, which are capable of regulating gene expression in only one direction. RNA devices that function as either ON or OFF switches that convert a molecular input signal to increased or decreased gene expression output, respectively, have been demonstrated in yeast and mammalian cells[19,42–46] (Figure 1.3). After initial demonstration of ribozyme switches responsive to theophylline and tetracycline in yeast[19], the framework was extended to provide a general approach for the engineering of multi-input, higher-order information processing devices, where two-input logic gates (AND, NOR, NAND, and OR gates), signal filters, band-pass filters, and programmed cooperativity operations were demonstrated[42]. These ribozyme switches were also used to control T-cell proliferation in mice[43], demonstrating phenotypic control in an animal model. Other investigators have demonstrated switching activity of a theophylline-responsive ribozyme switch coupled to the 5' UTR[45,46], but this strategy can lead to nonspecific reduction of translation initiation due to the high degree of secondary structure upstream of the start codon.

Ribozyme switches possess a significant advantage not shared by many other gene regulation platforms in that their mechanism of action does not require any cell-specific machinery. Ribozyme switches are therefore functional across different organisms, including bacteria[47], yeast[19], and mammalian systems[43]. This allows rapid screening of devices generated by both rational and directed evolution design strategies in simple organisms[48], optimizing device activity before transitioning to more complex organisms[44].

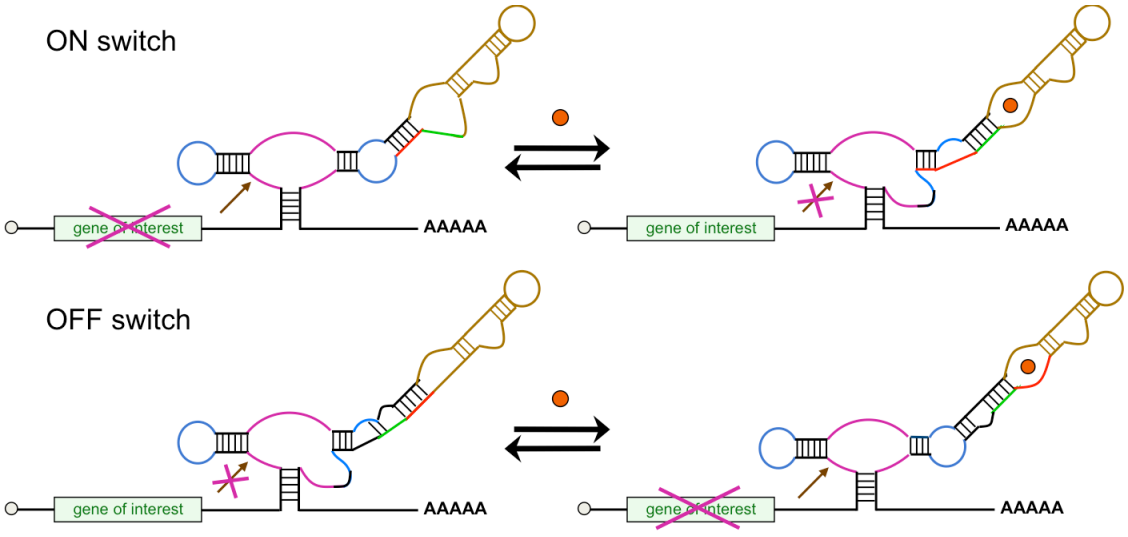

Figure 1.3. Ligand binding stabilizes the aptamer-formed conformation. In an ON switch, ligand (red disk) binding stabilizes the catalytically inactive conformation, preventing ribozyme self-cleavage and allowing translation of the gene of interest. In an OFF switch, the ligand stabilizes the catalytically active conformation, inducing cleavage and gene repression. The cleavage site is indicated with an arrow. Coloring is the same as in Figure 1.2.

However, ribozyme switches are somewhat limited in their effectiveness and range of capabilities. First, they are generally limited to the regulation of transgenes, with endogenous gene regulation achievable only through the utilization of targeted chromosomal integration strategies, which are cumbersome in mammalian systems[49,50]. In contrast, switch platforms based on RNAi enable facile ligand-responsive regulation of endogenous genes[27,28]. Second, to date only small-molecule-responsive ribozyme switches have been described, while other platforms have been shown to respond to protein ligands. Third, ribozyme switches have not yet been able to achieve the high dynamic ranges and input sensitivities of other gene regulation systems. Finally, the

mechanism of action, specifically the subcellular location where ribozyme cleavage occurs, has not been fully elucidated. For other switch platforms, such as those based on modulation of transcription, splicing, or RNAi processing, choice of ligand is constrained by the known subcellular location of the mechanism of action. It is desirable to elucidate similar details about ribozyme switches to determine which ligands the platform is capable of sensing.

Applications

Engineered RNA devices have been used for a variety of applications in eukaryotes. In reconstituting useful biosynthetic pathways in new host organisms, it is important to regulate the expression levels of the enzymes to maximize their activity while efficiently exploiting cellular resources. Ribozyme-based regulatory devices have been used as noninvasive sensors of enzymatic products. In one example in yeast, a ribozyme switch responsive to xanthine was used to control a fluorescent reporter gene[19]. When the yeast were fed xanthosine, the enzymatic conversion of xanthosine to xanthine was reported noninvasively by fluorescent output. In an extension of this concept, a theophylline-responsive ribozyme switch controlling a fluorescent reporter gene was used in a high-throughput screen of a large enzyme library of a caffeine demethylase, identifying a variant with 33-fold improvement in catalytic activity over eight rounds of directed evolution[51].

Synthetic RNA switches have demonstrated applications for medical purposes in human cells. In one notable example, an RNA switch controlling alternative splicing

modulated protein expression levels in response to nuclear factor κB (NF-κB) and β-catenin, two signaling proteins with important roles in disease[25]. The device was able to influence cell fate by controlling the levels of a gene conferring sensitivity to a drug that induces apoptosis. In another example, ribozyme switches responsive to small molecule drugs were used to regulate the expression of the cytokines IL-2 and IL-15 in engineered T cells, thereby imparting drug-modulated control over T-cell proliferation and survival *in vitro* and *in vivo*[43]. This latter system was demonstrated in a mouse model, highlighting the potential application of this technology to improving the safety and efficacy of adoptive immunotherapy strategies.

The future holds many more potential applications in biosensors, biofuels and drug compounds from synthetic metabolic pathways, diagnostic tools, and next-generation gene therapies. Additionally, all new applications, as well as all of the demonstrations described above, provide insight into the underlying biological mechanisms on which they rely, increasing our understanding of natural systems and how to better manipulate those systems in the future.

Scope of thesis

This thesis describes the development of a synthetic RNA device platform for the regulation of gene expression in response to molecular signals. As described in Chapter 2, we began with the cis-acting ribozyme switch platform developed by Win and Smolke[19], attempting to divide the structure into two RNA strands such that their annealing would reconstitute the functional device. These trans-ribozyme-based devices

were designed to target synthetic sequences inserted into the 3' UTR of the target gene and were expressed in human cells. After optimizing the molecular design for maximal *in vivo* cleavage activity using a cis-ribozyme-based model system, the improved trans-ribozyme was coupled with additional RNA elements intended to increase the likelihood of binding and cleavage of the target strand. However, *in vivo* activity of trans-ribozymes was not established, likely due to the inability of the two RNA strands to properly hybridize inside the cell. Chapter 3 describes the development of protein-responsive ribozyme switches. We designed a variety of device architectures intended to respond to the bacteriophage MS2 coat protein through different switching mechanisms. We developed a genetic system for quantitative characterization of the activity of these devices in human cells. After demonstrating a range of regulatory capabilities among the various device designs, we investigated the impact of different MS2 subcellular localizations on device activity and found that the switch platform is able to respond to both cytoplasmic- and nuclear-localized ligand. Finally, we designed ribozyme switches to respond to other protein ligands in order to demonstrate the versatility of our device platform. Chapter 4 discusses future directions for this work and its contributions to the field.

References

1. Khalil, A. & Collins, J. Synthetic biology: applications come of age. Nat. Rev. Genet. **11,** 367–379 (2010).

2. Chen, Y. Y. et al. Synthetic biology: advancing biological frontiers by building synthetic systems. Genome Biol. **13,** 240 (2012).

3. Chen, Y. Y. & Smolke, C. D. From DNA to targeted therapeutics: bringing synthetic biology to the clinic. Sci. Transl. Med. **3,** 106ps42 (2011).

4. Ma, S. et al. DNA synthesis, assembly and applications in synthetic biology. Curr. Opin. Chem. Biol. **16,** 260–267 (2012).

5. Chang, A. L. et al. Synthetic RNA switches as a tool for temporal and spatial control over gene expression. Curr. Opin. Biotechnol. **23,** 679–688 (2012).

6. Thodey, K. & Smolke, C. D. Bringing it together with RNA. Science. **333,** 412–413 (2011).

7. Liang, J. C. et al. Engineering Biological Systems with Synthetic RNA Molecules. Mol. Cell **43,** 915–926 (2011).

8. Isaacs, F. J. et al. RNA synthetic biology. Nat. Biotechnol. **24,** 545–554 (2006).

9. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene **361,** 13–37 (2005).

10. Winkler, W. C. & Breaker, R. R. Genetic control by metabolite-binding riboswitches. Chembiochem **4,** 1024–1032 (2003).

11. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. Nat. Rev. Mol. Cell Biol. **5,** 451–463 (2004).

12. Winkler, W. C. & Breaker, R. R. Regulation of bacterial gene expression by riboswitches. Annu. Rev. Microbiol. **59,** 487–517 (2005).

13. Sudarsan, N., Barrick, J. & Breaker, R. Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA **9,** 644–647 (2003).

14. Stefani, G. & Slack, F. J. Small non-coding RNAs in animal development. Nat. Rev. Mol. Cell Biol. **9,** 219–230 (2008).

15. Win, M. N. et al. Frameworks for programming biological function through RNA parts and devices. Chem. Biol. **16,** 298–310 (2009).

16. Mathews, D. H. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. RNA **10,** 1178–1190 (2004).

17. Mathews, D. H. & Turner, D. H. Prediction of RNA secondary structure by free energy minimization. Curr. Opin. Struct. Biol. **16,** 270–278 (2006).

18. Zadeh, J. et al. NUPACK: analysis and design of nucleic acid systems. J. Comput. Chem. **32,** 170–173 (2011).

19. Win, M. N. & Smolke, C. D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc. Natl. Acad. Sci. U. S. A. **104,** 14283–14288 (2007).

20. Kennedy, A. B. et al. A versatile cis-blocking and trans-activation strategy for ribozyme characterization. Nucleic Acids Res. **41,** e41 (2013).

21. Tucker, B. J. & Breaker, R. R. Riboswitches as versatile gene control elements. Curr. Opin. Struct. Biol. **15,** 342–348 (2005).

22. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science. **249,** 505–510 (1990).

23. Ellington, A. & Szostak, J. In vitro selection of RNA molecules that bind specific ligands. Nature **346,** 818–822 (1990).

24. Weigand, J. E. & Suess, B. Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. Nucleic Acids Res. **35,** 4179–4185 (2007).

25. Culler, S. J. et al. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. Science. **330,** 1251–1255 (2010).

26. Ramachandran, P. V. & Ignacimuthu, S. RNA interference—a silent but an efficient therapeutic tool. Appl. Biochem. Biotechnol. **169,** 1774–1789 (2013).

27. Saito, H. et al. Synthetic human cell fate regulation by protein-driven RNA switches. Nat. Commun. **2,** 160 (2011).

28. Beisel, C. L. et al. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. Mol. Syst. Biol. **4,** (2008).

29. Beisel, C. L. et al. Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. Nucleic Acids Res. **39,** 2981–2994 (2011).

30. Kötter, P. et al. A fast and efficient translational control system for conditional expression of yeast genes. Nucleic Acids Res. **37,** e120 (2009).

31. Paraskeva, E. et al. A translational repression assay procedure (TRAP) for RNA-protein interactions in vivo. Proc. Natl. Acad. Sci. U. S. A. **95,** 951–956 (1998).

32. Nie, M. & Htun, H. Different modes and potencies of translational repression by sequence-specific RNA-protein interaction at the 5'-UTR. Nucleic Acids Res. **34,** 5528–5540 (2006).

33. Saito, H. et al. Synthetic translational regulation by an L7Ae-kink-turn RNP switch. Nat. Chem. Biol. **6,** 71–78 (2010).

34. Babiskin, A. H. & Smolke, C. D. Engineering ligand-responsive RNA controllers in yeast through the assembly of RNase III tuning modules. Nucleic Acids Res. **39,** 5299–5311 (2011).

35. Grabowski, P. J. et al. The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of Tetrahymena. Cell **23,** 467–476 (1981).

36. Jacquier, A. & Rosbash, M. Efficient Trans-Splicing of a Yeast Mitochondrial RNA Group II Intron Implicates a Strong 5' Exon-Intron Interaction. Science. **234,** 1099–1104 (1986).

37. Rodnina, M. V. et al. How ribosomes make peptide bonds. Trends Biochem. Sci. **32,** 20–26 (2007).

38. Prody, G. et al. Autolytic Processing of Dimeric Plant Virus Satellite RNA. Science. **231,** 1577–1580 (1986).

39. Birikh, K. R. et al. The structure, function and application of the hammerhead ribozyme. Eur. J. Biochem. **245,** 1–16 (1997).

40. Chi, Y.-I. et al. Capturing hammerhead ribozyme structures in action by modulating general base catalysis. PLoS Biol. **6,** e234 (2008).

41. Flores, R. et al. Hammerhead ribozyme structure and function in plant RNA replication. Methods Enzymol. **341,** 540–552 (2001).

42. Win, M. N. & Smolke, C. D. Higher-order cellular information processing with synthetic RNA devices. Science. **322,** 456–460 (2008).

43. Chen, Y. Y. et al. Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. Proc. Natl. Acad. Sci. U. S. A. **107,** 8531–8536 (2010).

44. Wei, K. Y. et al. A yeast-based rapid prototype platform for gene control elements in mammalian cells. Biotechnol. Bioeng. **110,** 1201–1210 (2013).

45. Ausländer, S. et al. A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. Mol. Biosyst. **6,** 807–814 (2010).

46. Wieland, M. et al. Engineering of ribozyme-based riboswitches for mammalian cells. Methods **56,** 351–357 (2012).

47. Wieland, M. & Hartig, J. S. Improved aptazyme design and in vivo screening enable riboswitching in bacteria. Angew. Chemie **47,** 2604–2607 (2008).

48. Liang, J. C. et al. A high-throughput, quantitative cell-based screen for efficient tailoring of RNA device activity. Nucleic Acids Res. **40,** e154 (2012).

49. Urnov, F. D. et al. Genome editing with engineered zinc finger nucleases. Nat. Rev. Genet. **11,** 636–646 (2010).

50. Bogdanove, A. J. & Voytas, D. F. TAL effectors: customizable proteins for DNA targeting. Science. **333,** 1843–1846 (2011).

51. Michener, J. K. & Smolke, C. D. High-throughput enzyme evolution in Saccharomyces cerevisiae using a synthetic RNA switch. Metab. Eng. **14,** 306–316 (2012).