Chapter I: Introduction*

The proper regulation of gene expression is critical to many biological processes occurring in the cell. It is becoming increasingly apparent that post-transcriptional processing pathways play significant roles in regulating gene expression in both prokaryotic and eukaryotic organisms where they direct a variety of complex cellular functions. The directions regulating these processes are encoded within the genome, and many of them act at the level of *cis*- or *trans*-acting RNA elements that bind a wide range of biological molecules such as DNA, proteins, and other RNA molecules. Recent discoveries reveal larger roles of RNA as a sophisticated gene-regulatory molecule, which can implement diverse regulatory forms, such as riboswitches[1-5] and ribozymes[6, 7], in regulating target gene expression. Therefore, RNA represents an attractive platform for the development of novel gene-regulatory tools for a variety of biotechnological and medical applications and present opportunities for the modular design and construction of synthetic RNA switches and sensors to program different cellular functions.

1.1. RNA as a versatile and powerful gene-regulatory element

Ribonucleic acid (RNA) has traditionally been recognized for its role as a passive messenger of genetic information between the genome and the proteome in all living organisms. However, there are an increasing number of discoveries of naturally-occurring RNA molecules that act as regulatory elements, performing various cellular functions including gene expression regulation through sophisticated mechanisms, thereby expanding its traditional role as a genetic messenger and revealing it as a functionally versatile molecule.

*Sections 1.1.-1.4: Reproduced/adapted with permission from: M. N. Win and C. D. Smolke. (2006) "Regulating Gene Expression through Engineered RNA Technologies". Metabolic Engineering Protocols: Synthesis and Design Strategies. In: Walker J, series ed., Methods in Molecular Biology. Totawa: Humana Press. (book chapter in preparation)

Riboswitches[1-5], ribozymes[6, 7], antisense RNAs[8, 9], and small interfering and microRNAs (siRNAs and miRNAs, respectively)[10-13] are examples of RNA elements that exert their regulatory effects at different levels of the gene expression pathway such as transcription, translation, splicing, or decay. Unlike messenger RNAs (mRNAs), these regulatory RNAs are often noncoding RNAs (ncRNAs) or do not encode protein information. In addition, these regulatory elements are implemented through diverse physical compositions that can be grouped generally into *cis*- and *trans*-acting elements. In the former composition, the regulatory element is integrated within the transcript that harbors the target gene, whereas in the latter composition the regulatory element is a separate RNA molecule that acts on the transcript harboring the target gene through RNA-RNA binding interactions between the two individual molecules.

RNA exhibits a wide variety of functional properties, including catalytic, gene-regulatory, and ligand-binding activities. In addition, integrated RNA regulatory molecules have been characterized that achieve more sophisticated control over the expression of target proteins through a combination of these functional properties. These functional properties are encoded within the nucleotide sequence of an RNA molecule, which subsequently dictates its secondary and tertiary structure and ultimately its function. RNA adopts different conformations by folding into secondary and tertiary structures, which interact with various cellular constituents such as DNA, proteins, small molecules, and other RNA molecules[14, 15]. Furthermore, RNA molecules exhibit structural flexibility, which enable them to dynamically adopt different conformations. The binding of cellular and environmental molecules to particular RNA conformations has been demonstrated to regulate the equilibrium distribution between stable conformational states[16-18]. Unlike larger biomolecules such as proteins, the

functional activity of RNA is more directly defined by its secondary structure. This relationship between RNA secondary structure and function, in combination with predictive RNA secondary structure / energetic folding programs and rational and/or combinatorial design strategies, has enabled molecular engineers to construct synthetic 'designer' regulatory RNA elements[17-19]. In addition, technological advances in RNA engineering have demonstrated the programming of regulatory properties through alteration of nucleotide composition and ultimately RNA structure-function relationships[17-19].

Recent research in RNA biology and engineering supports the model of RNA as a versatile and powerful molecule possessing biologically-relevant gene-regulatory properties. Advances in RNA technology and nucleic acid engineering have allowed researchers to apply naturally-occurring RNAs as basic regulatory platforms and to develop more sophisticated regulatory RNA elements that involve integrated designs of multiple platforms. These synthetic riboregulators enable gene expression to be regulated in a more controlled manner and represent powerful tools for fundamental research and exhibit important applications in biotechnology and medical research.

1.2. RNA as sensory elements that exhibit universal sensing/binding properties

RNA molecules are functionally diverse and structurally flexible, and exhibit a wide range of regulatory properties such as catalytic, interactive, and allosteric binding properties. A unique property of RNA molecules is their sensing/binding capability to various types of inputs ranging from temperature to varied molecular ligands. Similar to other regulatory properties, RNA employs this property in exerting its diverse functional roles.

1.2.1. RNA elements that serve as thermosensors

RNA secondary structure is known to be highly dependent on temperature such that RNA assumes different structures in response to changes in temperature. Examples of such temperature-responsive RNA elements have been described[20]. For instance, genes that encode small heat-shock proteins and regulators of heat shock-responsive genes were found to contain sequences in their 5' untranslated regions (UTRs) that are capable of sequestering the Shine-Dalgarno (SD) sequence or prokaryotic ribosomal binding site (RBS) and the start codon AUG. The dependent adoption of RNA secondary structures at different temperatures is used to regulate the accessibility of these sequences by the ribosome, which subsequently modulates target gene expression levels, thereby enabling these RNA elements to serve as thermosensing gene expression regulators.

1.2.2. RNA elements that bind nucleic acids

Instances where RNA molecules exert their regulatory activities through base-pairing interactions with other RNA molecules are widespread in both natural and synthetic biological systems. For instance, a class of RNA molecules, called antisense RNAs, are single-stranded, *trans*-acting non-coding RNA elements, whose sequences are complementary to target transcripts and usually consist of 12-20 complementary nucleotides[8]. These RNA regulatory elements bind to their target transcripts through a sequence-specific hybridization event, resulting in the inhibition of gene expression from the bound mRNA[21]. As a natural example, SgrS is an antisense RNA that down-regulates the expression of the glucose transporter through base-pairing with the *ptsG* transcript when localized to the cell membrane, thereby lowering the accumulation of toxic phosphosugar metabolites[22, 23]. A

good synthetic example of RNA-RNA interactions through which a gene-regulatory function was achieved was described by Isaacs et al.[24]. In their engineered system, an RNA sequence segment was integrated into a location upstream of the RBS of a reporter gene to serve as a nucleic acid-sensing domain. This segment sequesters the RBS in the absence of the target RNA molecule, thereby suppressing the expression of the reporter gene. In the presence of the target RNA molecule, the sensor domain becomes bound to the target through RNA-RNA base-pairing interactions and releases the RBS for efficient translation, thereby serving as a nucleic acid-binding sensor domain.

1.2.3. RNA elements that bind molecular ligands

The ability of RNA structural elements to bind specific molecular ligands has been characterized in several natural systems[1]. However, researchers have also generated many examples of synthetic RNA ligand-binding elements, referred to as aptamers, in the laboratory. Synthetic aptamers can be generated through a standardized *in vitro* selection process known as SELEX (Systematic Evolution of Ligands by EXponential enrichment)[25, 26] (Figure 1.1). SELEX provides a very powerful selection method through which nucleic acid molecules exhibiting rare and specific binding properties to a ligand of interest can be generated *de novo* by selecting for functional binding activities from large randomized nucleic acid pools through iterative *in vitro* selection and amplification cycles. *In vitro* RNA aptamer selection schemes begin with a large pool of single-stranded RNA molecules generated through *in vitro* transcription from a DNA library. Aptamer pools are usually comprised of 30-70 randomized nucleotides, in order to generate an initial sequence diversity between 10^{14} to 10^{15} molecules[27]. The pool is incubated with the target ligand of interest and

subject to a partitioning event to separate bound members from unbound members. The most commonly used partitioning schemes are based on affinity chromatography. Bound (functional) members are recovered and then amplified through reverse transcription and polymerase chain reaction (PCR) to yield a pool enriched for target binding. This enriched pool will become the input pool for the next round of selection.

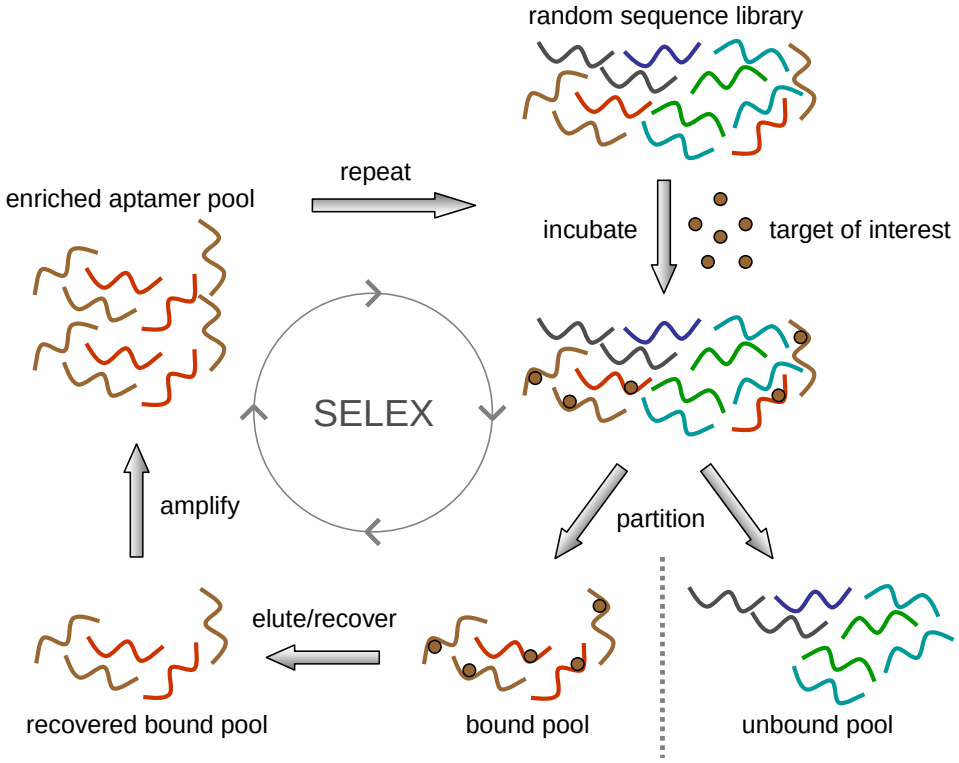

Figure 1.1. A schematic illustration of an *in vitro* selection process known as SELEX. The process starts with a large randomized pool of single-stranded RNA molecules transcribed from their DNA templates. The RNA pool is then incubated with target molecules of interest followed by separation of target-bound pool, which is reverse-transcribed to cDNA and amplified for the next selection cycle. The selection cycle is repeated typically for 8-15 cycles.

SELEX is particularly powerful in that RNA sequences that bind particular ligands can be generated *de novo* and the binding properties of the resulting aptamers can be programmed as desired. Specifically, aptamer binding properties such as affinities and specificities can be programmed by tailoring the stringency and counter-selections during

each selection. Aptamer affinities are tailored through the stringency of each selection cycle, normally by modifying wash volumes and target concentrations, whereas aptamer specificities are tailored through performed counter-selections with molecular analogues to the target. Typically eight to fifteen selection cycles are required to generate aptamers with high binding affinities and specificities. Recent work has demonstrated that protein aptamer selection schemes can be automated using standard robotics[28-30]. In addition, partitioning schemes for protein aptamer selections based on capillary electrophoresis have been recently developed that provide several advantages over conventional affinity-based partitioning schemes[31-33]. In particular, the efficiency of separation between the bound and unbound pools is significantly greater such that aptamers can be generated in fewer selection cycles. Synthetic aptamers have been generated to a wide range of target ligands, including small molecules, antibiotics, carbohydrates, amino acids, peptides, proteins[34], and even organelles such as phospholipid bilayers[35, 36], indicating that synthetic aptamers can be potentially generated to any targets of interest for user-specific applications.

1.3. Riboswitches are ligand-responsive RNA regulators of gene expression

All living organisms must manage the expression of many different genes in response to different signals such as metabolic demands and environmental changes[37]. This type of genetic management requires highly responsive sensors that accurately measure the magnitude of a particular signal and subsequently modulate the amount of the appropriate gene products to be synthesized. Proteins have traditionally been viewed as being the responsible sensor molecules for these signals. However, recent discoveries have

demonstrated that elements within mRNAs, termed riboswitches, are also capable of performing such tasks.

Riboswitches are *cis*-acting RNA elements that modulate the expression of target genes through integrated sensor and regulatory domains. This integration scheme enables riboswitches to sense their target ligands, typically cellular metabolites, through direct binding interactions and thus autonomously mediate their own functional activity in response to changing metabolite levels. While the majority of riboswitches characterized to-date have been discovered in bacteria, it has been shown more recently that these complex RNA regulatory elements are also present in eukaryotes[5, 38]. Riboswitches exhibiting unique mechanistic properties have also been recently identified. For example, the glycine riboswitch exhibits cooperative binding to its metabolite product, in which the metabolite binding turns on the gene expression of the enzyme responsible for the glycine cleavage system[3]. This cooperative-binding feature is present in the glycine riboswitch and is proposed to ensure that the metabolite is indeed in excess after consumption to provide carbon flux through the citric acid cycle while maintaining sufficient amounts of the amino acid available for protein synthesis. In another example, a tandem riboswitch system was characterized that exhibits a Boolean logic ability and functions as a two-input NOR logic gate, in which the two ligands (S-adenosylmethionine and coenzyme B_{12}) can independently suppress the target gene expression by binding their corresponding aptamers located upstream of a structure resembling an intrinsic transcription terminator[4]. Another tandem riboswitch system was very recently discovered that consists of two distinct riboswitches[39]. Unlike the glycine riboswitch and the two input-responsive logic-gate riboswitch systems, this riboswitch system does not exhibit cooperative ligand binding or detect two different metabolites,

respectively. This tandem riboswitch system responds independently to the same metabolite, thiamine pyrophosphate (TPP), and is predicted to function in concert to yield a more 'digital' gene control output response than a single riboswitch system.

1.3.1 General composition and conformational dynamics of riboswitches

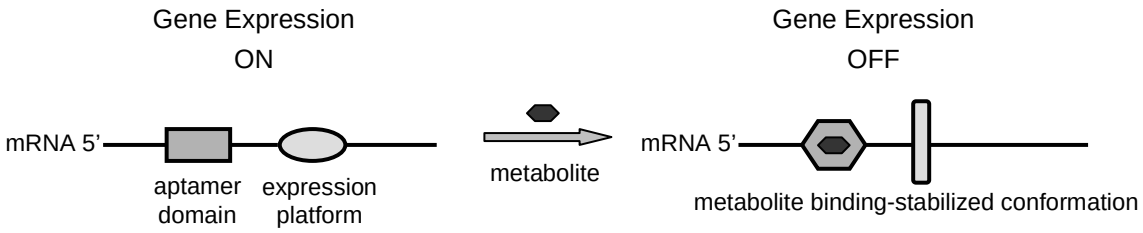

Figure 1.2. A schematic diagram of a typical riboswitch composed of two distinct domains: the ligand-binding domain known as the aptamer domain and the regulatory domain known as the expression platform (adapted from Winkler and Breaker[37]). Metabolite binding to the aptamer domain enables the stabilization of the rearranged conformation of the riboswitch (right), resulting in a shift in the equilibrium distribution between the two regulatory conformations that leads to the metabolite-dependent regulation of target gene expression.

Riboswitches are naturally-occurring, metabolite-responsive gene control elements primarily located within the 5' UTRs of cellular transcripts[37]. A riboswitch is typically comprised of two domains: the ligand-binding or sensor domain known as the aptamer domain, and the gene-regulatory domain known as the expression platform (Figure 1.2)[37]. Both domains are structurally flexible and capable of adopting different conformations. Riboswitches accomplish ligand-controlled regulation of gene expression through targeted dynamic switching between two primary conformations at equilibrium: one in which the regulatory domain is active and the other in which it is inactive. One of these conformational states is associated with the formation of the ligand-binding pocket within the aptamer domain, whereas the other carries an incorrectly formed binding pocket. Therefore, riboswitches can either repress or activate the expression of the target gene by assuming an

appropriate combination of different conformational states adopted by the aptamer and regulatory domains. Ligand binding to the riboswitch shifts the equilibrium distribution between these stable conformations to favor the ligand-bound form, thereby resulting in an allosteric gene regulation event. Most of the riboswitches characterized to date down-regulate the expression of the target gene; however, in a few exceptions such as the glycine[3] and adenine riboswitches[40], target gene expression is activated upon binding of the metabolite-ligand to the riboswitch.

1.3.2. Mechanisms of ligand-controlled gene regulation by riboswitches

Riboswitches regulate target gene expression *in cis* in response to changing metabolite levels through different mechanisms involving transcription termination[40, 41], translation initiation[42], mRNA processing[43], and splicing[5] (Figure 1.3). Transcription termination takes place through the mediated formation of a rho-independent terminator stem, which is usually GC rich, thereby destabilizing the transcription elongation complex[41]. Regulation can also target the disruption of the formation of a terminator stem upon metabolite binding, which allows proper transcription and thus up-regulation of target expression levels[40]. Riboswitches can also mediate translation initiation by undergoing adopting a secondary structure that interferes with ribosomal access to the target gene, such as sequestering the ribosome-binding site (RBS) or Shine-Dalgarno (SD) sequence in prokaryotic cells[42]. Regulation targeting transcript processing or deactivation can be achieved through expression platforms comprised of self-cleaving ribozymes, where the target transcripts undergo a ligand-directed cleavage event[43]. Regulation through splicing has recently been demonstrated in a filamentous fungus, in which metabolite binding to its

riboswitch can either repress or activate the expression of the main protein product by modulating the splice site choice through structural rearrangements[5]. Metabolite-binding domains have also been found within the 3' UTRs of transcripts in certain organisms, suggesting that riboswitch-mediated gene control may also occur through the regulation of mRNA stability[38].

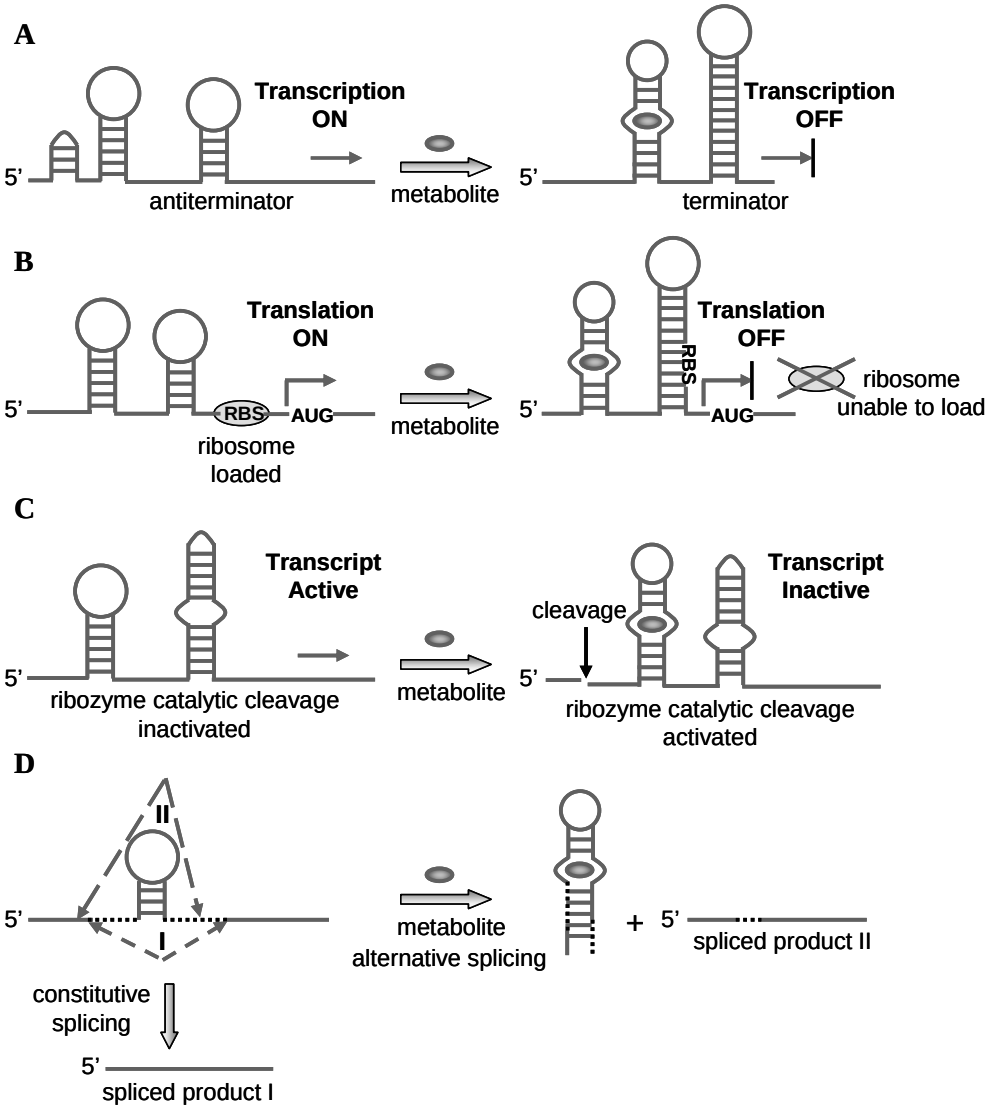

Figure 1.3. A schematic illustration of mechanisms by which riboswitches achieve gene expression regulation in response to their target metabolite binding (A and B adapted from Nudler and Mironov[44]). Ligand-regulated mechanisms involve (A) the formation of a transcription terminator stem, (B) sequestering the RBS and inhibiting translation initiation, (C) mRNA processing through catalytic cleavage of the transcript, and (D) alternative splicing using different sets of splice sites, I and II, respectively.

1.3.3. Riboswitch targets and implementation in metabolic networks

Most of the riboswitches characterized to-date have been discovered in bacteria. Cells employ these elements as genetic regulators in many fundamental metabolic pathways in response to changing metabolite levels. Target metabolites include various classes of small molecules such as amino acids, nucleotide bases, and coenzymes[45]. The presence of the integrated sensor domain enables riboswitches to sense intracellular metabolite concentrations through specific binding interactions and subsequently regulate expression levels of the associated gene product through allosteric conformational changes. Typically, this gene product is an enzyme directly involved in the biosynthesis, biodegradation, and/or transport of the target metabolite[45]. This mode of regulation provides a direct dynamic relationship between the intracellular metabolite concentration and the expression levels of the enzyme responsible for the metabolism, catabolism, and transport of the target metabolite. In addition, riboswitches are capable of binding their target metabolites with high specificities and affinities. Several different classes of natural riboswitches, their corresponding target metabolites, and functional roles in metabolic networks in various organisms are reviewed elsewhere in detail[45].

1.4. Current synthetic riboswitch systems for ligand-mediated regulation of target gene expression

Riboswitches are sophisticated gene control elements that achieve regulation by direct sensing of target metabolite levels and exhibit molecular recognition, high affinities, and precise control. Examples of the level of complexity achieved by these genetic regulatory elements include the self-cleaving ability of the *glmS* riboswitch[43], the alternative splicing

control of the TPP riboswitch[39], the cooperative binding property of the glycine riboswitch[3], and the NOR gate signal processing behavior of the SAM - coenzyme B_{12} riboswitch[4], demonstrating that riboswitches are powerful sensor-actuator control systems for autonomous gene expression control.

Inspired by the natural examples of riboswitches, numerous synthetic riboswitch systems have been engineered for artificial, ligand-mediated control of gene expression of target mRNAs. Many recent engineering efforts have focused on the construction of such synthetic ligand-controlled RNA-based gene regulatory elements through the integration of sensor and regulatory domains[18]. The flexibility in RNA regulatory systems, the programmability inherent in RNA design strategies, and the ability to generate sensor domains to potentially any molecular ligand of interest, enable such synthetic riboswitch systems to hold significant promise in transforming our ability to engineer cellular function. Therefore, engineered riboswitch control elements are attractive molecular tools for applications in synthetic network design where they can serve as powerful regulators of expression levels of genes of interest as well as *in vivo* biosensors of their corresponding target ligands through mechanisms of molecular recognition that go beyond base pairing. An overview of various engineered riboswitch systems that employ diverse gene-expression regulatory platforms is provided below to highlight the current technologies and challenges in the field of constructing synthetic riboswitches.

1.4.1. Riboswitch construction based on aptamer insertion within a transcript

Synthetic riboswitches have been constructed by inserting an aptamer or multiple aptamers directly into the 5’ UTR of a target mRNA in eukaryotes (Figure 1.4). Although the insertion location may be any region in the 5’ UTR, it is often chosen to be in the vicinity of

the cap region or the start codon of the transcript. This insertion strategy is a trail-and-error strategy, as the inserted aptamer(s) often does not result in ligand-mediated regulation of gene expression and may even cause substantial knockdown of the target gene in the absence of ligand. This strategy requires that the insertion of the aptamer itself and its associated secondary structure does not interfere with translation in the absence of the target ligand. The binding of ligand to the aptamer results in structural stabilization due to the molecular binding interaction between the aptamer and its target[34, 46]. Similar to binding of a protein to the 5' UTR[47, 48], this stabilized secondary structure can repress translation[49] presumably by interfering with ribosomal scanning or ribosome-mRNA interactions required for effective translation.

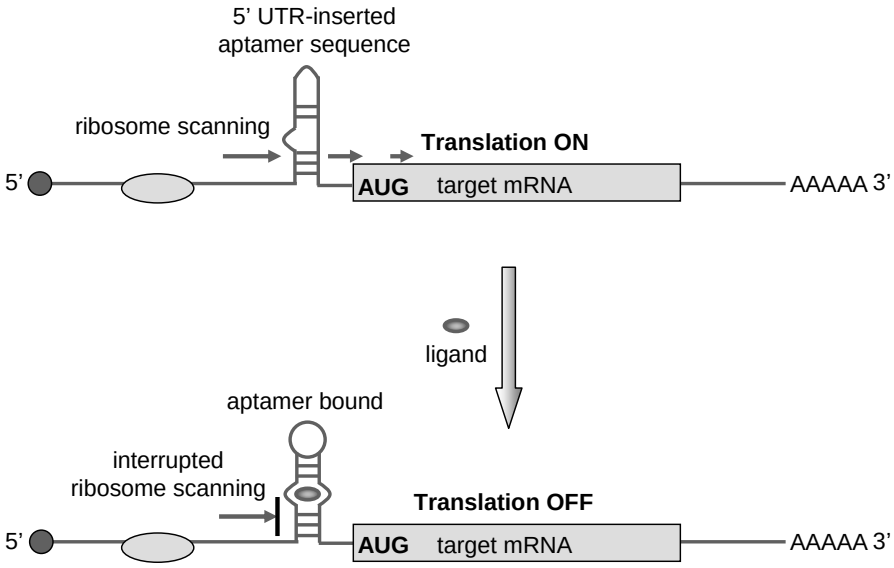

Figure 1.4. A schematic illustration of riboswitches constructed based on aptamer insertion within a target transcript. An aptamer or multiple aptamers can be inserted into the 5' untranslated region of an mRNA transcript near the 5' cap region or the start codon. Such an insertion may allow the aptamer-fused transcript region to adopt two primary conformations, one in which the aptamer binding pocket is disrupted (top) and the other in which the aptamer is correctly formed to reside its ligand (bottom), and the ligand binding shifts the equilibrium towards the latter conformation by stabilizing this conformation. This insertion strategy yet requires that the former conformation not introduce steric hindrance to ribosome for proper translation and that the ligand-bound latter conformation effectively inhibit translation through its stabilized structure.

Werstuck and Green[50] constructed the first examples of such riboswitches by inserting small molecule-binding RNA aptamers into the 5' UTR of transcripts. Translation was demonstrated to be repressed by the addition of the appropriate ligands both *in vitro* and *in vivo* in mammalian cells. Following this initial work, different research groups have constructed synthetic riboswitches through this design strategy using theophylline-[51], biotin-[52], and tetracycline-binding[53] aptamers, and demonstrated similar ligand-controlled gene regulation in different systems, including wheat germ extracts[54], *Xenopus oocyte*[54], and the budding yeast *S. cerevisiae*[55, 56].

Synthetic riboswitches have also been constructed to regulate translation of target genes in prokaryotes through a similar aptamer insertion strategy. Although still located in the 5' UTR of the target transcript, prokaryotes do not exhibit the same type of ribosomal scanning as eukaryotic organisms. Therefore, the physical implementation of these switches requires slightly different design strategies. In bacteria, the sequence distance between the ribosomal binding site (RBS) and the start codon is relatively short and varies between 5 to 13 nucleotides[57]. As a result, targeted insertion of an aptamer in this region to interfere with ribosomal scanning through a ligand-induced secondary structure is generally not applicable. In most bacteria, translation initiation relies on ribosomal accessibility to the RBS and the start codon, and thus mRNA secondary structure in the translational initiation region can dictate the efficiency of translation[58-60]. Desai and Gallivan developed a synthetic riboswitch system in *E. coli* where they inserted the theophylline aptamer[51] to a location five base-pairs upstream of the RBS to modulate ribosomal access to the RBS through ligand binding[61]. The theophylline-dependent up-regulation of gene expression by this synthetic riboswitch was demonstrated through plate-based screening and liquid culture assays. In addition, the extent

of up-regulation from this synthetic riboswitch was observed to be dramatically affected when the aptamer was moved to a slightly different location, two or eight base-pairs upstream of the RBS, indicating the functional sensitivity of this riboswitch system to the aptamer insertion location.

1.4.2. Riboswitch construction based on direct attachment of the aptamer to a regulatory element

Synthetic riboswitches have been constructed by directly attaching an aptamer to the regulatory domain, such that ligand binding to the aptamer inhibits the activity of the regulatory domain through some mechanism (Figure 1.5). This construction strategy is also a trial-and-error strategy, since the desired ligand-responsive regulatory activity may be highly specific to the location of attachment, the mechanism of action, and the specific aptamer-regulator pair.

A riboswitch system based on this strategy was developed by An et al. to modulate Dicer processing of an shRNA molecule through a small molecule ligand-aptamer interaction within this RNAi substrate[62]. In this example, the theophylline aptamer[51] was directly fused to an shRNA molecule, in place of the loop sequence. This shRNA construct silenced reporter gene expression in mammalian cells in a theophylline dose-dependent manner. Dicer cleavage of the aptamer-fused shRNA for subsequent generation of siRNAs was demonstrated to be modulated *in vitro* and *in vivo* by theophylline. This ligand-mediated regulation of Dicer processing was likely achieved due to locating the ligand-binding site of the aptamer sufficiently close to the Dicer processing site, such that theophylline binding to its aptamer blocks Dicer cleavage of the shRNA molecule, resulting in regulatable siRNA-based gene silencing. This proposed mechanism is supported by the observation that the

ligand-mediated regulatory effect was abolished when the shRNA stem was extended by one or two base-pairs, resulting in a small shift in the aptamer fusion point compared to the initial fusion design.

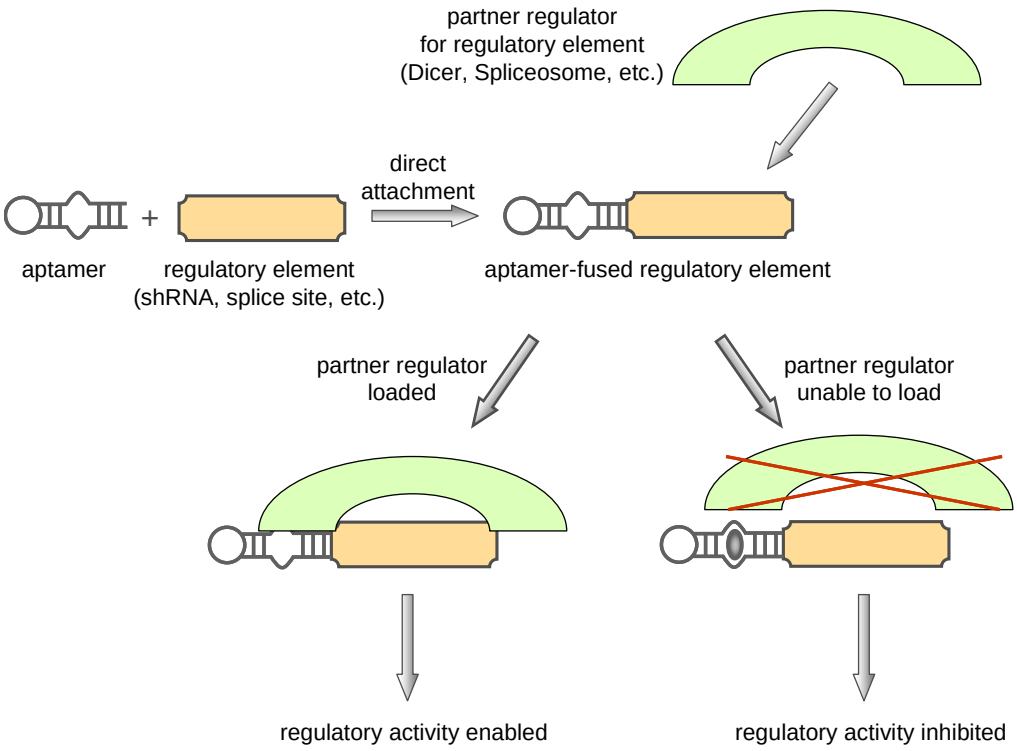

Figure 1.5. A schematic illustration of riboswitches constructed based on direct attachment between the aptamer and regulatory domains. An aptamer is directly attached to the regulatory platform in a way that the ligand-binding pocket within the aptamer is sufficiently close to the regulatory platform. In the absence of ligand, the partner regulatory element is capable of loading onto its platform and enabling the corresponding regulatory event to occur. In the presence of ligand, ligand binding to the aptamer and residing within the binding pocket create steric hindrance for the partner regulatory element loading to its platform, resulting in the inhibition of the normal regulatory event.

This direct attachment strategy has also been employed in constructing a synthetic riboswitch that regulates gene expression at the level of splicing. This riboswitch was designed by insertion of the theophylline aptamer[51] near the 3' consensus splice site region of a model pre-mRNA to modulate the splicing of the pre-mRNA through ligand-aptamer complex interactions[63]. The addition of theophylline was shown to repress the *in vitro*

splicing of the pre-mRNA harboring the aptamer and have no regulatory effect on the pre-mRNA without the aptamer. In addition, the aptamer's effect on splicing was demonstrated to be location-dependent and explained by modulating the spliceosome's accessibility to the splice site. The pre-mRNA harboring an aptamer with a stable base stem, inserted to encompass the 3' splice site AG within the ligand-binding sequence exhibited the most efficient splicing inhibition, as the splice site becomes less accessible when theophylline resides in the aptamer binding pocket.

Synthetic riboswitches constructed through the direct attachment strategy between the aptamer and regulatory domains exhibit functional dependence on the attachment location of the aptamer to the regulatory domain. This is because the ligand-mediated regulatory mechanism relies solely on how effective the ligand-aptamer complex interaction is in affecting the functional activity of the regulatory domain. This mechanism is not standardized and is highly specific to the particular aptamer-regulator pair. Therefore, such engineered riboswitches lack a reliable composition framework for integrating sensor and regulatory domains that results in allosteric binding properties.

1.4.3. Riboswitch construction based on an evolved linker between the aptamer and regulatory domains

Synthetic riboswitches have also been constructed by using a linker region that couples the aptamer and regulatory domains and serves as an element that translates the ligand-binding event in the aptamer domain to the adjacent regulatory domain. Early examples of evolved linker regions implemented a mechanism of information transmission known as 'helix slipping', in which a nucleotide shift event within the element is translated to a small-scale change in the conformation of the regulatory domain in a ligand-dependent

manner[64]. Such functional elements are often evolved through *in vitro* selection procedures and referred to as 'communication modules'[64]. These dynamic elements are typically three to five base-pairs long and often contain non-Watson-Crick base pairing.

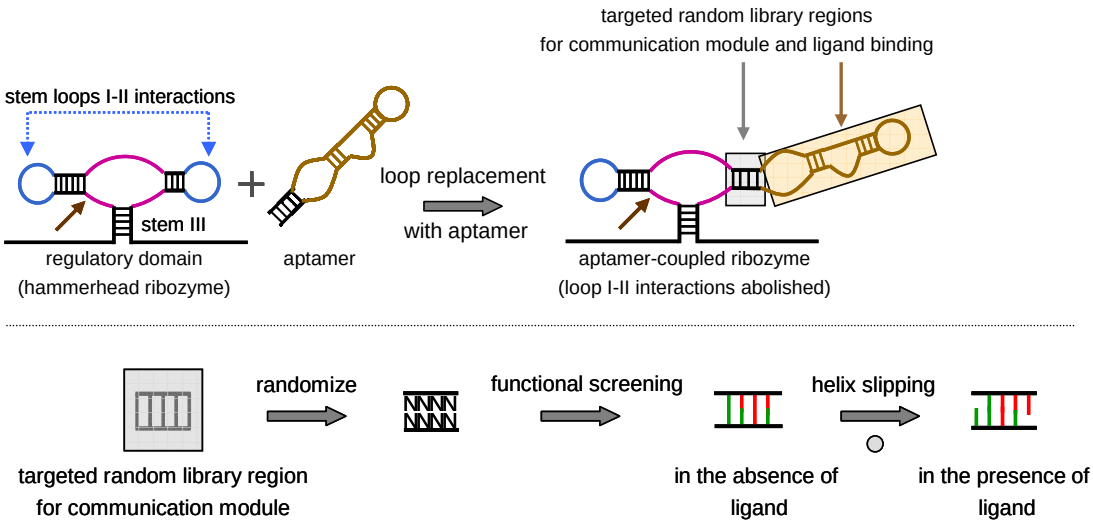

Figure 1.6. A schematic illustration of riboswitches constructed based on an evolved linker between the aptamer and regulatory domains. An *in vitro* functional communication module-based allosteric hammerhead ribozyme system is shown as an example. In general, an aptamer is attached to the regulatory domain through a linker (top), whose function is evolved to be ligand-dependent through selection from a random sequence library (bottom). The functional linkers are called 'communication modules', which employ the helix slipping mechanism in mediating the activity of the regulatory domain. An existing functional linker can also be used to mediate the activities of other regulatory platforms such as the RBS. In addition, an existing functional linker can also be used to couple a regulatory domain to an aptamer domain comprised of a random sequence library to evolve the latter to bind a new ligand of interest and function in this ligand-dependent manner (top). In this particular example of the *in vitro* allosteric ribozyme system, part of the aptamer domain replaces one loop of the ribozyme domain, thereby abolishing loop I-II interactions required for *in vivo* functionality.

Significant effort has been directed towards the construction of communication module-based riboswitches that use a hammerhead ribozyme as the regulatory domain (Figure 1.6), as ribozymes have proven to be a powerful platform for controlling gene expression. Several research groups have engineered a class of *in vitro* riboswitches called allosteric ribozymes[64-70]. Allosteric ribozymes resemble allosteric enzymes in that binding of

specific effectors, typically small molecule ligands, modulate the functional activities of the molecule[71]. An allosteric ribozyme contains two separate domains, a catalytic, or regulatory, domain and a ligand-binding, or aptamer, domain, which interact in a ligand-dependent manner to control the catalytic activity of the molecule[71]. Thus, the allosteric property of these ribozymes enables their catalytic activity to be regulated through specific ligands, and therefore may represent a modular design platform that can directly make use of different ligand-aptamer pairs.

Different strategies including rational design, library screening, and combinatorial approaches have been employed to generate allosteric hammerhead ribozymes[71, 72]. Rational design strategies involve integration of an existing aptamer domain directly to the catalytic domain of the ribozyme through different linkers, followed by examination of the activities of the resulting integrated constructs[65-67]. Library screening strategies involve screening randomized sequence libraries for novel aptamer domains (sometimes including communication modules) that function allosterically with the attached catalytic domain[73, 74]. Finally, the combined approach involves integration of an existing aptamer domain to the ribozyme's catalytic domain through a randomized linker region, and screening for functional linker sequences from this library that result in allosteric binding[64, 68, 69]. The majority of synthetic allosteric hammerhead ribozymes constructed to-date are responsive to small molecule ligands such as theophylline[51, 64], adenosine triphosphate (ATP)[65, 75], and flavin mononucleotide (FMN)[66, 76].

Researchers have also developed communication module-based riboswitches using a different regulatory platform such as the RBS. Suess et al. engineered a synthetic riboswitch comprised of a theophylline aptamer[51] and a previously developed communication module[71]

placed at a position proximal to the RBS[77]. This linker element had been proposed to perform helix slipping by one nucleotide between the ligand-bound and unbound states[71]. In their design, the communication module served as a helix bridge between the aptamer and the RBS such that binding of theophylline to its aptamer causes a single-nucleotide shift in the communication module, thereby enabling ribosome binding to the RBS without steric interference, and thus efficient translation in the presence of theophylline. This design scheme is similar to a direct coupling design between a theophylline aptamer and RBS described above, except that a distinct communication module was incorporated between the aptamer and regulatory domains.

Linker regions have also been evolved, which implemented a different mechanism of information transmission known as ‘strand displacement’, a functionally similar mechanism to ‘helix slipping’. Gallivan and colleagues developed a second riboswitch system as an extension of their initial direct attachment riboswitch design, using a combined rational and library screening design strategy[78]. A linker region adjoining the theophylline aptamer and the RBS was randomized, and sequences that translated a ligand-binding event in the aptamer domain to a structural change in the RBS, thereby regulating ribosomal access to the RBS, were screened through plate-based assays. These sequences are functionally similar to communication modules in that they translate ligand-binding events at the aptamer domain to the regulatory domain, but they are compositionally and mechanistically distinct. Gene expression regulation through this second class of linker regions takes place through the strand-displacement mechanism instead of a helix slipping mechanism. The functional sequences are complementary to regions of the theophylline aptamer such that base-pairing with a region of the aptamer sequesters the RBS and thus inhibits ribosomal access to the

RBS, whereas binding of theophylline to its aptamer disrupts this conformation and releases the RBS, resulting in up-regulation of target gene expression. As such, this regulatory mechanism is specific to the theophylline aptamer employed in this system and is not functionally independent. Consequently, this riboswitch system is not readily amenable to the insertion of different aptamers and thus lacks modularity, such that new linker regions would need to be generated by screening for specific aptamer-regulator pairs.

1.5. Further advancing the current field of engineering synthetic riboswitch systems

Numerous synthetic riboswitch systems have been developed for ligand-mediated regulation of functional activities both *in vitro* and *in vivo* as described above. These systems have made remarkable contributions in advancing the fielding of engineering synthetic RNA switches and sensors. During the past decade, the field has rapidly emerged and gained tremendous interest, as engineered riboswitches are effective regulatory elements that hold significant promise in transforming our ability to engineer cellular functions for various biotechnological applications through the functionally versatile biological substrate, RNA. Nevertheless, previous examples of synthetic riboswitches face one or more limitations such as failure to function in the cellular environment (*in vivo* functionality), requirement of specific cellular machinery (portability across different cellular systems), functional dependence among components within the switch molecule (modularity), programmability of the components (response tunability), and limited availability of molecular ligands or inputs that can be employed (scalability).

In order to further advance the promising field of riboswitch engineering, we set out to develop a riboswitch platform that exhibits the above-described functional properties.

Allosteric hammerhead ribozymes with *in vivo* functionality are highly attractive substrates for the development of a functionally versatile riboswitch platform for the following reasons: first, their regulatory mechanism employs self-cleavage and thus does not require cell-specific machinery (portability); and second, since RNA aptamers can be *de novo* generated for potentially any target ligands of interest using a standardized procedure known as SELEX[25, 26] and the sensor domain of allosteric hammerhead ribozymes is comprised of an RNA aptamer, it is amenable to many user-specified molecular inputs that can be employed (scalability). Therefore, we employed allosteric hammerhead ribozymes as platform substrates, incorporated engineering design principles and strategies into our design, and developed an *in vivo* functional riboswitch platform, called the ribozyme switch[79], that is modular in design, tunable in regulation, scalable in molecular input, and portable in regulatory mechanism across diverse cellular systems, thereby further advancing the current field of riboswitch engineering.

1.6. Interrelationship among the thesis projects

Chapter I provides an overview of RNA as a functionally versatile molecule, and current technologies and challenges in the field of riboswitch engineering, from which my thesis projects were developed. Chapter II describes work on one of my thesis projects, which provides a high-throughput method for functional characterization of small molecule-binding RNA aptamers. This method will enable robust, accurate, and rapid characterization of such RNA aptamers and can be very useful as we (and others) develop RNA aptamers for small molecules of specific interest that are to be integrated into the ribozyme switch platform as sensing elements for specific applications. Chapter III describes the detailed

work in the development of the above-mentioned extensible ribozyme switch platform for the reliable design and construction of ligand-controlled gene-regulatory systems applicable across different cellular systems. Chapter IV describes a sophisticated application aspect of our ribozyme switch platform through which higher-order RNA devices were built to achieve complex cellular information processing operations, including logic control, advanced computation, and cooperativity. Chapter V describes the functional extension of the small molecule-responsive ribozyme switch platform to respond to a different class of ligand molecules, proteins, in developing protein-responsive gene regulators and cellular biosensors. This extension broadens the platform utility to a wider range of biotechnological applications. These research projects synergistically support each other such that 'designer' gene-regulatory systems for various biotechnological and medical applications can be reliably and effectively constructed.

References

1. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. Nat Rev Mol Cell Biol 5, 451-63 (2004).
2. Breaker, R. R. Complex riboswitches. Science 319, 1795-7 (2008).
3. Mandal, M. et al. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science 306, 275-9 (2004).
4. Sudarsan, N. et al. Tandem riboswitch architectures exhibit complex gene control functions. Science 314, 300-4 (2006).

5. Cheah, M. T., Wachter, A., Sudarsan, N. & Breaker, R. R. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature 447, 497-500 (2007).

6. Doudna, J. A. & Cech, T. R. The chemical repertoire of natural ribozymes. Nature 418, 222-8 (2002).

7. Fedor, M. J. & Williamson, J. R. The catalytic diversity of RNAs. Nat Rev Mol Cell Biol 6, 399-412 (2005).

8. Simons, R. W. Naturally occurring antisense RNA control--a brief review. Gene 72, 35-44 (1988).

9. Inouye, M. Antisense RNA: its functions and applications in gene regulation--a review. Gene 72, 25-34 (1988).

10. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. Nature 431, 343-9 (2004).

11. Dykxhoorn, D. M., Novina, C. D. & Sharp, P. A. Killing the messenger: short RNAs that silence gene expression. Nat Rev Mol Cell Biol 4, 457-67 (2003).

12. Novina, C. D. & Sharp, P. A. The RNAi revolution. Nature 430, 161-4 (2004).

13. Tang, G. siRNA and miRNA: an insight into RISCs. Trends Biochem Sci 30, 106-14 (2005).

14. Breaker, R. R. Natural and engineered nucleic acids as tools to explore biology. Nature 432, 838-45 (2004).

15. Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. Nat Rev Mol Cell Biol 5, 908-19 (2004).

16. Soukup, J. K. & Soukup, G. A. Riboswitches exert genetic control through metabolite-induced conformational change. Curr Opin Struct Biol 14, 344-9 (2004).

17. Bauer, G. & Suess, B. Engineered riboswitches as novel tools in molecular biology. J Biotechnol 124, 4-11 (2006).

18. Isaacs, F. J., Dwyer, D. J. & Collins, J. J. RNA synthetic biology. Nat Biotechnol 24, 545-54 (2006).

19. Davidson, E. A. & Ellington, A. D. Engineering regulatory RNAs. Trends Biotechnol 23, 109-12 (2005).

20. Lai, E. C. RNA sensors and riboswitches: self-regulating messages. Curr Biol 13, R285-91 (2003).

21. Kurreck, J. Antisense technologies. Improvement through novel chemical modifications. Eur J Biochem 270, 1628-44 (2003).

22. Kawamoto, H., Morita, T., Shimizu, A., Inada, T. & Aiba, H. Implication of membrane localization of target mRNA in the action of a small RNA: mechanism of post-transcriptional regulation of glucose transporter in Escherichia coli. Genes Dev 19, 328-38 (2005).

23. Vanderpool, C. K. & Gottesman, S. Noncoding RNAs at the membrane. Nat Struct Mol Biol 12, 285-6 (2005).

24. Isaacs, F. J. et al. Engineered riboregulators enable post-transcriptional control of gene expression. Nat Biotechnol 22, 841-7 (2004).

25. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249, 505-10 (1990).

26. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. Nature 346, 818-22 (1990).

27. Rimmele, M. Nucleic acid aptamers as tools and drugs: recent developments. Chembiochem 4, 963-71 (2003).

28. Cox, J. C. & Ellington, A. D. Automated selection of anti-protein aptamers. Bioorg Med Chem 9, 2525-31 (2001).

29. Cox, J. C. et al. Automated acquisition of aptamer sequences. Comb Chem High Throughput Screen 5, 289-99 (2002).

30. Cox, J. C. et al. Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer. Nucleic Acids Res 30, e108 (2002).

31. Berezovski, M. et al. Nonequilibrium capillary electrophoresis of equilibrium mixtures: a universal tool for development of aptamers. J Am Chem Soc 127, 3165-71 (2005).

32. Drabovich, A., Berezovski, M. & Krylov, S. N. Selection of smart aptamers by equilibrium capillary electrophoresis of equilibrium mixtures (ECEEM). J Am Chem Soc 127, 11224-5 (2005).

33. Krylov, S. N. Nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM): A novel method for biomolecular screening. J Biomol Screen 11, 115-22 (2006).

34. Hermann, T. & Patel, D. J. Adaptive recognition by nucleic acid aptamers. Science 287, 820-5 (2000).

35. Khvorova, A., Kwak, Y. G., Tamkun, M., Majerfeld, I. & Yarus, M. RNAs that bind and change the permeability of phospholipid membranes. Proc Natl Acad Sci U S A 96, 10649-54 (1999).

36. Vlassov, A., Khvorova, A. & Yarus, M. Binding and disruption of phospholipid bilayers by supramolecular RNA complexes. Proc Natl Acad Sci U S A 98, 7706-11 (2001).

37. Winkler, W. C. & Breaker, R. R. Genetic control by metabolite-binding riboswitches. Chembiochem 4, 1024-32 (2003).

38. Sudarsan, N., Barrick, J. E. & Breaker, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes. Rna 9, 644-7 (2003).

39. Welz, R. & Breaker, R. R. Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. Rna 13, 573-82 (2007).

40. Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat Struct Mol Biol 11, 29-35 (2004).

41. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. An mRNA structure that controls gene expression by binding S-adenosylmethionine. Nat Struct Biol 10, 701-7 (2003).

42. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419, 952-6 (2002).

43. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. Nature 428, 281-6 (2004).

44. Nudler, E. & Mironov, A. S. The riboswitch control of bacterial metabolism. Trends Biochem Sci 29, 11-7 (2004).

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

46. Patel, D. J. et al. Structure, recognition and adaptive binding in RNA aptamer complexes. J Mol Biol 272, 645-64 (1997).

47. Stripecke, R., Oliveira, C. C., McCarthy, J. E. & Hentze, M. W. Proteins binding to 5' untranslated region sites: a general mechanism for translational regulation of mRNAs in human and yeast cells. Mol Cell Biol 14, 5898-909 (1994).

48. Paraskeva, E., Atzberger, A. & Hentze, M. W. A translational repression assay procedure (TRAP) for RNA-protein interactions in vivo. Proc Natl Acad Sci U S A 95, 951-6 (1998).

49. Pelletier, J. & Sonenberg, N. Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. Cell 40, 515-26 (1985).

50. Werstuck, G. & Green, M. R. Controlling gene expression in living cells through small molecule-RNA interactions. Science 282, 296-8 (1998).

51. Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. Science 263, 1425-9 (1994).

52. Wilson, C., Nix, J. & Szostak, J. Functional requirements for specific ligand recognition by a biotin-binding RNA pseudoknot. Biochemistry 37, 14410-9 (1998).

53. Berens, C., Thain, A. & Schroeder, R. A tetracycline-binding RNA aptamer. Bioorg Med Chem 9, 2549-56 (2001).

54. Harvey, I., Garneau, P. & Pelletier, J. Inhibition of translation by RNA-small molecule interactions. Rna 8, 452-63 (2002).

55. Suess, B. et al. Conditional gene expression by controlling translation with tetracycline-binding aptamers. Nucleic Acids Res 31, 1853-8 (2003).

56. Hanson, S., Berthelot, K., Fink, B., McCarthy, J. E. & Suess, B. Tetracycline-aptamer-mediated translational regulation in yeast. Mol Microbiol 49, 1627-37 (2003).

57. Kozak, M. Initiation of translation in prokaryotes and eukaryotes. Gene 234, 187-208 (1999).

58. de Smit, M. H. & van Duin, J. Control of prokaryotic translational initiation by mRNA secondary structure. Prog Nucleic Acid Res Mol Biol 38, 1-35 (1990).

59. de Smit, M. H. & van Duin, J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc Natl Acad Sci U S A 87, 7668-72 (1990).

60. de Smit, M. H. & van Duin, J. Control of translation by mRNA secondary structure in Escherichia coli. A quantitative analysis of literature data. J Mol Biol 244, 144-50 (1994).

61. Desai, S. K. & Gallivan, J. P. Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. J Am Chem Soc 126, 13247-54 (2004).

62. An, C. I., Trinh, V. B. & Yokobayashi, Y. Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. Rna 12, 710-6 (2006).

63. Kim, D. S., Gusti, V., Pillai, S. G. & Gaur, R. K. An artificial riboswitch for controlling pre-mRNA splicing. Rna 11, 1667-77 (2005).

64. Soukup, G. A. & Breaker, R. R. Engineering precision RNA molecular switches. Proc Natl Acad Sci U S A 96, 3584-9 (1999).

65. Tang, J. & Breaker, R. R. Rational design of allosteric ribozymes. Chem Biol 4, 453-9 (1997).

66. Araki, M., Okuno, Y., Hara, Y. & Sugiura, Y. Allosteric regulation of a ribozyme activity through ligand-induced conformational change. Nucleic Acids Res 26, 3379-84 (1998).

67. Soukup, G. A. & Breaker, R. R. Design of allosteric hammerhead ribozymes activated by ligand-induced structure stabilization. Structure 7, 783-91 (1999).

68. Soukup, G. A., Emilsson, G. A. & Breaker, R. R. Altering molecular recognition of RNA aptamers by allosteric selection. J Mol Biol 298, 623-32 (2000).

69. Kertsburg, A. & Soukup, G. A. A versatile communication module for controlling RNA folding and catalysis. Nucleic Acids Res 30, 4599-606 (2002).

70. Link, K. H. et al. Engineering high-speed allosteric hammerhead ribozymes. Biol Chem 388, 779-86 (2007).

71. Soukup, G. A. & Breaker, R. R. Nucleic acid molecular switches. Trends Biotechnol 17, 469-76 (1999).

72. Wilson, D. S. & Szostak, J. W. In vitro selection of functional nucleic acids. Annu Rev Biochem 68, 611-47 (1999).

73. Koizumi, M., Soukup, G. A., Kerr, J. N. & Breaker, R. R. Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. Nat Struct Biol 6, 1062-71 (1999).

74. Piganeau, N., Thuillier, V. & Famulok, M. In vitro selection of allosteric ribozymes: theory and experimental validation. J Mol Biol 312, 1177-90 (2001).

75. Sassanfar, M. & Szostak, J. W. An RNA motif that binds ATP. Nature 364, 550-3 (1993).

76. Burgstaller, P. & Famulok, M. Isolation of RNA apatmers for biological cofactors by in vitro selection. Angew. Chem. Int. Ed. Engl. 33, 1084-1087 (1994).

77. Suess, B., Fink, B., Berens, C., Stentz, R. & Hillen, W. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. Nucleic Acids Res 32, 1610-4 (2004).

78. Lynch, S. A., Desai, S. K., Sajja, H. K. & Gallivan, J. P. A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. Chem Biol 14, 173-84 (2007).

79. Win, M. N. & Smolke, C. D. From the Cover: A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc Natl Acad Sci U S A 104, 14283-8 (2007).