Chapter 1

Introduction

Advancing the construction of biological systems

Biological engineering is entering a new, unprecedented era of design and implementation. DNA synthesis technologies currently allow the generation of entire genomes, while continuing insights into the diversity and underlying mechanisms of life are providing the basis to construct autonomous cells and organisms with programmed behaviors and functions. Initial engineered systems resulting from these advances are already addressing global challenges, with prominent examples in sustainable energy production and inexpensive pharmaceutical synthesis (Keasling, 2008; Savage *et al*, 2008). However, these simple systems pale in comparison to the complexity, scale, and diversity of function observed in the natural world. To continue advancing the sophistication and scale of engineered systems, a common need exists to improve the availability and encoded functionalities of biological components that form the basis of these systems. With each development, new tools will emerge that will yield further insights into biological function and address current challenges in our world. While tool development has primarily focused on protein design, a new player – RNA – has gained prominence and is poised to reshape how we approach the construction of biological systems.

The many faces of RNA

RNA is a rich and versatile biomolecule composed of a linear chain of nucleotide units. Each unit includes one of four chemical bases (G, A, U, and C), a ribose sugar, and a negatively-charged phosphodiester backbone. Individual bases can pair through specific hydrogen bonding interactions (G to C, A to U), such that a strand of RNA can identify its sequence complement with high resolution. Base pairing can also occur internally, lending to the formation of complex secondary and tertiary structures. These structures can form unique epitopes for protein recognition, catalyze the cleavage or ligation of nucleic acid molecules, and bind diverse biomolecules with high affinity and specificity. Furthermore, the structure and activity of RNA sequences are amenable to computational prediction based on the hierarchical folding of RNA (Shapiro *et al*, 2007). Finally, the pivotal localization of RNA in gene expression translates the myriad of functionalities associated with RNA into potential gene regulatory strategies in cells.

Within biological systems, RNA is traditionally viewed as a passive information carrier in gene expression. However, the identification of endogenously-expressed RNAs displaying these different functionalities has radically redefined the natural role of RNA. The identified RNAs do not encode proteins – garnering the title noncoding RNAs – yet act as critical regulators in diverse cellular processes. Prominent examples include microRNAs in development (Erson and Petty, 2008), piwi-interacting RNAs in transposon defense and methylation patterning (Klattenhoff and Theurkauf, 2008), bacterial small RNAs in stress responses (Majdalani *et al*, 2005), self-cleaving ribozymes in viral reproduction (Lilley, 1999; Shih and Been, 2002), and riboswitches in dynamic metabolic control (Henkin, 2008). The current list may represent the tip of the iceberg, as

a plethora of recently identified noncoding RNAs still awaits characterization (Guttman *et al*, 2009).

Engineering riboswitches

Inspired by the emerging regulatory role of RNA in nature, biological engineers have begun designing synthetic RNA sequences that serve as genetic regulators in biological systems (Isaacs *et al*, 2006). Design efforts have predominantly focused on riboswitches, naturally occurring noncoding RNAs generally located in the 5' untranslated region (UTR) of genes involved in metabolite biosynthesis and breakdown (Winkler, 2005). Riboswitches dynamically activate (ON behavior) or repress (OFF behavior) the expression of downstream genes in response to changing metabolite concentrations (Figure 1.1). Two domains with prescribed functionalities are responsible for riboswitch activity: an aptamer domain that recognizes and binds a cognate metabolite and a gene regulatory domain that controls gene expression. Metabolites of diverse compositions are specifically recognized, including vitamins, amino acids, nucleic acid bases, metals, and secondary messengers. In addition, natural riboswitches utilize diverse regulatory mechanisms, including alternative splicing, transcriptional termination, translational repression, mRNA destabilization, and combinations thereof (Barrick and Breaker, 2007). In almost all instances of characterized riboswitches, metabolite binding stabilizes an alternative conformation associated with differential expression levels of the downstream gene. Therefore, riboswitches mediate an indirect link between molecular information in the cell and the genetic program primarily through conformational

stabilization, providing a natural RNA-based strategy to coordinate cellular processes in response to environmental perturbations.

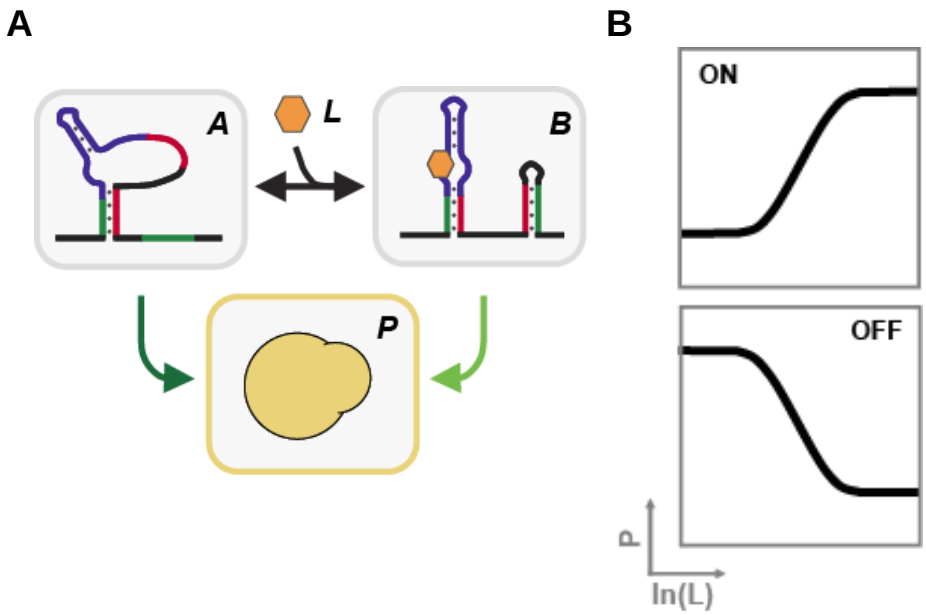

Figure 1.1 General schematic of riboswitch function. Riboswitches are generally composed of two domains that encode a ligand-binding aptamer and a genetic regulatory element. Ligand binding to the aptamer modulates the activity of the regulatory element, thereby affecting the expression levels of genes under the control of the riboswitch. (**A**) Most riboswitches adopt two distinct conformations designated A and B associated with differential expression levels of the encoded protein (P). Only one of the two conformations (conformation B) contains a formed aptamer shown in blue. Interaction between the aptamer and the ligand (L) stabilizes conformation B, thereby biasing conformational partitioning and subsequent protein expression levels. (**B**) Response curves relating ligand concentration (L) and protein levels (P). Ligand addition either increases (ON behavior) or decreases (OFF behavior) protein levels when the conformation with the formed aptamer is associated with higher or lower expression levels, respectively.

Initial efforts to design synthetic riboswitches began after the discovery of aptamers, the RNA elements responsible for molecular recognition. Novel aptamers can be generated in vitro through a procedure termed SELEX (systematic evolution of ligands by exponential enrichment) that isolates RNA sequences with specific ligand binding activity (Ellington and Szostak, 1990; Tuerk and Gold, 1990). Following the development of SELEX, aptamers have been selected against a wide range of molecules – well beyond the known number and types of molecules recognized by natural riboswitches. In an effort to couple genetic control and molecular recognition by aptamers, Werstruck and Green introduced aptamers selected against Hoechst dye into the 5' UTR of a mammalian reporter gene (Werstuck and Green, 1998). Addition of the dye to cells harboring the engineered reporter construct greatly reduced reporter production by perceivably inhibiting ribosomal loading. This seminal demonstration of ligand control of gene expression catalyzed the exploration of synthetic riboswitch design.

Since the communication of this seminal example in 1998, researchers have devised ingenious ways to combine aptamers and regulatory elements in order to introduce ligand control of gene expression. Riboswitch designs now include diverse regulatory mechanisms and operate in bacteria, yeast, and mammalian cells (Suess and Weigand, 2008), all working toward the design of riboswitches to control cellular processes in response to a selected ligand in different applications. Through the design process, a few desirable properties of riboswitches were identified that facilitate the implementation process. The first property is aptamer modularity – namely, the ability to introduce new aptamers in order to change the recognized ligand without riboswitch

redesign. By limiting the necessity for redesign, a single riboswitch design can be widely applied without significant time investment. A second property is tuning of the response curve, the quantitative relationship between ligand concentration and gene expression levels. By modifying the characteristics of riboswitch activity, such as aptamer affinity or the stability of the alternative conformation, the associated response curve could be rationally tuned to meet the regulatory demands of the desired application. A final property is the ability to control endogenous genes. Most riboswitch designs act in cis such that the riboswitch is included in the transcript of the regulated gene. However, directed recombination techniques necessary to couple endogenous genes with synthetic riboswitches are either technically challenging or unavailable in many organisms, limiting most riboswitch designs to the control of heterologous genes. At the start of this thesis work, only one riboswitch design attained regulation in trans, which relied on antisense inhibition as the regulatory mechanism (Bayer and Smolke, 2005). The challenges associated with antisense technology and the advent of a recently discovered trans regulatory mechanism, RNA interference (RNAi), provided the motivation to pursue the design of riboswitches that exploit this powerful mechanism of control (Sahu *et al*, 2007; Scherer and Rossi, 2003).

The emergence and importance of RNA interference

RNAi represents a powerful trans-targeting mechanism originally discovered only ten years ago in the worm *Caenorhabditis elegans* (Figure 1.2). After injecting *C. elegans* with double-stranded (ds)RNA homologous to a target gene, Fire and Mello observed dramatic silencing of the encoded protein from even a few molecules of the applied RNA

(Fire *et al*, 1998). The observed knockdown was sequence-specific, suggesting a targeted approach to silence any individual gene in the genome. Ensuing studies revealed an endogenous enzymatic pathway that conducts gene silencing through a process generally called RNA interference or RNAi. The general pathway was later identified in diverse organisms including insects, fungi, plants, and mammals, and unified seemingly unrelated silencing mechanisms such as quelling in fungi and post-transcriptional gene silencing in plants (Sharp, 1999).

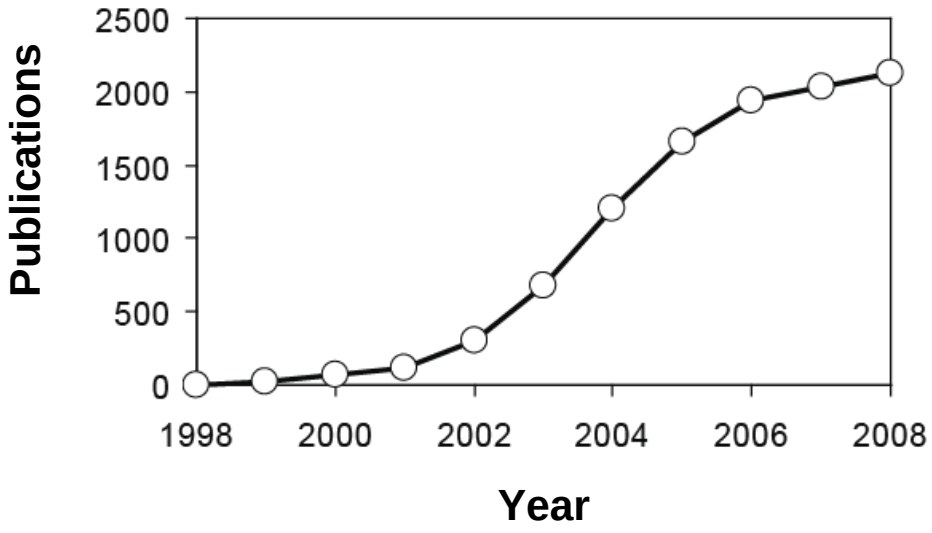

Figure 1.2 Number of RNAi publications each year since the discovery of RNAi in 1998. The search was conducted on the online NCBI PubMed database. Publications with the phrase "RNA interference" in the title or abstract were included in the tally. Notable landmarks in the timeline include the discovery of RNAi in mammalian cells in 2001 and the awarding of the Nobel Prize in Physiology or Medicine for the discovery of RNAi in 2006.

The RNAi pathway encompasses a linear processing cascade that sequentially reduces the applied dsRNA to a single strand of 21 – 24 nucleotides. This strand, termed

the guide strand, is loaded into the RNA induced silencing complex (RISC) and serves as a template to recruit complementary transcripts to RISC through base pairing interactions. Recruited transcripts that are partially or fully complementary to the guide strand are translationally inhibited or cleaved by RISC, thereby suppressing levels of the encoded protein.

RNAi can generally accept four different forms of dsRNA at distinct points in the processing cascade: microRNAs (miRNAs) at the beginning, small hairpin (sh)RNAs and long dsRNA in the middle, and small interfering (si)RNAs at the end (Figure 1.3). miRNAs are large imperfect hairpins present in pol II transcripts and exist as naturally-occurring effectors of RNAi. Processing by the nuclear RNase III Drosha (Lee *et al*, 2003) releases an shRNA with a characteristic 2-nt 3'overhang recognized for nuclear export by the shuttle protein Exportin-5 (Lund *et al*, 2004; Yi *et al*, 2003). Processing of shRNAs and long dsRNA by the cytoplasmic RNase III Dicer produces a duplex siRNA with symmetric 2-nt 3'overhangs. The siRNA is loaded into RISC, where one of the two strands is cleaved and discarded (Matranga *et al*, 2005; Rand *et al*, 2005) based on a collection of factors that include loading orientation and the relative thermodynamic stability on the 3' end of the retained strand (Amarzguioui *et al*, 2006). The remaining piece of the siRNA serves as the guide strand for transcript targeting. Note that the transcribed miRNA, the produced shRNA, and the final guide strand are also called primary, precursor, and mature miRNAs, respectively, within the natural miRNA field due to the difference in nomenclature used for naturally-occurring and synthetic RNAi effectors.

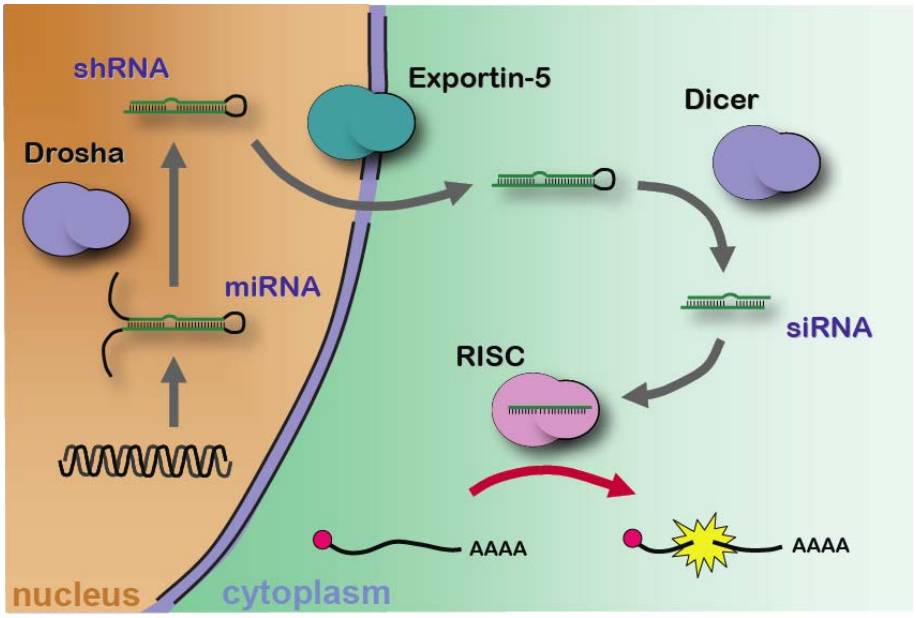

Figure 1.3 RNAi processing pathway and post-transcriptional gene silencing mechanism in mammals. The linear cascade terminates with the loading of RISC with the guide strand responsible for transcript targeting. The guide strand can be generated from three different RNAs: miRNAs, shRNAs, and siRNAs. miRNAs are processed to shRNAs by Drosha located in nucleus. Following export to the cytoplasm by Exportin-5, shRNAs are processed to siRNAs by Dicer. One of the two strands is retained by RISC as the guide strand.

With a greater understanding of the RNAi pathway came the advent of synthetic RNAi effectors for targeted gene silencing. Within mammalian cells, siRNAs were first developed for exogenous delivery in order to circumvent the non-specific interferon response (Elbashir *et al*, 2001). This response is induced by long dsRNA containing at least 30 consecutive base pairs and triggers a global shutdown of protein synthesis. shRNAs were developed after siRNAs as single RNA molecules that induce RNAi, where their use was bolstered by reports that Dicer processing improved RISC loading and subsequent gene silencing (Kim *et al*, 2005; Siolas *et al*, 2005). shRNAs have been

primarily used to endogenously activate RNAi since only one promoter is necessary to transcribe the molecule. With the discovery and characterization of the miRNA-processing enzyme Drosha, synthetic miRNAs dubbed second-generation shRNAs were developed (Silva *et al*, 2005). These RNAi effectors showed improved silencing over shRNAs presumably due to earlier entry into the RNAi processing pathway.

The ability to specifically silence any individual gene through RNAi has had a profound impact on fundamental biological research and holds tremendous potential in the treatment of genetic diseases. Through the facile design and implementation of RNAi effectors targeting individual genes, researchers are capable of rapidly deciphering genetic functions. Moving beyond single-gene studies, comprehensive libraries have been developed to target every expressed coding region in a genome (Haney, 2007). Associated large-scale screens have elucidated entire regulatory networks and identified novel members of critical signaling pathways (Berns *et al*, 2004; Mullenders *et al*, 2009; Tang *et al*, 2008).

The characteristic high specificity and efficiency of silencing has motivated the consideration of RNAi as a therapeutic strategy. RNAi effectors have been tested in model systems for cancer (Takeshita and Ochiya, 2006), HIV infection (Singh, 2008), neurodegenerative diseases (Farah, 2007), and macular degeneration (Whitehead *et al*, 2009) to silence genes associated with the disease state, with promising results that have spurred a flurry of clinical trials (Castanotto and Rossi, 2009). Related research efforts have revealed two major safety and efficacy concerns: off-target effects and cytotoxicity. Off-target effects result from the suppression of non-target genes with partial sequence complementarity to the guide strand, thereby compromising the specificity of targeting

and the potential therapeutic benefit (Kim and Rossi, 2009). Cytotoxicity has been attributed to induction of the interferon response and saturation of the RNAi machinery. Apart from dsRNA with 30 contiguous base pairs, high concentrations of dsRNA, the presence of 5’ phosphates, or some dsRNA sequences can efficiently trigger the innate immune response (Kim and Rossi, 2009). Saturating the RNAi machinery appears to impair the regular function of endogenously-expressed miRNAs, which may be critical for cellular homeostasis. Mouse studies revealed that constitutive over-expression of some shRNAs induced death within a month (Grimm *et al*, 2006). The toxic effect was relieved by co-expressing Exportin-5 or avoided by expressing synthetic miRNAs, suggesting alternative routes to improve the safety of RNAi therapeutics. Through further research efforts to overcome these challenges, the use of RNAi will progress rapidly in the clinic and may revolutionize the treatment of a broad range of genetic diseases.

Design of riboswitches that act through RNA interference

The work described in this thesis exploits the regulatory prowess of RNAi in riboswitch design and evaluates the quantitative relationship between riboswitch function and performance. Chapters 2 through 4 discuss the design of riboswitches that introduce ligand control of siRNA, shRNA, and miRNA processing. Designs rely on known structural requirements for efficient processing by the RNAi machinery elucidated through structural studies. A key pursuit in these chapters is achieving both aptamer modularity and tuning of the response curve to facilitate the widespread implementation of these designs toward different applications. The design of ligand-responsive shRNAs also integrated mathematical modeling and RNA folding algorithms toward the

development of sequence-function relationships for the forward design of riboswitches. Chapter 5 extends on these modeling efforts by investigating how the dynamics of riboswitch function dictate performance for various cis-acting regulatory mechanisms. A collection of design principles were elucidated to guide future design efforts and highlight areas for future experimental studies to better understand the design of synthetic riboswitches and the evolutionary optimization of natural riboswitches. Finally, Chapter 6 provides a perspective on where future research should be focused and the role RNA may play in the construction of biological systems.

REFERENCES

Amarzguioui M, Lundberg P, Cantin E, Hagstrom J, Behlke MA, Rossi JJ (2006) Rational design and in vitro and in vivo delivery of Dicer substrate siRNA. *Nat Protoc* **1:** 508-517.

Barrick JE, Breaker RR (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* **8:** R239.

Bayer TS, Smolke CD (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* **23:** 337-343.

Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B, Agami R, Ge W, Cavet G, Linsley PS, Beijersbergen RL, Bernards R (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428:** 431-437.

Castanotto D, Rossi JJ (2009) The promises and pitfalls of RNA-interference-based therapeutics. *Nature* **457:** 426-433.

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411:** 494-498.

Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346:** 818-822.

Erson AE, Petty EM (2008) MicroRNAs in development and disease. *Clin Genet* **74:** 296-306.

Farah MH (2007) RNAi silencing in mouse models of neurodegenerative diseases. *Curr Drug Deliv* **4:** 161-167.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391:** 806-811.

Grimm D, Streetz KL, Jopling CL, Storm TA, Pandey K, Davis CR, Marion P, Salazar F, Kay MA (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441:** 537-541.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458:** 223-227.

Haney SA (2007) Increasing the robustness and validity of RNAi screens. *Pharmacogenomics* **8:** 1037-1049.

Henkin TM (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev* **22:** 3383-3390.

Isaacs FJ, Dwyer DJ, Collins JJ (2006) RNA synthetic biology. *Nat Biotechnol* **24:** 545-554.

Keasling JD (2008) Synthetic biology for synthetic chemistry. *ACS Chem Biol* **3:** 64-76.

Kim DH, Behlke MA, Rose SD, Chang MS, Choi S, Rossi JJ (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat Biotechnol* **23:** 222-226.

Kim DH, Rossi JJ (2009) Overview of gene silencing by RNA interference. *Curr Protoc Nucleic Acid Chem* **Chapter 16:** Unit 16 11.

Klattenhoff C, Theurkauf W (2008) Biogenesis and germline functions of piRNAs. *Development* **135:** 3-9.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425:** 415-419.

Lilley DM (1999) Structure, folding and catalysis of the small nucleolytic ribozymes. *Curr Opin Struct Biol* **9:** 330-338.

Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U (2004) Nuclear export of microRNA precursors. *Science* **303:** 95-98.

Majdalani N, Vanderpool CK, Gottesman S (2005) Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol* **40:** 93-113.

Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123:** 607-620.

Mullenders J, Fabius AW, Madiredjo M, Bernards R, Beijersbergen RL (2009) A large scale shRNA barcode screen identifies the circadian clock component ARNTL as putative regulator of the p53 tumor suppressor pathway. *PLoS ONE* **4:** e4798.

Rand TA, Petersen S, Du F, Wang X (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123:** 621-629.

Sahu NK, Shilakari G, Nayak A, Kohli DV (2007) Antisense technology: a selective tool for gene expression regulation and gene targeting. *Curr Pharm Biotechnol* **8:** 291-304.

Savage DF, Way J, Silver PA (2008) Defossiling fuel: how synthetic biology can transform biofuel production. *ACS Chem Biol* **3:** 13-16.

Scherer LJ, Rossi JJ (2003) Approaches for the sequence-specific knockdown of mRNA. *Nat Biotechnol* **21:** 1457-1465.

Shapiro BA, Yingling YG, Kasprzak W, Bindewald E (2007) Bridging the gap in RNA structure prediction. *Curr Opin Struct Biol* **17:** 157-165.

Sharp PA (1999) RNAi and double-strand RNA. *Genes Dev* **13:** 139-141.

Shih IH, Been MD (2002) Catalytic strategies of the hepatitis delta virus ribozymes. *Annu Rev Biochem* **71:** 887-917.

Silva JM, Li MZ, Chang K, Ge W, Golding MC, Rickles RJ, Siolas D, Hu G, Paddison PJ, Schlabach MR, Sheth N, Bradshaw J, Burchard J, Kulkarni A, Cavet G, Sachidanandam R, McCombie WR, Cleary MA, Elledge SJ, Hannon GJ (2005) Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet* **37:** 1281-1288.

Singh SK (2008) RNA interference and its therapeutic potential against HIV infection. *Expert Opin Biol Ther* **8:** 449-461.

Siolas D, Lerner C, Burchard J, Ge W, Linsley PS, Paddison PJ, Hannon GJ, Cleary MA (2005) Synthetic shRNAs as potent RNAi triggers. *Nat Biotechnol* **23:** 227-231.

Suess B, Weigand JE (2008) Engineered riboswitches: overview, problems and trends. *RNA Biol* **5:** 24-29.

Takeshita F, Ochiya T (2006) Therapeutic potential of RNA interference against cancer. *Cancer Sci* **97:** 689-696.

Tang W, Dodge M, Gundapaneni D, Michnoff C, Roth M, Lum L (2008) A genome-wide RNAi screen for Wnt/beta-catenin pathway components identifies unexpected roles for TCF transcription factors in cancer. *Proc Natl Acad Sci U S A* **105:** 9697-9702.

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

Werstuck G, Green MR (1998) Controlling gene expression in living cells through small molecule-RNA interactions. *Science* **282:** 296-298.

Whitehead KA, Langer R, Anderson DG (2009) Knocking down barriers: advances in siRNA delivery. *Nat Rev Drug Discov* **8:** 129-138.

Winkler WC (2005) Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr Opin Chem Biol* **9:** 594-602.

Yi R, Qin Y, Macara IG, Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17:** 3011-3016.