

Copyright
by
Jay Richard Hesselberth
2003

**The Dissertation Committee for Jay Richard Hesselberth Certifies that this is
the approved version of the following dissertation:**

Identification and application of functional RNAs

Committee:

Andrew D. Ellington, Supervisor

Robin Gutell

Brent L. Iverson

Edward M. Marcotte

Jon D. Robertus

Identification and application of functional RNAs

by

Jay Richard Hesselberth, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2003

Dedication

Thank you, Laurie, for supporting me these last five years. This work would certainly not have been possible without the love and support of my family and friends.

Acknowledgements

I would like to acknowledge the members of the Ellington lab for their endless support. In particular, Andy let go of my reins for several years, allowing me to explore scientific avenues I would not have considered otherwise. Discussions I have had with Matthew Levy, Michael Robertson, Scott Knudsen, Jeff Tabor and Edward Marcotte have been pivotal in much of the work presented in this dissertation. Vlad Codrea completed the work described on using primary sequence homology for the identification of novel riboswitch elements. I would like to thank Jeffrey Barrick (Ronald Breaker Lab, Yale University) for providing annotated *Bacillus subtilis* structured RNAs (Table 5.1).

Identification and application of functional RNAs

Publication No. _____

Jay Richard Hesselberth, Ph.D.
The University of Texas at Austin, 2003

Supervisor: Andrew D. Ellington

RNA assumes many roles in nature. As a structural entity, it is responsible for proxying genetic information, ferrying amino acids and organizing ribosome architecture. RNA also participates in several genetic regulatory mechanisms. Catalytic RNAs are responsible for the faithful modification of unprocessed transfer, ribosomal and messenger RNAs, and the assembly of proteins from amino acid monomers may be performed by RNA-derived catalytic residues in the ribosome.

Previously researchers have developed methods for the isolation of unnatural RNA ligands and enzymes from random sequence pools. We have used these techniques in order to develop novel biosensors and biosensor components. RNA ligands (aptamers) were isolated from a random sequence pool that bind and inhibit ricin A-chain, a ribosome-inactivating protein and proposed biological

warfare agent. In addition, ribozymes that were previously engineered to respond to a set of ligands were incorporated into a sensor platform for the first demonstration of an RNA-based detection system for the simultaneous detection of diverse analyte classes including DNA, proteins, peptides and small-molecules.

We have also developed a set of computational tools in order to identify functional RNAs from both random sequence pools and genomic sequence databases. A computationally-accessible model was conceived in order to describe and predict the structural rearrangements undergone by allosteric ribozymes upon ligand binding, and the model was in turn used to select novel allosteric ribozymes from random sequence pools. In addition, we have used a structure-based search strategy in order to characterize well-folded regions in mRNAs with the goal of discovering additional examples of a novel class of genetic regulatory element, the riboswitches.

Table of Contents

List of Tables.....	xii
List of Figures	xiii
Chapter 1: Introduction	1
RNA in nature	1
RNA in technology.....	4
A method for the identification of functional nucleic acids.....	4
<i>In vitro</i> selection and application of aptamers.....	6
<i>In vitro</i> selection and application of ribozymes	8
Allosteric ribozymes	8
References	9
Chapter 2: <i>In vitro</i> selection of RNA molecules that inhibit the activity of ricin A-chain.....	14
Introduction	14
Ribosome inactivating proteins.....	14
<i>In vitro</i> selection of aptamers	15
Results and Discussion.....	16
<i>In vitro</i> selection of aptamers specific for ricin a-chain.....	16
Minimization of an anti-RTA aptamer.....	19
Anti-RTA aptamers bind RTA with high affinity.....	21
Anti-RTA aptamers compete with the SRL but are not depurinated ..	23
31RA inhibits ribosomal inactivation by RTA	25
Anti-RTA aptamers represent a novel class of nucleic acid inhibitors.....	26
Conclusions	28
Experimental Procedures.....	28
Pools and oligonucleotides.....	28

<i>In vitro</i> selection.....	29
Competition Assay	30
Filter Binding assays	31
Assay for RNA Depurination.....	32
Protein Synthesis Inhibition Assay	32
References	33
Chapter 3: An aptazyme ligase array for the detection of diverse analyte classes.....	38
Introduction	38
Biosensor design	38
Allosteric ribozymes as biosensor elements.....	40
Results and Discussion.....	42
Construction of an aptazyme ligase array.	42
Response characteristics of the aptazyme array	43
Sensitivity of the aptazyme array	45
Improvement of the signal-to-noise ratio	47
Alternative detection strategies	48
Aptamer derivatives as biosensor components	49
Conclusion.....	52
Experimental Protocols	52
Aptazymes and analytes	52
RNA preparation	53
Aptazyme arrays.....	53
References	54
Chapter 4: Computational engineering of allosteric ribozymes.....	58
Introduction	58
Computational design of biopolymers	58
The RNA folding problem	59
Aptazyme engineering.....	60

Results and Discussion.....	61
Computational assessment of slip structures.....	61
Computational selection of allosteric ribozymes	65
Varying the length of the joining region	67
Generalization of the method	70
A second design iteration	71
Aptamer structures could dictate design success	76
Other applications of the design algorithm	77
Conclusion.....	78
Experimental methods.....	79
Computational methods.....	79
<i>In vitro</i> transcription.....	80
Ribozyme assays	80
References	81
Chapter 5: Computational methods for the identification of riboswitches	85
Introduction	85
Results and discussion.....	87
Functional RNAs can be identified by structural signatures.....	87
Structural identification of <i>B. subtilis</i> riboswitches	88
Optimization of the ed-scan window size for functional RNA elements.....	92
Discovering structurally conserved RNA elements using orthology ..	94
Comparing local versus genomic scans	100
Using conservation of sequence and position for identifying riboswitch elements.....	101
Structure versus sequence based identification.....	105
Phylogenetic distribution of riboswitches	105
Conclusions	108
Experimental methods.....	108

Computer software	108
Comparing ed-scan scores with COGs.....	109
References	110
Appendix	113
Bibliography	114
Vita... ..	136

List of Tables

Table 2.1. Summary of the <i>in vitro</i> selection procedure used to isolate anti-RTA aptamers.	17
Table 3.1. A summary of available techniques for the construction of biosensors.	39
Table 3.2. Summary of aptazyme characteristics relevant to biosensor design ...	41
Table 4.1. Summary of computationally selected aptazymes.	69
Table 5.1 Summary of RNA structural elements found in <i>Bacillus subtilis</i>	90
Table 5.2. Summary of COGs containing structurally unique leader regions.	99

List of Figures

Figure 1.1. The roles of RNA are more complex than stated in the Central Dogma.	1
Figure 1.2. Generalized scheme for feedback in genetic regulation.	3
Figure 1.3. Generalized <i>in vitro</i> selection scheme.	5
Figure 1.4. Recognition in an Aptamer-protein complex.	7
Figure 2.2. The sequence of the anti-RTA aptamer is nearly optimal for its structure.	18
Figure 2.3. Comparison of anti-RTA aptamer conformers.	20
Figure 2.4. Binding isotherms for anti-RTA aptamers.	22
Figure 2.6. The anti-RTA aptamer inhibits depurination of the SRL by RTA. ...	24
Figure 2.7. The anti-RTA aptamer inhibits RTA activity in an <i>in vitro</i> protein synthesis assay.	26
Figure 3.1. Schematic of an aptazyme ligase array.	42
Figure 3.2. Aptazyme ligase array.	43
Figure 3.3. Response curves for arrayed aptazyme ligases.	44
Figure 3.4. Limits of detection determined using the aptazyme array.	46
Figure 3.5. Thorough washing leads to an increase in signal-to-noise ratio.	48
Figure 3.6. Aptamer beacon design strategy utilizing a pseudoknot to induce a conformation change.	51
Figure 4.1. Energetic profile model.	62
Figure 4.2. Computational selection of aptazymes.	63
Figure 4.3. Application of the energetic profile model.	66

Figure 4.4. Generality of the energetic profile model.	71
Figure 4.5. Incorporation of occupancy measures in the design algorithm.	73
Figure 5.1. Ed-scan trace for the purE-purK-purB operon of <i>Bacillus subtilis</i>	91
Figure 5.2. Optimization of the ed-scan window size for the hisS-asgS operon of <i>Bacillus subtilis</i>	93
Figure 5.3. Distributions of <i>B. subtilis</i> ed-scan scores from shuffled and native sequences.	94
Figure 5.4. RNA structural elements are not distinguished in the context of unrelated sequence.	101
Figure 5.5. Algorithm for identifying homologous positionally conserved sequence elements.	102

Chapter 1: Introduction

RNA IN NATURE

The central dogma was developed in order to explain the progression of the genetic program (**Figure 1.1**). In a cell, DNA encodes genetic units that are transcribed into RNA which are subsequently translated into proteins. This initial description has proven too simple, as several additional layers of complexity have since been found, many involving the role of RNA in the process.

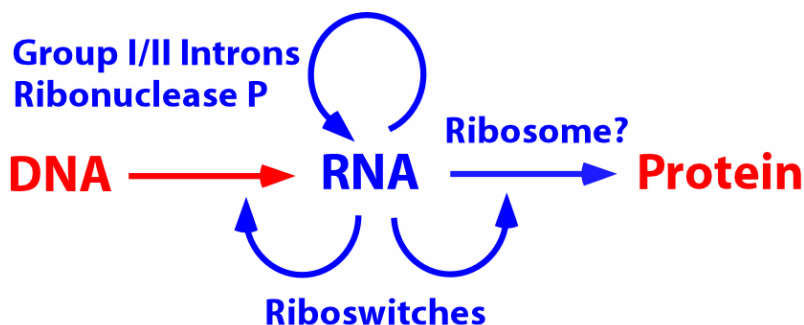


Figure 1.1. The roles of RNA are more complex than stated in the Central Dogma. As initially posed, the Central Dogma stated that RNA acted as an intermediate to transmit genetic information from genes encoded in DNA to the assembly of proteins. Since then, several additional roles for RNA in genetic regulation have been described (blue) including RNA enzymes that process mRNA, tRNA and rRNA (introns and ribonuclease P), mRNA elements that coordinate genetic regulation (riboswitches) and ribosomal RNA, which may contain the catalytic residues used by the ribosome to direct protein synthesis.

The first discoveries that RNA could act as a catalyst were made nearly 20 years ago. Thomas Cech and Sidney Altman won a Nobel Prize for their proof that RNA was capable of catalyzing specific reactions without the help of proteins. The systems they studied, a Group I self-splicing intron and Ribonuclease P, serve important regulatory functions in the processing of ribosomal and transfer RNAs, respectively (Guerrier-Takada et al., 1983; Kruger et al., 1982). These initial demonstrations were instrumental in not only illustrating the role of catalytic RNA in common cellular processes, but also in challenging the notion that proteins are the sole catalytic biological molecules.

Since those characterizations, several other RNAs have been characterized that play critical roles in genetic regulation. Simple ribozymes are used by several viruses for processing of their genomic material. The HDV, VS, hairpin and hammerhead ribozymes were all initially identified by studying the viral RNAs from which they derive (Symons, 1992). The structure of the ribosome has been solved to atomic detail (Ban et al., 2000; Yusupov et al., 2001). Perhaps the most studied RNA-protein complex, it is responsible for the mRNA-directed assembly of proteins from amino acid monomers. While it is unclear whether RNA residues are responsible for catalyzing the peptidyl-transferase reaction (Moore and Steitz, 2003), putative active-site nucleobases have been identified (Muth et al., 2000), and it has been noted that, in the crystal structure, the peptidyl transferase site of the ribosome is nearly devoid of amino side-chains (Ban et al., 2000; Moore and Steitz, 2003). While confirmation that the ribosome is a ribozyme remains elusive, it is interesting to consider that the machine largely responsible for protein synthesis in the cell may actually be an RNA enzyme.

While not catalytic entities in and of themselves, small non-coding RNA (ncRNAs) play a huge role in the processing of eukaryotic RNAs (the RNAi pathway) (Schwarz and Zamore, 2002; Zamore, 2002). These small ~21 nucleotide RNAs derive from larger RNA precursors that undergo a protein-complex directed processing step. The processed RNAs subsequently bind mRNA molecules, targeting them for degradation (Banerjee and Slack, 2002). In addition, other intriguing roles for these small RNAs have been suggested: it appears that they may be involved in chromatin remodeling and expression (Reinhart and Bartel, 2002) and to some extent may be responsible for developmental processes such as the cell cycle and patterning (Carmell et al., 2002). Similarly, a large number of ncRNAs in prokaryotes have also been discovered, and perform a variety of regulatory functions such as osmoregulation (OxyS) and targeted protein degradation (tmRNA) (Wassarman, 2002).

Very recently it has been shown that RNA has the capacity for a novel type of genetic regulation (**Figure 1.2**). Here mRNA elements dubbed 'riboswitches' serve as receptors for small-molecule metabolites and this interaction causes a change in a genetic program. Unlike previously described genetic feedback systems, this process goes on without the need for intervening proteins. For example, the metabolite thiamine pyrophosphate has been shown to bind the 5' untranslated regions of mRNAs coding for thiamine biosynthetic enzymes (Winkler et al., 2002a). These elements are evolutionarily conserved across all three major kingdoms of life, regulating thiamine biosynthesis and transport in a huge number of organisms (Rodionov et al., 2002). In addition to thiamine, riboswitches specific for riboflavin (Winkler et al., 2002b), cobalamin (Vitamin B12) (Nahvi et al., 2002) and S-adenosyl methionine (McDaniel et al., 2003) have also been described. The fact that these elements exist is not particularly surprising; the mechanism of regulation is simple and elegant, typically employing either translational attenuation or transcriptional termination. However, because the mechanism has remained undiscovered until now, discovery and characterization of these elements has become a burgeoning field (Hesselberth and Ellington, 2002).

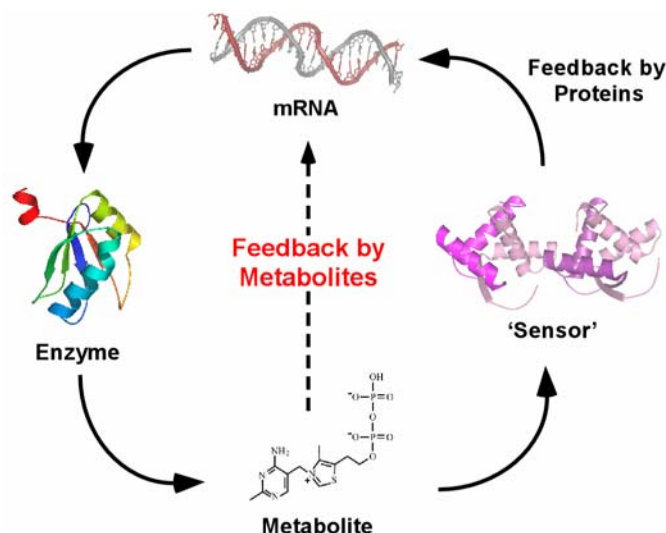


Figure 1.2. Generalized scheme for feedback in genetic regulation. Both RNA and protein serve as sensors which can control gene expression levels (Hesselberth and Ellington, 2002).

Because RNA is such a versatile molecule, a hypothesis in which an 'RNA World' was the precursor to modern life has been proposed (Gilbert, 1986). In this conceptual world, RNA takes center stage as the biopolymer which performs all of the reactions necessary for a sustainable metabolism. As the lone class of molecule, it would need to perform a variety of functions in order to survive, including, perhaps most intriguingly, the ability to self-replicate (Gesteland and Atkins, 1993).

RNA IN TECHNOLOGY

Nature has learned the utility of RNA and exploited its properties to serve a multitude of different functions. It is only fitting that researchers have taken their cue and developed a large number of RNA-based technologies, some which are inspired by nature, while others push the capabilities of RNA itself. There is still much undiscovered territory in RNA biology, and as researchers learn of its novel capabilities so will technologies appear that employ the utility of RNA.

A method for the identification of functional nucleic acids

In vitro selection was pioneered as a methodology for the isolation of active nucleic acid molecules from a randomized pool. Two major advances facilitated the widespread use of *in vitro* selection: chemical DNA synthesis (Caruthers et al., 1987) and the polymerase chain reaction (PCR) (Saiki et al., 1988). This ability to design and construct nucleic acid pools and primers coupled with the ability to amplify the resulting sequences using PCR has enabled researchers to construct customized nucleic acid pools for the purpose of identifying functional molecules from them. Functional molecules are identified by an iterative procedure involving application of a selection pressure (i.e. binding to a target), removal of unwanted species and regeneration of molecules capable of responding to the initial challenge (**Figure 1.3**).

Some of the first examples of *in vitro* selection were designed to identify RNA and DNA molecules that could specifically interact with a dye molecule (Ellington and Szostak, 1990; Ellington and Szostak, 1992) or bind a protein (Tuerk and Gold, 1990). In addition, RNAs able to catalyze an intermolecular cleavage reaction were also identified from a randomized library based on the Tetrahymena Group I ribozyme (Robertson and Joyce, 1990). After these initial demonstrations that functional nucleic acid sequences could be isolated from randomized pools, a host of selection schemes have developed for the identification of novel RNA and DNA molecules able to perform a desired function.

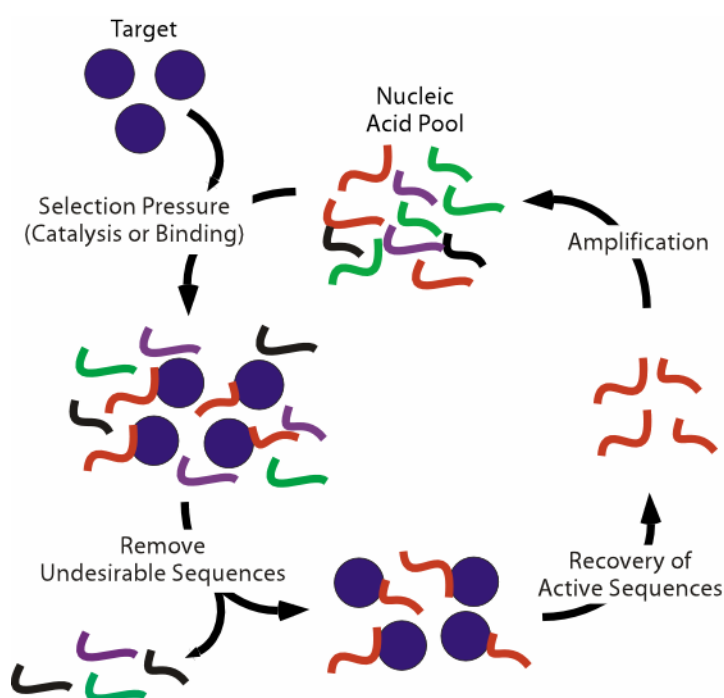


Figure 1.3. Generalized *in vitro* selection scheme. Beginning with a pool of randomized RNA or DNA (top), a selective pressure is placed on the pool. This pressure is the hallmark of a given selection scheme: one can look for nucleic acid species able to bind other molecules, or sequences capable of performing a catalytic reaction. Sequences unable to respond to the selective pressure are discarded, and the 'winning' molecules are regenerated to form a new pool. Iterations of this process yield sequences capable of performing the desired reaction.

Nucleic acid pools are typically designed such as to maximize the probability of finding a desired functionality, and have been empirically tuned for the type of functionality desired. For example, nucleic acid binding species are typically selected

from pools containing ~40 randomized nucleotides containing upwards of 10^{15} unique sequences, and it has been proposed that functional nucleic acid binding species can be isolated from much smaller pools (~600 unique sequences, or 1 zeptomole) (Knight and Yarus, 2003). *In vitro* selections for catalytic species typically employ more sequence diversity in the form of larger pools. Previous selections for catalytic RNA ligases have used pools containing 90-220 randomized nucleotide positions (Ekland et al., 1995; Robertson and Ellington, 1999), and from these pools highly active RNA enzymes can be isolated with complex sequence and structural requirements (Ekland and Bartel, 1995; Ekland et al., 1995).

***In vitro* selection and application of aptamers**

Aptamers are RNA or DNA molecules that bind ligands; they are conceptual equivalent of protein antibodies. RNA aptamers have been shown to discriminate between a single methyl group on a small-molecule (caffeine vs. theophylline) (Jenison et al., 1994) and typically bind their ligands with high affinity (pM-nM K_d s). In addition, they are capable of binding incredibly diverse classes of ligands: scores of anti-protein and small-molecule aptamers have been selected (<http://aptamer.icmb.utexas.edu>), and aptamers have even been selected against entire organisms (Homann and Goring, 1999). Several structural studies of aptamers have been carried out, including the crystal structures of several protein:aptamer complexes (Convery et al., 1998) and protein:small-molecule complexes, (Baugh et al., 2000; Sussman et al., 2000) in addition to detailed NMR experiments that illustrate the conformational changes undergone by an aptamer upon ligand binding (Hermann and Patel, 2000). The crystal structure of an aptamer specific to a coat protein from MS2 bacteriophage (**Figure 1.4**) illustrates the capability of these non-natural molecules to interact with their natural counterparts.

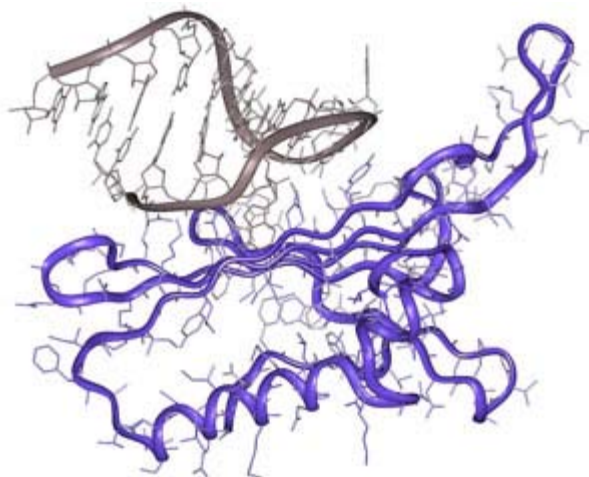


Figure 1.4. Recognition in an Aptamer-protein complex. An RNA aptamer (gray) selected against MS2 bacteriophage coat protein (blue) was crystallized with its protein target (Convery et al., 1998).

Because nucleic acid aptamers bind with high affinity and are extremely specific for their ligands, they are ideal candidates for adaptation to therapeutic and sensor platforms. Coupled with the fact that they can be synthetically manipulated and made in large quantities, these molecules have yielded several novel applications of nucleic acid in technology. For example, conformational changes associated with the binding of an aptamer to its target can be exploited to yield immediate, 'real-time' signals. By careful engineering or using a modified *in vitro* selection procedure, fluorescently labeled aptamers can be created that transduce ligand binding directly to change in their optical properties (Jhaveri et al., 2000a; Jhaveri et al., 2000b). The adaptation of aptamers to sensor platforms has been pioneered for the detection of small amounts of proteins in solution, (Fredriksson et al., 2002) as well as for medical diagnostic purposes (Brody and Gold, 2000). RNA aptamers have also been shown to have therapeutic value: aptamers that bind to HIV-1 Rev also inhibit viral replication when expressed in a cell (Symensma et al., 1996), and aptamers that bind coagulation factor IXa are potent anticoagulants in humans (Rusconi et al., 2002).

***In vitro* selection and application of ribozymes**

Since the discovery that RNA is capable of catalysis, significant interest has been focused on the isolation and characterization of novel ribozymes and *in vitro* selection has been used to isolate RNA molecules able to perform many diverse reactions. While all naturally derived ribozymes involve phosphodiester bonds in their reaction mechanisms, other ribozymes have been selected that are able to catalyze a variety of chemical reactions such as carbon-carbon bond formation via a Diels-Alder cycloaddition (Tarasow et al., 1997) and porphyrin metalation (Conn et al., 1996; Li and Sen, 1996). In addition, ribozymes have been selected that mimic protein-catalyzed reactions like the charging of tRNAs with activated amino acids (Lee et al., 2000) or the production of dipeptides using aminoacyl adenylated substrates (Sun et al., 2002).

Researchers have also used *in vitro* selection to study ribozymes in terms of their fitness to do a particular function. For example, selections for thermostable Group I intron variants have been performed (Guo and Cech, 2002) and active site optimizations can be done (Ekland and Bartel, 1995; Robertson et al., 2001), yielding RNA progeny that are more capable of catalyzing a reaction than their parental sequence.

Allosteric ribozymes

While *in vitro* selected ribozymes are useful for illustrating that particular reactions can be accomplished using simple systems, or for comparison with the naturally occurring ribozymes, many times they are not directly applicable to biosensor platforms. To that end, several researchers have attempted to combine the reactivities of catalytic RNAs with the recognition properties of aptamers in an effort to create the RNA equivalent of allosteric proteins.

By combining the unique properties of a given aptamer and ribozyme, one can create novel RNA enzymes that are capable of responding to a particular analyte. Aptazymes have been made using both rational design (Tang and Breaker, 1998) and *in*

vitro selection strategies (Soukup and Breaker, 1999) and the resulting chimeric, catalytic RNAs have been shown to be dependent upon their ligands for activity (Soukup and Breaker, 2000).

Because such a wealth of chemistry can be explored using RNA, these molecules are easily adapted to diagnostic platforms. For example, a ribozyme ligase has been previously selected that was subsequently shown to be dependent upon a short oligonucleotide (the 3' primer used in the selection) for activity (Robertson and Ellington, 1999). The ribozyme could thus 'sense' the presence of its oligonucleotide cofactor, and could detect its presence at low nanomolar concentrations. Subsequent experiments have shown that by joining aptamer domains with existing ribozymes, chimeric molecules can be generated in which the ribozyme's activity is directly regulated by the presence of the cognate ligand for the aptamer. Such molecules are easily adapted to biosensor platforms, and indeed a significant amount of work has been done to show they are as capable as protein antibodies for the detection and quantitation of a wide variety of molecules.

RNA is a remarkable biopolymer. Herein we have attempted to exploit its properties in order to select and design novel examples of biosensors and biosensor reagents. In addition, we have developed computational methods for the identification of functional RNAs: in one instance we conceive and implement a method for the *in silico* selection of allosteric ribozymes, and also present several methods for the discovery and characterization of a novel class of RNA regulatory element, the riboswitches.

REFERENCES

Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905-920.

- Banerjee, D. and Slack, F. (2002) Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. *Bioessays*, **24**, 119-129.
- Baugh, C., Grate, D. and Wilson, C. (2000) 2.8 Å crystal structure of the malachite green aptamer. *J Mol Biol*, **301**, 117-128.
- Brody, E.N. and Gold, L. (2000) Aptamers as therapeutic and diagnostic agents. *J Biotechnol*, **74**, 5-13.
- Carmell, M.A., Xuan, Z., Zhang, M.Q. and Hannon, G.J. (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev*, **16**, 2733-2742.
- Caruthers, M.H., Barone, A.D., Beaucage, S.L., Dodds, D.R., Fisher, E.F., McBride, L.J., Matteucci, M., Stabinsky, Z. and Tang, J.Y. (1987) Chemical synthesis of deoxyoligonucleotides by the phosphoramidite method. *Methods in Enzymology*, **154**, 287-313.
- Conn, M.M., Prudent, J.R. and Schultz, P.G. (1996) Porphyrin Metalation Catalyzed by a Small RNA Molecule. *Journal of the American Chemical Society*, **118**, 7012-7013.
- Convery, M.A., Rowsell, S., Stonehouse, N.J., Ellington, A.D., Hirao, I., Murray, J.B., Peabody, D.S., Phillips, S.E. and Stockley, P.G. (1998) Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution. *Nat Struct Biol*, **5**, 133-139.
- Ekland, E.H. and Bartel, D.P. (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res*, **23**, 3231-3238.
- Ekland, E.H., Szostak, J.W. and Bartel, D.P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science*, **269**, 364-370.
- Ellington, A.D. and Szostak, J.W. (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818-822.
- Ellington, A.D. and Szostak, J.W. (1992) Selection *in vitro* of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature*, **355**, 850-852.
- Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gustafsdottir, S.M., Ostman, A. and Landegren, U. (2002) Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol*, **20**, 473-477.
- Gesteland, R.F. and Atkins, J.F. (eds.). (1993) *The RNA World*. Cold Spring Harbor Laboratory Press.
- Gilbert, W. (1986) The RNA world. *Nature*, **319**, 618.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849-857.

- Guo, F. and Cech, T.R. (2002) Evolution of Tetrahymena ribozyme mutants with increased structural stability. *Nat Struct Biol*, **9**, 855-861.
- Hermann, T. and Patel, D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820-825.
- Hesselberth, J.R. and Ellington, A.D. (2002) A (ribo) switch in the paradigms of genetic regulation. *Nat Struct Biol*, **9**, 891-893.
- Homann, M. and Goring, H.U. (1999) Combinatorial selection of high affinity RNA ligands to live African trypanosomes. *Nucleic Acids Res*, **27**, 2006-2014.
- Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425-1429.
- Jhaveri, S., Rajendran, M. and Ellington, A.D. (2000a) *In vitro* selection of signaling aptamers. *Nat Biotechnol*, **18**, 1293-1297.
- Jhaveri, S.D., Kirby, R., Conrad, R., Maglott, E.J., Bowser, M., Kennedy, R.T., Glick, G. and Ellington, A.D. (2000b) Designed Signaling Aptamers that Transduce Molecular Recognition to Changes in Fluorescence Intensity. *Journal of the American Chemical Society*, **122**, 2469-2473.
- Knight, R. and Yarus, M. (2003) Finding specific RNA motifs: Function in a zeptomole world? *Rna*, **9**, 218-230.
- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147-157.
- Lee, N., Bessho, Y., Wei, K., Szostak, J.W. and Suga, H. (2000) Ribozyme-catalyzed tRNA aminoacylation. *Nat Struct Biol*, **7**, 28-33.
- Li, Y. and Sen, D. (1996) A catalytic DNA for porphyrin metallation. *Nat Struct Biol*, **3**, 743-747.
- McDaniel, B.A., Grundy, F.J., Artsimovitch, I. and Henkin, T.M. (2003) Transcription termination control of the S box system: Direct measurement of S-adenosylmethionine by the leader RNA. *Proc Natl Acad Sci U S A*.
- Moore, P.B. and Steitz, T.A. (2003) After the ribosome structures: How does peptidyl transferase work? *Rna*, **9**, 155-159.
- Muth, G.W., Ortoleva-Donnelly, L. and Strobel, S.A. (2000) A single adenosine with a neutral pKa in the ribosomal peptidyl transferase center. *Science*, **289**, 947-950.
- Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L. and Breaker, R.R. (2002) Genetic control by a metabolite binding mRNA. *Chem Biol*, **9**, 1043.

- Reinhart, B.J. and Bartel, D.P. (2002) Small RNAs correspond to centromere heterochromatic repeats. *Science*, **297**, 1831.
- Robertson, D.L. and Joyce, G.F. (1990) Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, **344**, 467-468.
- Robertson, M.P. and Ellington, A.D. (1999) *In vitro* selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat Biotechnol*, **17**, 62-66.
- Robertson, M.P., Hesselberth, J.R. and Ellington, A.D. (2001) Optimization and optimality of a short ribozyme ligase that joins non-Watson-Crick base pairings. *Rna*, **7**, 513-523.
- Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. and Gelfand, M.S. (2002) Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *J Biol Chem*, **277**, 48949-48959.
- Rusconi, C.P., Scardino, E., Layzer, J., Pitoc, G.A., Ortel, T.L., Monroe, D. and Sullenger, B.A. (2002) RNA aptamers as reversible antagonists of coagulation factor IXa. *Nature*, **419**, 90-94.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491.
- Schwarz, D.S. and Zamore, P.D. (2002) Why do miRNAs live in the miRNP? *Genes Dev*, **16**, 1025-1031.
- Soukup, G.A. and Breaker, R.R. (1999) Engineering precision RNA molecular switches. *Proc Natl Acad Sci U S A*, **96**, 3584-3589.
- Soukup, G.A. and Breaker, R.R. (2000) Allosteric nucleic acid catalysts. *Curr Opin Struct Biol*, **10**, 318-325.
- Sun, L., Cui, Z., Gottlieb, R.L. and Zhang, B. (2002) A selected ribozyme catalyzing diverse dipeptide synthesis. *Chem Biol*, **9**, 619-628.
- Sussman, D., Nix, J.C. and Wilson, C. (2000) The structural basis for molecular recognition by the vitamin B 12 RNA aptamer. *Nat Struct Biol*, **7**, 53-57.
- Symensma, T.L., Giver, L., Zapp, M., Takle, G.B. and Ellington, A.D. (1996) RNA aptamers selected to bind human immunodeficiency virus type 1 Rev *in vitro* are Rev responsive *in vivo*. *J Virol*, **70**, 179-187.
- Symons, R.H. (1992) Small catalytic RNAs. *Annu Rev Biochem*, **61**, 641-671.
- Tang, J. and Breaker, R.R. (1998) Mechanism for allosteric inhibition of an ATP-sensitive ribozyme. *Nucleic Acids Res*, **26**, 4214-4221.

- Tarasow, T.M., Tarasow, S.L. and Eaton, B.E. (1997) RNA-catalysed carbon-carbon bond formation. *Nature*, **389**, 54-57.
- Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.
- Wassarman, K.M. (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell*, **109**, 141-144.
- Winkler, W., Nahvi, A. and Breaker, R.R. (2002a) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952-956.
- Winkler, W.C., Cohen-Chalamish, S. and Breaker, R.R. (2002b) An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci U S A*, **99**, 15908-15913.
- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. and Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883-896.
- Zamore, P.D. (2002) Ancient pathways programmed by small RNAs. *Science*, **296**, 1265-1269.

Chapter 2: *In vitro* selection of RNA molecules that inhibit the activity of ricin A-chain

INTRODUCTION

Ribosome inactivating proteins

Ribosome-inactivating proteins (RIPs) inhibit protein synthesis by disabling the cellular translation machinery. Class I RIPs have a catalytic A-chain that recognizes and depurinates a universally conserved adenosine found in a GAGA tetraloop in the sarcin-ricin loop (SRL) of eukaryotic 23-28S rRNA. (Endo et al., 1987; Endo and Tsurugi, 1987; Endo and Tsurugi, 1988) Class II RIPs are functionally identical to the Class I proteins, but possess an additional lectin B-chain that is required for cell-surface attachment and subsequent endocytosis of the toxin domain (Lord et al., 1991).

Ricin, a member of the class II RIPs, is an enzyme derived from the castor bean plant (*Ricinus communis*). Detailed structures of the ribosome have recently become available, (Ban et al., 2000; Schlutzen et al., 2000; Yusupov et al., 2001) resulting in the structural elucidation of the SRL. The loop protrudes from the ribosome like a finger, and the action of ricin is similarly analogous to clipping a nail. Removal of the crucial adenine ring results in structural rearrangements in the RNA-protein complex, precluding binding of additional elongation factors. While detailed structural information of RIP-inactivated ribosomes does not exist, the site has been probed by chemical means and implicates interaction with either eEF-1 or eEF-2 (Larsson et al., 2002). This mechanism of inactivation is extremely effective and is well-conserved among a large group of RIP proteins (Lord et al., 1991).

Because ricin is an extremely efficient, potentially lethal enzyme and can be produced in large quantities inexpensively, it has been classified by the United States Centers for Disease Control as a 'Class B' biological agent, being 'moderately easy to

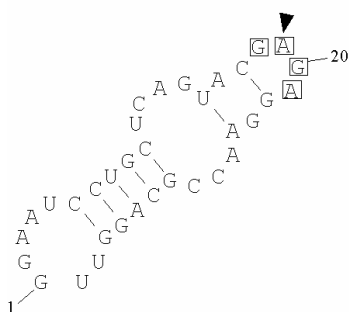
disseminate' and resulting in moderate to low morbidity rates (<http://www.bt.cdc.gov/Agent/agentlist.asp>, 2003). Ricin's effectiveness as a weapon is related to its efficiency as an enzyme: once inside a cell, one ricin molecule is sufficient to cause death, and the protein has been reported to inactivate 1777 ribosomes min⁻¹ (Endo and Tsurugi, 1988).

***In vitro* selection of aptamers**

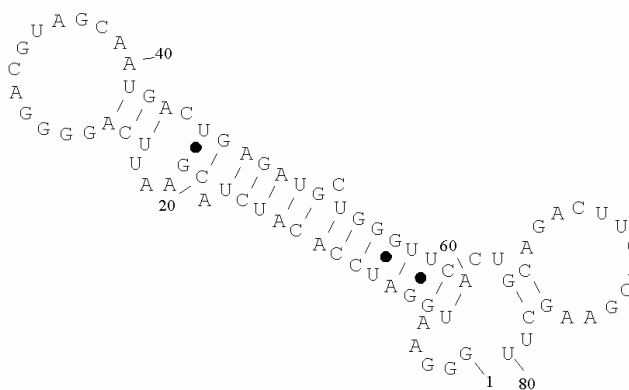
In vitro selection is a powerful molecular tool that can be used to generate nucleic acid molecules that bind with both specificity and affinity to a wide variety of analytes (Burgstaller et al., 2002; Famulok, 1999). RNA aptamers have been shown to discriminate between a single methyl group on a small-molecule (caffeine vs. theophylline) (Jenison et al., 1994) and have been selected against entire organisms (Homann and Goring, 1999). These nucleic acid ligands can be readily adapted to function as either therapeutics or sensors. By careful engineering or using a modified *in vitro* selection procedure, fluorescently labeled aptamers can be created that transduce ligand binding directly to change in their optical properties (Jhaveri et al., 2000a; Jhaveri et al., 2000b). The adaptation of aptamers to biosensor platforms has also been pioneered for the detection of small amounts of proteins in solution, (Fredriksson et al., 2002) as well as for medical diagnostic purposes (Brody and Gold, 2000). RNA aptamers have also been shown to have therapeutic value: aptamers that bind to HIV-1 Rev also inhibit viral replication when expressed in a cell (Symensma et al., 1996). In consequence, aptamers that recognize and potentially inhibit ricin might be useful prophylactic or therapeutic agents, or could be used as biosensor elements for the detection of aerosolized ricin or areas contaminated by the toxin.

***In vitro* selection of aptamers specific for ricin a-chain**

(a)



(b)



16

Following nine rounds of selection the resulting pool was cloned and sequenced. A survey of 24 individual sequences indicated that a single species dominated the population. This sequence was subsequently folded using the mfold algorithm (Zuker, 2000) in order to ascertain its secondary structure. The initial indication suggested that the most stable conformer consisted of a long stem capped by a large loop. This structure is indicated in **Figure 2.1b**.

N30 selection					
	Input	Input			
Round	[RNA]	[RTA]	[RNA]:[RTA]	% bound to RTA	% bound to filter
	nM	nM			
1	2500	500	5:1		
2	2500	500	5:1		
3	800	26.7	30:1		
4	800	26.7	30:1	1.06	0.73
5	800	26.7	30:1		
6	800	26.7	30:1		
7	800	26.7	30:1	19.92	7.88
8	800	800	1:1		
9	800	800	1:1	51.09	2.41
D30 selection					
	Input	Input			
Round	[RNA]	[RTA]	[RNA]:[RTA]	% bound to RTA	% bound to filter
	nM	nM			
1	1000	200	5:1	5.62	2.13
2	1000	200	5:1	15.52	2.86
3	2000	200	10:1	27.65	2.52
4	2000	200	10:1	28.72	1.38

Table 2.1. Summary of the *in vitro* selection procedure used to isolate anti-RTA aptamers. Two separate selection procedures are summarized; the first (N30) was used to select anti-RTA aptamers, while the D30 selection was used to optimize the originally selected sequence.

In order to determine which sequences in the aptamer contributed to recognition of RTA, a doped sequence population was prepared based on the apparent secondary structure of the aptamer. The pool was designed such that every position of the original random region in addition to a portion of the 5' constant region was doped at 85%. In other words, as a population the pool contained sequences that had 85% the wild-type residue at each position, and 5% each of the other three nucleotides. Previous selections utilizing doped pools have identified sequences that were more capable of binding their ligands than the originally selected sequence (Bartel et al., 1991). This pool was

amplified as before, except that the 5' primer used for PCR was three nucleotides shorter to prevent re-incorporation of the original sequence.

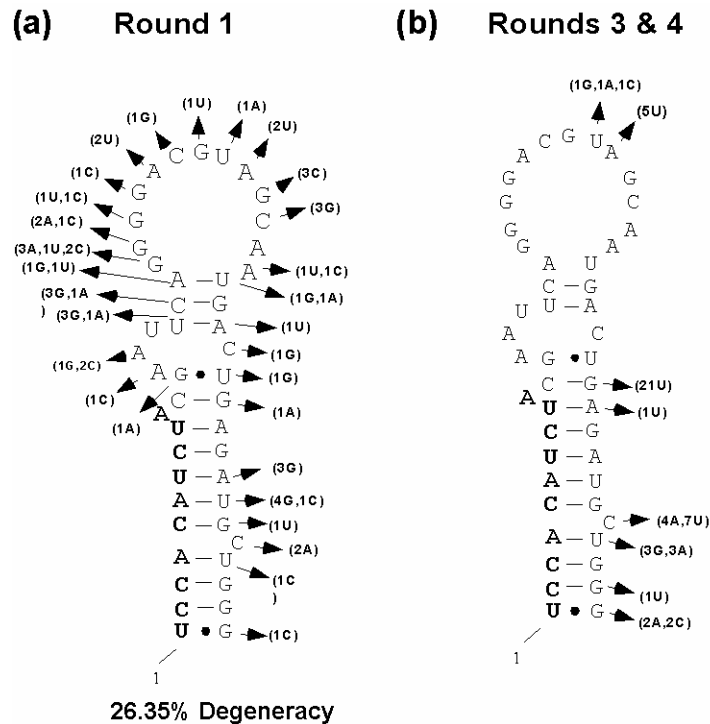


Figure 2.2. The sequence of the anti-RTA aptamer is nearly optimal for its structure. (a) Summary of mutations present following the first round of selection. The numbers of mutations are displayed in the context of the hypothetical secondary structure. Residues which were not doped in the pool are shown in bold. (b) Following several rounds of selection, the isolated sequences were near the wild-type. One mutation (G37U) was clearly preferred in the final populations.

After four cycles of selection and amplification the population bound RTA as well as the parental aptamer. Clones from the doped pool were sequenced in the early rounds of selection to ensure that the pool was sufficiently diverse (**Figure 2.1a**). Following several rounds of selection, many of the positions that contained mutations in the first round of selection had completely returned to their wild-type progenitors by the third round of selection (**Figure 2.2b**). Overall, the distribution of mutations within the variants suggests that mutations in the stem portion of the doped region were tolerated, whereas the loop region was much more highly conserved. These results were to some extent expected, since the loop region was derived completely from the random sequence

region, while the stem was established in part by pairing between the random sequence region and the 5' constant region.

Minimization of an anti-RTA aptamer

Neither the sequence of the selected aptamer nor its predicted secondary structure showed any gross similarity to 28S rRNA. In order to determine if more subtle similarities existed, we compared the aptamer to the SRL of 28S rRNA. The conformation of the SRL has been extensively studied (Correll et al., 1998; Munishkin and Wool, 1997; Szewczak et al., 1993) and both NMR and crystallography studies of small SRL mimics have indicated that it is essentially a continuous helix interrupted by a G-bulge and topped by a GAGA tetraloop. The sequence of the tetraloop has been shown to be critical for toxin recognition and binding (Chen et al., 1998). Following the detailed structural characterization of the ribosome, several groups noted that in its ribosomal context the SRL is aggressively protruded; this is not surprising as this loop is not only the site of ricin action but also a critical region for binding of translation factors (Ban et al., 2000).

The sequence required for ricin action is a conserved GAGA motif (Endo et al., 1991). The original aptamer contained a GAGA motif, but it is buried within a stem structure in the minimum free energy structure (**Figure 2.2b**). However, an alternative, less stable conformer was predicted to contain a putative GAGA recognition site in a loop (**Figure 2.3**). In order to ascertain whether the original (RA80.1) or alternative (RA80.2) conformer was the active ricin-binding species, several deletion variants were prepared. The deletion variants that were predicted to favor the most stable conformer were active; those that were predicted to favor the alternative conformer were inactive. The selection of a completely non-native, anti-RTA aptamer is not necessarily unusual; many aptamers bear little or no resemblance to corresponding wild-type RNA ligands (Giver et al., 1993a; Tuerk et al., 1992) with strikingly few exceptions (Cox et al., 2002). While ricin

is perhaps the most studied RIP-family protein, other RIPs possess similar toxicity towards the ribosome, and aptamer selections have been carried out against other RIP proteins (Hirao et al., 2000; Hirao et al., 1997). The RIP pepocin from *Curcubita pepo* is nearly as potent as ricin in terms of its ability to inactivate cellular ribosomes, though is unable to cross the cellular membrane (Yoshinari et al., 1996). As is the case with the anti-RTA aptamer, anti-pepocin aptamers bear little resemblance to ribosomal RNA or the SRL, although they are effective in inhibiting the action of pepocin on rat liver ribosomes (Hirao et al., 2000).

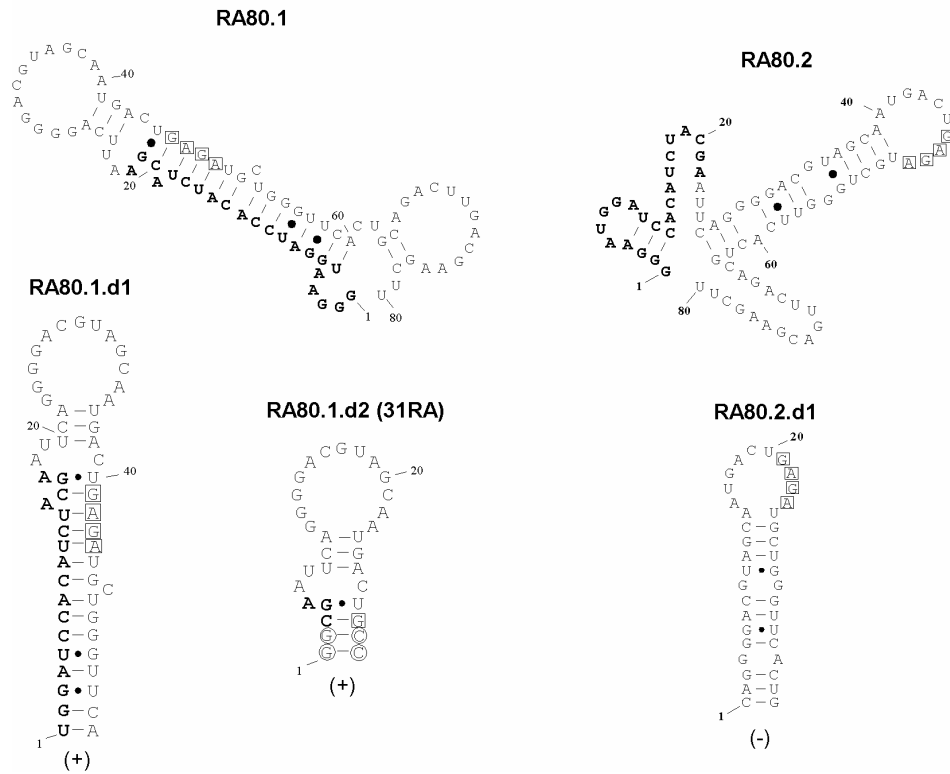


Figure 2.3. Comparison of anti-RTA aptamer conformers. RA80.1 is the originally predicted conformer; RA80.2 is an alternative conformer that presents a single-stranded GAGA sequence (boxed) that might be recognized by ricin. Deletion constructs that were predicted to enforce either the RA80.1 conformer (RA80.1.d1 and RA80.1.d2) or the RA80.2 conformer (RA80.2.d1) were synthesized and assayed for their ability to bind ricin. Residues in the constant regions are bolded, and those that differ from the original aptamer are circled. Binding is indicated by '+' (observed binding) and '-' (no observed binding).

Multiple mutations occurred in the stem region following mutagenesis and re-selection, indicating that many of the residues in the stem might not contribute to RTA-binding. To test whether the loop was primarily responsible for interactions with RTA, we prepared a short, 31 nucleotide RNA that contained most of the residues that were conserved on re-selection (31RA, **Figure 2.3**). The secondary structure of the minimal aptamer was stabilized by the inclusion of several G:C base-pairs that were not present in the selected aptamer. This minimized aptamer was fully competent to bind RTA. The fact that a functional, quadruple mutant (the two designed G:C pairs) could be designed based on the predicted aptamer secondary structure is further evidence in favor of our nascent structural model. Interestingly, the minimal aptamer lacked the GAGA sequence, but did contain a GGGG sequence in its loop. It is possible that this altered run of purines encourages tight interactions with RTA while avoiding the presentation of an adenosine substrate.

Anti-RTA aptamers bind RTA with high affinity

The interaction between anti-RTA aptamers and RTA was assessed as a function of protein concentration using a filter-binding assay similar to that used for selection. The RTA:RA80.1 complex was found to have a K_d of 240 nM while the RTA:31RA complex was found to have a K_d of 7.4 nM (**Figure 2.4**). Since the binding ability of the aptamer actually improved following truncation it can again be argued that sequences in the loop region are primarily responsible for RTA-recognition. Indeed, one of the reasons that the binding ability may have improved was that the shorter structure could no longer fold into alternate conformations (such as RA80.2).

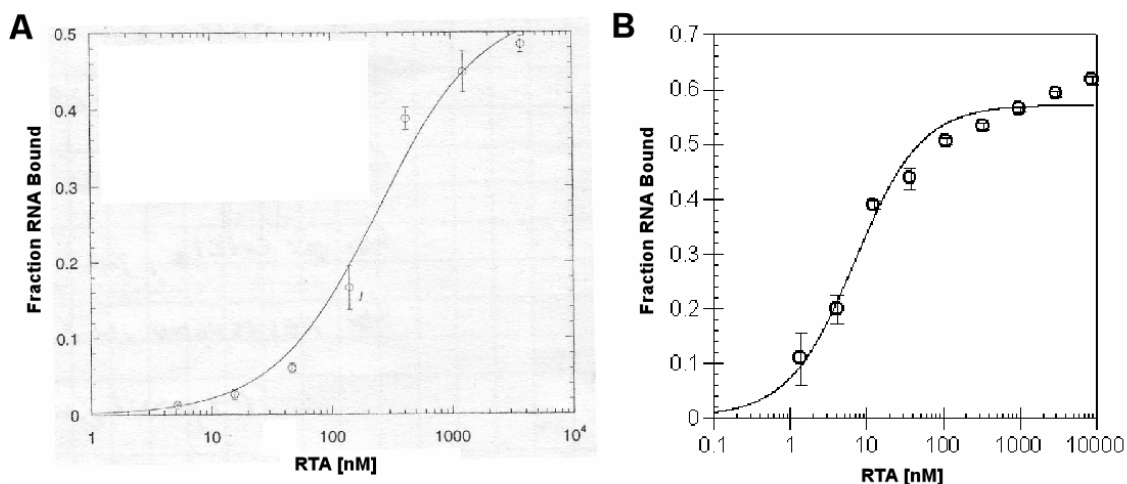


Figure 2.4. Binding isotherms for anti-RTA aptamers. Different concentrations of protein were mixed with a fixed concentration (2 nM) of either (A) 80RA or (B) 31RA. The binding reactions were allowed to come to equilibrium at room temperature over one hour, and RNA:protein complexes were separated by filtration. 100% binding is likely not observed at saturation because of inefficient capture of complexes. This phenomena has been previously been noted for other aptamer:protein complexes (Jellinek et al., 1994; Jellinek et al., 1993). The calculated dissociation constants (K_d) were 243 ± 47 nM for 80RA and 7.4 ± 1.1 nM for 31RA.

Although the affinity of the anti-RTA aptamer for RTA is reminiscent of the affinities of other anti-protein aptamers for their targets, (Gold et al., 1995) it is actually quite extraordinary considering that the SRL itself has a K_m of 13.55 μ M in a depurination assay (Endo et al., 1991). Even a small, SRL-like RNA containing a pre-designed transition-state analogue has been reported to inhibit RTA activity with a K_i of only 0.73 μ M (Chen et al., 1996). Molecular modeling studies have identified pteronic acid as an inhibitor of RTA activity, but the K_i for pteronic acid is around 0.6 mM (Robertus et al., 1996; Yan et al., 1997). The anti-RTA aptamer therefore binds at least three orders of magnitude better than the best-known small-molecule inhibitor of ricin, and two orders of magnitude better than a designed RNA analogue. The high affinity of the aptamer relative to other ligands exemplifies the dramatic possibilities afforded by *in vitro* selection.

Anti-RTA aptamers compete with the SRL but are not depurinated

In order to assess whether the anti-RTA aptamers bound to the active site of RTA, a competition assay was employed. For the assay, 31RA and 34SRL were incubated with a limiting amount of RTA. A portion of the reaction was filtered and RNAs retained on the filter membrane were recovered and separated by gel electrophoresis. As Figure 2.5 shows, 31RA completely displaced the 34SRL substrate from RTA. This competition assay can also be used to quantitate the relative degree of binding (Giver et al., 1993a). The specific activities of the samples prior to filtration were compared with the specific activities following filtration. Within the limits of the assay, 31RA binds RTA at least 9-fold better than 34SRL.

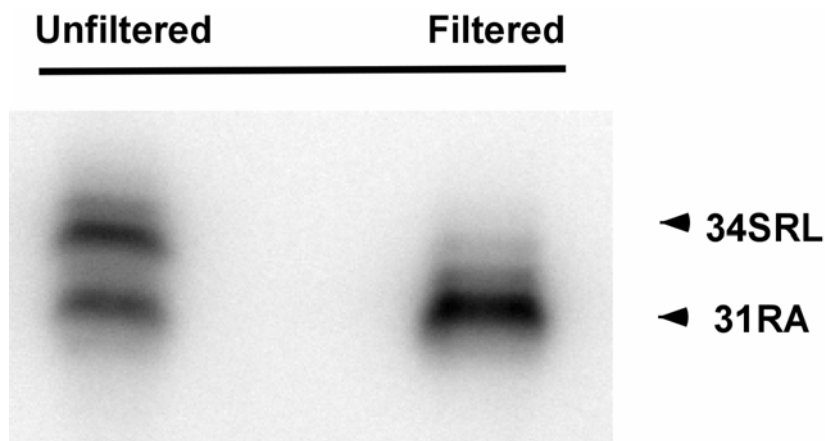


Figure 2.5. The anti-RTA aptamer competes with the SRL for the same site on RTA. Body-labeled 31RA and 34SRL were incubated in equal amounts with a limiting amount of RTA. A portion of the binding reaction was filtered over nitrocellulose, and filtered and unfiltered RNAs were precipitated and analyzed by electrophoresis. No detectable 34SRL is observed following filtration, and 31RA out-competes 34SRL by at least a factor of 9.

Although 31RA does not contain a canonical GAGA recognition sequence, it might still be depurinated (for example, at A32, A36, or even A39). To assess this possibility, we employed a depurination assay in which the accumulation of radiolabeled cleavage fragments is monitored (Endo et al., 1991). When 31RA was incubated with RTA and the same depurination assay used for the SRL was carried out no RTA-specific

radiolabeled cleavage products were observed (**Figure 2.6**). In contrast, when the SRL was incubated with RTA, a characteristic depurination and cleavage product was observed (Figure 2.6). When 31RA and the SRL were incubated together (5:1 31RA:SRL), the cleavage of SRL was inhibited, as would be expected if the two RNAs were binding to the same site on the enzyme. No inhibition of depurination was observed with a non-specific RNA molecule, tRNA. Because of the many steps associated with the depurination assay, it was not feasible to attempt a shorter time-course nor could a K_i value for the aptamer be derived. Overall, though, the data strongly support the nascent hypothesis originally suggested by the sequence and structure of 31RA: the aptamer binds to RTA, competes with the SRL, but is not itself a substrate for RTA.

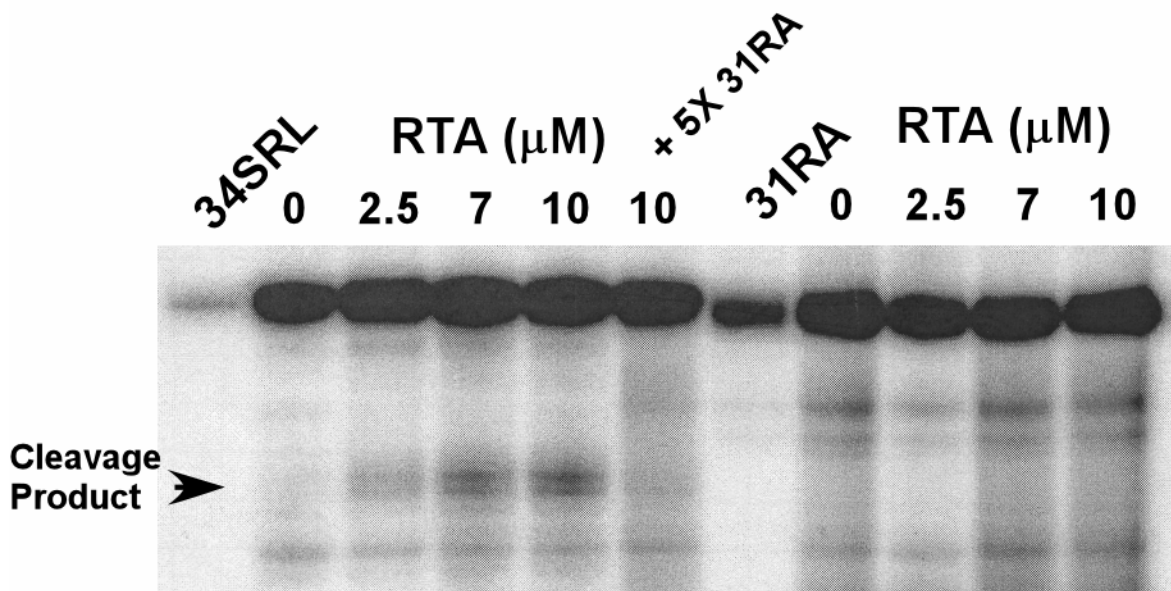


Figure 2.6. The anti-RTA aptamer inhibits depurination of the SRL by RTA. Different concentrations of RTA were mixed and incubated with a fixed amount of labeled substrate RNA for one hour. The resulting RNA was treated with aniline to cleave the SRL at depurinated, abasic sites. The products were then analyzed on a 20% gel and quantitated using a PhosphorImager. The expected cleavage product appears in a concentration dependent manner relative to RTA. The doubling of the band representing the cleavage product is likely due to the presence of untemplated, 3' additions of nucleotides to the full-length transcription product (Milligan et al., 1987). Competition with a five molar excess of unlabeled 31RA resulted in a 50% decrease in cleavage product. Upon incubation of 31RA with RTA no RTA-specific cleavage products were present.

31RA inhibits ribosomal inactivation by RTA

RTA activity can also be measured by observing the disruption of *in vitro* protein synthesis by *Artemia salina* ribosomes. This sensitive method has previously been used to assess the activity of other RTA inhibitors (Day et al., 1996). The inhibition of RTA activity was assayed as a function of 31RA concentration and Figure 2.7 shows that as the aptamer saturates RTA, the enzyme loses about 60% of its activity. The amount of 31RA that gives 50% of this maximal RTA inhibition (IC_{50}) was approximately 100 nM. It may be that the bound aptamer partially occludes the RTA active site, retarding activity about 60%; partial active site occlusion has previously been observed in aptamers that bind to protein tyrosine phosphatases (Bell et al., 1998). In the absence of RTA, the aptamer had no toxic effect on ribosomes. A non-cognate, anti-basic fibroblast growth factor aptamer, PS8 (Jhaveri et al., 1998) was used as a negative control. PS8 was not an inhibitor of RTA function and did not have an effect on *Artemia* ribosomes.

Figure 7.

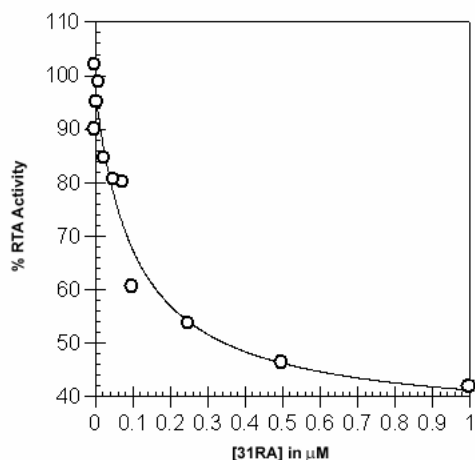


Figure 2.7. The anti-RTA aptamer inhibits RTA activity in an *in vitro* protein synthesis assay. The degradation of the translation capacity of *Artemia salinas* ribosomes was judged by determining the amount of radiolabeled protein produced. (Day et al., 1996) The minimal anti-RTA aptamer 31RA was added at varying concentrations. The % RTA inhibition was normalized to the amount of protein produced when no aptamer was added. Points are the average of triplicate experiments; a curve-fitting algorithm was used to derive the IC_{50} for 31RA, which was calculated to be 104 ± 25 nM.

The minimal aptamer may have protected translation more effectively than it protected the SRL from depurination because it was in greater excess over RTA in the translation assays. In addition, RTA acted on ribosomes for only 5 minutes in the *in vitro* translation assay, while it acted for 60 minutes on the isolated SRL. Nonetheless, the ability of the aptamer to inhibit the toxic effects of RTA at relatively low concentrations suggests that it may have promise as a scaffold for the development of gene or more conventional therapies for ricin exposure. For example, aptamers or their derivatives could be conjugated to pteric acid to produce a highly efficient chimeric inhibitor in which the aptamer binds tightly and the pteric acid completely inactivates RTA.

Anti-RTA aptamers represent a novel class of nucleic acid inhibitors

Aptamers have previously been shown to inhibit a variety of protein targets (Famulok and Jenne, 1998; Osborne et al., 1997; Rusconi et al., 2002). In general,

though, there was little or no possibility that previous aptamers could be acted on by their targets. To our knowledge, RTA is the first selection target that could catalyze the cleavage of individual species in a RNA pool.

It has been previously shown that short stems capped by GAGA tetraloops serve as substrates for RTA (Chen et al., 1998). Given these minimal requirements for interaction with RTA it is still somewhat surprising that the most bountiful anti-RTA aptamers were not also substrates. Minimal binding motifs frequently overrun *in vitro* selection experiments, a phenomenon that has been termed the ‘tyranny of small motifs’ (Jhaveri et al., 1997). However, any depurinated anti-RTA aptamers would have been inefficiently copied and would not have contained an adenosine in future generations, likely obviating the tyranny in these experiments. In addition, the selection of a novel sequence and structure is somewhat less startling when it is realized that the natural target of RTA is the ribosome. The SRL is surrounded by other RNA structures in the ribosome (Ban et al., 1999) and there are multiple amino acid residues in or adjacent to the RTA active site that could interact with these RNA structures (Mlsna et al., 1993). Differences between rRNA interactions with RTA and SRL interactions with RTA are further highlighted by the fact that rRNA is depurinated 77,000-fold faster than the SRL (Endo et al., 1991; Endo and Tsurugi, 1988; Mlsna et al., 1993). Since the relatively short RNA molecules present in the N30 pools lack the structural context of rRNA, very different sequences may be required to bind tightly to RTA.

The anti-RTA aptamer contrived to bind to the active site of the enzyme, but in a way that eluded depurination. RNA molecules that inhibit a RNA glycosidase are reminiscent of proteins that inhibit proteases (Laskowski and Kato, 1980). For example, even though bovine pancreatic trypsin inhibitor and soybean trypsin inhibitor are cleaved by trypsin, the inhibitor remains bound to the active site, allowing resynthesis. In addition, new evidence suggests that these inhibitors may bind in a ‘nonproductive’ mode that is substantially different than the way in which substrates are bound and that

preferentially stabilizes the ground state of the reaction (Coombs et al., 1999). Similarly, structural studies of bovine pancreatic trypsin inhibitor (BPTI) complexed with a thrombin mutant reveal that surface loops of the protease are rearranged, leading to inactivation (van de Loch et al., 1997). Using this co-crystal structure as a model, mutagenesis studies have shown that a lack of productive electrostatic contacts may also contribute to the ability of plasminogen activator inhibitor-1 to inhibit thrombin (Rezaie, 1998). Indeed, the ‘unnatural’ interaction of BPTI with a mutant thrombin may be especially germane to understanding 31RA function, in that it is possible that the ‘unnatural’ aptamer similarly induces unproductive structural transitions in RTA.

Conclusions

The development of novel diagnostic procedures for the purpose of detecting biologically relevant molecules, whether hazardous or not, is imperative. Because aptamers afford selective and high affinity binding, they are excellent candidates for incorporation into existing detection platforms, or can be used in novel diagnostic techniques. One powerful approach for detection using nucleic acids has been the use of microarray technology for the detection of complementary oligonucleotides, as in the assessment of cellular gene expression by the use of cDNA. Similarly, one can imagine employing aptamer sequences in a similar system that are used not for purposes of hybridization, but rather in a manner similar to protein antibodies.

EXPERIMENTAL PROCEDURES

Pools and oligonucleotides

The random sequence pool (N30) has been previously described (Bell et al., 1998; Conrad et al., 1995). Following chemical synthesis, the single-stranded DNA pool was purified on a 6% denaturing polyacrylamide gel and eluted by diffusion. The pool was

then amplified via the polymerase chain reaction using primers 41.30 (5'-GATAATACGACTCACTATAGGGAATGGATCCACATCTACG) and 24.30 (5'-AAGCTTCGTCAAGTCTGCAGTGAA). The selected aptamer 80RA was re-synthesized as a doped sequence pool (D30). The sequence of the D30 pool was 5'-GGGAATGGATCCACATCTACGAATTCAGGGGACGTAGCAATGACTGAGATGCTGGGTTCACTGCACTTGACGAAGCTT, where italicized residues indicate positions that were doped at 85% wild-type and 5% of each non-wild-type nucleotide. All other residues were kept constant. The 5' primer used for amplification was three nucleotides shorter from the 3' end than 41.30, and the 3' primer was 24.30.

A small version of the ribosomal sarcin-ricin loop (SRL) binds RTA and has previously been used as a substrate in depurination assays. (Endo et al., 1991) The sequence of the sarcin-ricin loop (34SRL) was 5'-GGAAUCCUGCUCAGUACGAGAGGAACCGCAGGUU; a double-stranded DNA template containing a T7 RNA polymerase promoter was used to transcribe 34SRL.

***In vitro* selection**

The RNA pool (2.2×10^{14} molecules) was transcribed from the amplified DNA template (1.1×10^{14} molecules) using an Ampliscribe T7 *in vitro* transcription kit (Epicentre Technologies, Madison, WI). Following purification on an 8% denaturing polyacrylamide gel, the RNA pellet was dissolved in 1 X PBS and 5 mM $MgCl_2$, then heated to 65 °C for 3 min and allowed to cool to room temperature over ten minutes. To exclude filter-binding RNA sequences from the pool, the RNA was passed over a 0.45- μ m HAWP filter (Millipore, Bedford, MA) and washed with an equal volume of buffer. RTA was added to the binding reaction (200 μ L final volume) in varying amounts throughout the course of the selection (Table I). Binding reaction mixtures were incubated for 1 hour at room temperature. After one hour, the solution was vacuum-filtered over a HAWP filter at 5 p.s.i., and washed twice with 0.5 ml of the selection

buffer. RNA retained on the filter was eluted twice with 0.2 ml of 7 M urea, 100 mM sodium citrate (pH 5.0) and 3 mM EDTA for 3 min at 100 °C, and the eluted RNA was precipitated with isopropanol. For the 8th and 9th rounds, the eluted RNA pool was passed over a HAWP filter following selection and precipitation to remove filter-binding species. Selected RNAs were reverse-transcribed using SuperScript II reverse transcriptase (Gibco BRL, Rockville, MD) and the 3' primer (24.30). The cDNA products were PCR amplified following addition of 5' primer (41.30) and DisplayTaq (PGC Scientific, Gaithersburg, MD). The PCR products were transcribed and gel-purified to begin the next round of selection.

Following round 9, the pool was cloned (Topo TA cloning kit, Invitrogen, Carlsbad, CA) and sequenced using standard dideoxy methods. The dominant clone (80RA) and its derivatives (31RA) were characterized using a nitrocellulose filtration assay similar to that used for selection.

The doped sequence selection was carried out using procedures identical to those used in the initial selection. A total of 1.84×10^{13} DNA molecules were used for transcription, and a total of 1.2×10^{14} RNA molecules were introduced at the first round of selection. Based on the rate of mutagenesis and the size of the pool, this should have represented all possible single to hextuple mutations. Table I summarizes selection conditions.

Competition Assay

31RA and 34SRL were body labeled in a transcription reaction containing [α - ^{32} P] UTP (3000 Ci/mmol, NEN Life Sciences). Following labeling, products were gel purified and precipitated. In competition assays, the minimal aptamer, 31RA and 34SRL RNAs were incubated at equimolar amounts (0.5 μ M) in the presence of limiting RTA (0.1 μ M) in the selection buffer. The binding reaction mixture was incubated at 25 °C for one hour and 80% of the volume was filtered over nitrocellulose and washed twice with

300 μ L selection buffer. Retained RNA molecules were eluted and precipitated. The remaining 20% of the reaction mixture was also precipitated. Both filtered and unfiltered samples were dissolved in 5 μ l of stop dye (7M urea, 1X TBE, 0.1% bromophenol blue) and analyzed on a denaturing 12% acrylamide gel. The amount of radioactivity in individual bands was quantitated using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). The relative binding activity was calculated based on the formula: $[(\text{counts filtered, 31RA}/\text{counts unfiltered, 31RA})] / [(\text{counts filtered, 34SRL}) / (\text{counts unfiltered, 34SRL})]$. This formula has previously been used to assess the relative binding activities of RNA ligands. (Giver et al., 1993b; Hirao et al., 1998) Assuming the binding reaction is at equilibrium (a likely assumption, given the length of the binding reaction), the binding ratio represents a ratio of the dissociation constants of individual RNA:protein complexes.

Filter Binding assays

Following rounds 4, 7 and 9 of the initial selection, and at every round of the doped selection, the RNA pool was body labeled in a transcription reaction using [α -³²P] UTP (3000 Ci/mmol, NEN Life Sciences). Products were gel purified and precipitated. The labeled RNA (0.5 μ M final concentration) was incubated with RTA (0.5 μ M final concentration) in 0.1 ml selection buffer for 1 hour at room temperature. The solution was filtered and washed three times with selection buffer. The filter was exposed to a PhosphorImager plate and the amount of retained radioactivity was determined. The fraction of the pool or clone that bound RTA was calculated by comparing the counts retained on the filter with the total number of counts in the original RNA.

For the determination of dissociation constants, 1.5 pmol RNA was end-labeled with [γ -³²P] ATP (0.03 mCi, 6000 Ci/mmol, ICN Biomedicals, Costa Mesa, CA) following 5'-dephosphorylation. Labeled RNA was purified by two successive precipitations using ammonium acetate. The labeled RNA (2 nM final concentration)

was incubated with increasing concentrations of RTA (1.2 – 450 nM) in 50 μ l of the selection buffer for 1 hour at room temperature. The mixture was filtered on a vacuum manifold (Schleicher & Schuell, Keene, NH) equipped with a piece of pure nitrocellulose membrane (Schleicher & Schuell) over a piece of Hybond (Amersham Pharmacia Biotech, Piscataway, NJ) as previously described (Conrad et al., 1995; Weeks and Cech, 1995). The radioactivity on the nitrocellulose and nylon filter was quantitated, and the dissociation constant (K_d) for aptamer:RTA complexes was determined using a curve-fitting algorithm on the program Kaleidagraph (Adelbeck Software, Reading, PA).

Assay for RNA Depurination

The assay was performed as previously described. (Endo et al., 1991) RNA (2 μ M) was denatured at 65 °C in 1 X PBS, 5 mM $MgCl_2$ and allowed to cool to room temperature for ten minutes. DTT and EDTA were added to 1 mM final concentration (20 μ L final volume) and reaction mixtures were incubated with varying concentrations of RTA for 60 minutes at 30 °C. After a phenol/ $CHCl_3$ extraction and ethanol precipitation, pellets were resuspended in 20 μ l of a buffered aniline solution and incubated for 15 minutes on ice. Following a second precipitation, RNA molecules were 5' end labeled using T4 polynucleotide kinase and [γ -32P] ATP (0.03 mCi, 6000 Ci/mmol, ICN Biomedicals, Costa Mesa, CA) in a 5 μ L reaction volume. The RNA was precipitated and resuspended in loading dye, and products were analyzed on a 20% denaturing polyacrylamide gel. The gel was fixed and counts were quantitated using a PhosphorImager.

Protein Synthesis Inhibition Assay

Recombinant RTA was prepared as previously described (Ready et al., 1991). A standard assay for RTA function is to observe the degradation of translation ability by isolated *Artemia salina* ribosomes (Day et al., 1996). When RTA (2 nM) was added to

an *A. salinas* *in vitro* translation system, the synthesis of ¹⁴C-labeled poly-phenylalanine was reduced by 80%. Using this basal assay as our standard, different concentrations of 31RA were then assayed for their ability to inhibit RTA degradation of translation ability. Prior to use, solutions containing 31RA were heated at 65 °C for 3 min, and allowed to cool at 25°C for 10 minutes.

REFERENCES

- Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P.B. and Steitz, T.A. (1999) Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature*, **400**, 841-847.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905-920.
- Bartel, D.P., Zapp, M.L., Green, M.R. and Szostak, J.W. (1991) HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell*, **67**, 529-536.
- Bell, S.D., Denu, J.M., Dixon, J.E. and Ellington, A.D. (1998) RNA molecules that bind to and inhibit the active site of a tyrosine phosphatase. *J Biol Chem*, **273**, 14309-14314.
- Brody, E.N. and Gold, L. (2000) Aptamers as therapeutic and diagnostic agents. *J Biotechnol*, **74**, 5-13.
- Burgstaller, P., Girod, A. and Blind, M. (2002) Aptamers as tools for target prioritization and lead identification. *Drug Discov Today*, **7**, 1221-1228.
- Chen, X.Y., Link, T.M. and Schramm, V.L. (1998) Ricin A-chain: kinetics, mechanism, and RNA stem-loop inhibitors. *Biochemistry*, **37**, 11605-11613.
- Chen, X.-Y., Link, T.M. and Schramm, V.L. (1996) Inhibition of Ricin by an RNA Stem-Loop Containing a Ribo-Oxocarbenium Mimic. *Journal of the American Chemical Society*, **118**, 3067-3068.
- Conrad, R.C., Baskerville, S. and Ellington, A.D. (1995) *In vitro* selection methodologies to probe RNA function and structure. *Mol Divers*, **1**, 69-78.
- Coombs, G.S., Rao, M.S., Olson, A.J., Dawson, P.E. and Madison, E.L. (1999) Revisiting catalysis by chymotrypsin family serine proteases using peptide substrates and inhibitors with unnatural main chains. *J Biol Chem*, **274**, 24074-24079.

- Correll, C.C., Munishkin, A., Chan, Y.L., Ren, Z., Wool, I.G. and Steitz, T.A. (1998) Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc Natl Acad Sci U S A*, **95**, 13436-13441.
- Cox, J.C., Hayhurst, A., Hesselberth, J., Bayer, T.S., Georgiou, G. and Ellington, A.D. (2002) Automated selection of aptamers against protein targets translated *in vitro*: from gene to aptamer. *Nucleic Acids Res*, **30**, e108.
- Day, P.J., Ernst, S.R., Frankel, A.E., Monzingo, A.F., Pascal, J.M., Molina-Svinth, M.C. and Robertus, J.D. (1996) Structure and activity of an active site substitution of ricin A chain. *Biochemistry*, **35**, 11098-11103.
- Endo, Y., Gluck, A. and Wool, I.G. (1991) Ribosomal RNA identity elements for ricin A-chain recognition and catalysis. *J Mol Biol*, **221**, 193-207.
- Endo, Y., Mitsui, K., Motizuki, M. and Tsurugi, K. (1987) The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28 S ribosomal RNA caused by the toxins. *J Biol Chem*, **262**, 5908-5912.
- Endo, Y. and Tsurugi, K. (1987) RNA N-glycosidase activity of ricin A-chain. Mechanism of action of the toxic lectin ricin on eukaryotic ribosomes. *J Biol Chem*, **262**, 8128-8130.
- Endo, Y. and Tsurugi, K. (1988) The RNA N-glycosidase activity of ricin A-chain. *Nucleic Acids Symp Ser*, 139-142.
- Famulok, M. (1999) Oligonucleotide aptamers that recognize small molecules. *Curr Opin Struct Biol*, **9**, 324-329.
- Famulok, M. and Jenne, A. (1998) Oligonucleotide libraries--variatio delectat. *Curr Opin Chem Biol*, **2**, 320-327.
- Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gustafsdottir, S.M., Ostman, A. and Landegren, U. (2002) Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol*, **20**, 473-477.
- Giver, L., Bartel, D., Zapp, M., Pawul, A., Green, M. and Ellington, A.D. (1993a) Selective optimization of the Rev-binding element of HIV-1. *Nucleic Acids Res*, **21**, 5509-5516.
- Giver, L., Bartel, D.P., Zapp, M.L., Green, M.R. and Ellington, A.D. (1993b) Selection and design of high-affinity RNA ligands for HIV-1 Rev. *Gene*, **137**, 19-24.
- Gold, L., Polisky, B., Uhlenbeck, O. and Yarus, M. (1995) Diversity of oligonucleotide functions. *Annu Rev Biochem*, **64**, 763-797.
- Hesselberth, J.R., Miller, D., Robertus, J. and Ellington, A.D. (2000) *In vitro* selection of RNA molecules that inhibit the activity of ricin A-chain. *J Biol Chem*, **275**, 4937-4942.

- Hirao, I., Madin, K., Endo, Y., Yokoyama, S. and Ellington, A.D. (2000) RNA aptamers that bind to and inhibit the ribosome-inactivating protein, pepocin. *J Biol Chem*, **275**, 4943-4948.
- Hirao, I., Spingola, M., Peabody, D. and Ellington, A.D. (1998) The limits of specificity: an experimental analysis with RNA aptamers to MS2 coat protein variants. *Mol Divers*, **4**, 75-89.
- Hirao, I., Yoshinari, S., Yokoyama, S., Endo, Y. and Ellington, A.D. (1997) *In vitro* selection of aptamers that bind to ribosome-inactivating toxins. *Nucleic Acids Symp Ser*, 283-284.
- Homann, M. and Goring, H.U. (1999) Combinatorial selection of high affinity RNA ligands to live African trypanosomes. *Nucleic Acids Res*, **27**, 2006-2014.
- Centers for Disease Control and Prevention (<http://www.bt.cdc.gov/Agent/agentlist.asp>)
- Jellinek, D., Green, L.S., Bell, C. and Janjic, N. (1994) Inhibition of receptor binding by high-affinity RNA ligands to vascular endothelial growth factor. *Biochemistry*, **33**, 10450-10456.
- Jellinek, D., Lynott, C.K., Rifkin, D.B. and Janjic, N. (1993) High-affinity RNA ligands to basic fibroblast growth factor inhibit receptor binding. *Proc Natl Acad Sci U S A*, **90**, 11227-11231.
- Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425-1429.
- Jhaveri, S., Olwin, B. and Ellington, A.D. (1998) *In vitro* selection of phosphorothiolated aptamers. *Bioorganic & Medicinal Chemistry Letters*, **8**, 2285-2290.
- Jhaveri, S., Rajendran, M. and Ellington, A.D. (2000a) *In vitro* selection of signaling aptamers. *Nat Biotechnol*, **18**, 1293-1297.
- Jhaveri, S.D., Hirao, I., Bell, S., Uphoff, K.W. and Ellington, A.D. (1997) Landscapes for molecular evolution: lessons from *in vitro* selection experiments with nucleic acids. *Annual Reports in Combinatorial Chemistry and Molecular Diversity*, **1**, 169-191.
- Jhaveri, S.D., Kirby, R., Conrad, R., Maglott, E.J., Bowser, M., Kennedy, R.T., Glick, G. and Ellington, A.D. (2000b) Designed Signaling Aptamers that Transduce Molecular Recognition to Changes in Fluorescence Intensity. *Journal of the American Chemical Society*, **122**, 2469-2473.
- Larsson, S.L., Sloma, M.S. and Nygard, O. (2002) Conformational changes in the structure of domains II and V of 28S rRNA in ribosomes treated with the translational inhibitors ricin or alpha-sarcin. *Biochim Biophys Acta*, **1577**, 53-62.
- Laskowski, M., Jr. and Kato, I. (1980) Protein inhibitors of proteinases. *Annu Rev Biochem*, **49**, 593-626.

- Lord, J.M., Hartley, M.R. and Roberts, L.M. (1991) Ribosome inactivating proteins of plants. *Semin Cell Biol*, **2**, 15-22.
- Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res*, **15**, 8783-8798.
- Mlsna, D., Monzingo, A.F., Katzin, B.J., Ernst, S. and Robertus, J.D. (1993) Structure of recombinant ricin A chain at 2.3 Å. *Protein Sci*, **2**, 429-435.
- Munishkin, A. and Wool, I.G. (1997) The ribosome-in-pieces: binding of elongation factor EF-G to oligoribonucleotides that mimic the sarcin/ricin and thiostrepton domains of 23S ribosomal RNA. *Proc Natl Acad Sci U S A*, **94**, 12280-12284.
- Osborne, S.E., Matsumura, I. and Ellington, A.D. (1997) Aptamers as therapeutic and diagnostic reagents: problems and prospects. *Curr Opin Chem Biol*, **1**, 5-9.
- Ready, M.P., Kim, Y. and Robertus, J.D. (1991) Site-directed mutagenesis of ricin A-chain and implications for the mechanism of action. *Proteins*, **10**, 270-278.
- Rezaie, A.R. (1998) Elucidation of the structural basis for the slow reactivity of thrombin with plasminogen activator inhibitor-1. *Biochemistry*, **37**, 13138-13142.
- Robertus, J.D., Yan, X., Ernst, S., Monzingo, A., Worley, S., Day, P., Hollis, T. and Svinth, M. (1996) Structural analysis of ricin and implications for inhibitor design. *Toxicon*, **34**, 1325-1334.
- Rusconi, C.P., Scardino, E., Layzer, J., Pitoc, G.A., Ortel, T.L., Monroe, D. and Sullenger, B.A. (2002) RNA aptamers as reversible antagonists of coagulation factor IXa. *Nature*, **419**, 90-94.
- Schlutzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. and Yonath, A. (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, **102**, 615-623.
- Symensma, T.L., Giver, L., Zapp, M., Takle, G.B. and Ellington, A.D. (1996) RNA aptamers selected to bind human immunodeficiency virus type 1 Rev *in vitro* are Rev responsive *in vivo*. *J Virol*, **70**, 179-187.
- Szewczak, A.A., Moore, P.B., Chang, Y.L. and Wool, I.G. (1993) The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc Natl Acad Sci U S A*, **90**, 9581-9585.
- Tuerk, C., MacDougall, S. and Gold, L. (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci U S A*, **89**, 6988-6992.
- van de Locht, A., Bode, W., Huber, R., Le Bonniec, B.F., Stone, S.R., Esmon, C.T. and Stubbs, M.T. (1997) The thrombin E192Q-BPTI complex reveals gross structural

- rearrangements: implications for the interaction with antithrombin and thrombomodulin. *Embo J*, **16**, 2977-2984.
- Weeks, K.M. and Cech, T.R. (1995) Efficient protein-facilitated splicing of the yeast mitochondrial b15 intron. *Biochemistry*, **34**, 7728-7738.
- Yan, X., Hollis, T., Svinth, M., Day, P., Monzingo, A.F., Milne, G.W. and Robertus, J.D. (1997) Structure-based identification of a ricin inhibitor. *J Mol Biol*, **266**, 1043-1049.
- Yoshinari, S., Yokota, S., Sawamoto, H., Koresawa, S., Tamura, M. and Endo, Y. (1996) Purification, characterization and subcellular localization of a type-1 ribosome-inactivating protein from the sarcocarp of *Cucurbita pepo*. *Eur J Biochem*, **242**, 585-591.
- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. and Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883-896.
- Zuker, M. (2000) Calculating nucleic acid secondary structure. *Curr Opin Struct Biol*, **10**, 303-310.

Chapter 3: An aptazyme ligase array for the detection of diverse analyte classes

INTRODUCTION

Biosensor design

The development of biosensors is an important venture which could have broad impact in arenas ranging from medicine to warfare defense. Many types of sensor platforms have been developed using various sensing components and detection techniques. Each of the current implementations has their own unique advantages and disadvantages, but all biosensors are ultimately similar in spirit: biologically derived or inspired components are used to perform the two core functions of a sensor, event acquisition and signal generation. In addition, researchers in the biosensor field are also concerned with issues such as a given sensor's utility and usability, as well as its limits of detection and specificity of response to a given set of analytes. **Table 3.1** summarizes the issues relevant to the construction and use of biosensors.

Event Acquisition		
<i><u>Biopolymer</u></i>	<i><u>Application</u></i>	<i><u>Reference</u></i>
DNA	Microarray	(DeRisi et al., 1997)
RNA	Aptamer / Ribozyme	(Hesselberth et al., 2003; Seetharaman et al., 2001)
Proteins	ELISA, proteomics	(de Wildt et al., 2000)
Metabolites	Small-molecule (drug) screening	(Winssinger et al., 2002)
Signal Generation		
<i><u>Type</u></i>	<i><u>Application</u></i>	<i><u>Reference</u></i>
Optical	Fluorescence	(de Wildt et al., 2000; DeRisi et al., 1997; Winssinger et al., 2002)
Electrical	SPR, FET	(Fritz et al., 2002; McDonnell, 2001)
Mechanical	Cantilever	(Fritz et al., 2000)

Table 3.1. A summary of available techniques for the construction of biosensors.

While the detection of an event is useful information, the ability to monitor several different events at once can be very powerful. Thus, the creation of arrays of biopolymers each capable of monitoring different events have been developed. Existing array technologies have proven extraordinarily useful for the detection of a wide variety of analytes. Spotting of cDNAs on glass slides has enabled the measurement of cellular expression levels (DeRisi et al., 1997), and the use of antibody arrays has enabled the assessment of proteins directly and in parallel (de Wildt et al., 2000). More recently, the use of small-molecule arrays was employed to examine the activity and specificity of cellular proteases, and enabled the identification of a novel caspase inhibitor (Winssinger et al., 2002). One of the first examples using RNA aptamers in array technology employed protein specific aptamers in an ELISA-like system utilizing a sandwich assay for detection (Drolet et al., 1996). In addition, RNA aptamers specific for several small molecule metabolites have also been incorporated into arrays for cellular diagnostic

purposes (Brody et al., 1999). One of the most impressive demonstrations of detection using aptamers was a strategy in which a ligand brings to chimeric aptamer molecules close together, and they are subsequently ligated together to form a molecule suitable for amplification using PCR. Using a pair of PDGF specific DNA aptamers, the authors report detecting attomoles of protein (Fredriksson et al., 2002).

Allosteric ribozymes as biosensor elements

Ribozymes have been shown to be very much like their enzymatic protein counterparts in that they are capable of allostery. It is possible to convert an existing ribozyme into one whose activity is dependent upon an external factor. These allosteric ribozymes (or ‘aptazymes’) are typically made by either fusion with a pre-existing aptamer domain or by selection from naïve pools that preserve the core of the ribozyme. Several examples of these aptazyme classes now exist. One of the first examples was a hammerhead ribozyme (a self-cleaving RNA) whose activity was dependent upon an external oligonucleotide (Porta and Lizardi, 1995). Following this example, hammerhead variants that incorporated an ATP aptamer domain were shown to be regulated by ATP (Tang and Breaker, 1997). *In vitro* selection was then employed for the identification of allosteric ribozymes from pools in which a stem joining ribozyme and aptamer domains was randomized. Variants of the hammerhead motif derived from this pool were dependent upon FMN, ATP and theophylline (Soukup and Breaker, 1999). Since those initial examples, hammerhead aptazymes have been identified that respond up to ~5000-fold to a wide variety of small-molecules (e.g. cGMP) as well as protein and peptide analytes (Hartig et al., 2002; Koizumi et al., 1999; Soukup and Breaker, 1999; Vaish et al., 2002; Wang and Sen, 2002).

Previously our laboratory developed aptazymes based upon a ribozyme ligase that was initially selected from a pool of random sequences. This ribozyme, termed the L1 ligase, catalyzes the self-attachment of a short oligonucleotide substrate at rates

approaching 1 min^{-1} (Robertson and Ellington, 1999). It was originally discovered that the ligase is dependent upon the 3' primer used during the selection, as its activity increases 10,000 fold in its presence. Allosteric variants of the ligase identified through *in vitro* selection are activated by 350- to 1500- fold in the presence of small molecules, and variants identified after selection against peptide (Rev) and protein (lysozyme, Cyt-18) targets have also been shown to be highly active in the presence of cognate analyte (19,000, 3,100 and $\sim 100,000$ fold, respectively) (Robertson and Ellington, 1999; Robertson and Ellington, 2000; Robertson and Ellington, 2001). This set of aptazymes was particularly appropriate for incorporation into a biosensor, as they could detect a variety of ligands in solution with large concomitant changes in their reaction rates. **Table 3.2** summarizes the characteristics of the set of aptzyme ligases used in the array.

Aptzyme	Ligand	Ligand M.W. (g / mol)	Rate (k_{active} , k_{inactive}) (hr^{-1})	Activation (k_{on} / k_{off})	Reference
L1	18.90A	5822.6	0.71, ~ 0.00007	$\sim 10,000$	(Robertson and Ellington, 1999)
L1-ATP	ATP	551.1	0.26, 0.0017	150	(Robertson and Ellington, 2000)
L1-FMN	FMN	478.3	0.43, 0.002	260	(Robertson and Ellington, 2000)
L1-lysozyme	lysozyme	14,400	0.6, 0.0002	3,100	(Robertson and Ellington, 2001)
L1-Rev	Rev peptide	2437.8	0.86, 0.000046	19,000	(Robertson, 2001)
L1-theophylline	theophylline	180.2	0.52, 0.002	350	(Robertson and Ellington, 2000)

Table 3.2. Summary of aptzyme characteristics relevant to biosensor design

RESULTS AND DISCUSSION

Construction of an aptazyme ligase array.

In order to adapt these aptazymes to a sensing platform, a scheme was developed in which the ligand-dependent ligation of a biotinylated substrate to a ribozyme resulted in the immobilization of the product on streptavidin coated, 96-well polystyrene plates (**Figure 3.1**) (Hesselberth et al., 2003). The levels of radiolabeled ribozyme could then be monitored as a function of a variety of conditions, such as incubation time and analyte class.

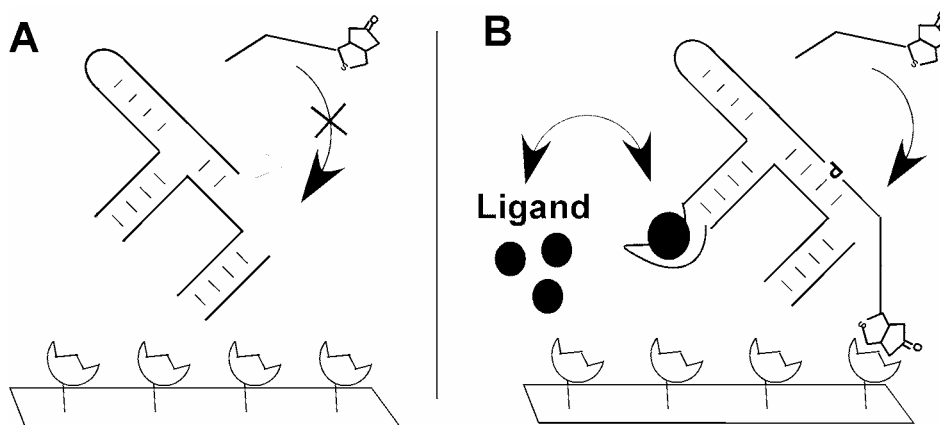


Figure 3.1. Schematic of an aptazyme ligase array. (A) In the absence of analyte, the ribozyme is unable to catalyze the ligation of a biotinylated substrate and remains in the supernatant. (B) The presence of analyte activates the aptazyme, enabling the self-attachment of the substrate which results in immobilization of the product on a streptavidin coated surface.

By employing the above strategy, we showed that multiple classes of analytes were able to be detected on a single platform. By incubating radiolabeled ribozymes (rows) with cognate and non-cognate analytes (columns) (Figure 3.2), the ability of a given aptazyme to detect and report on its analyte partner was borne out by the diagonal line of spots indicating identity between a ribozyme-analyte pair. In addition, two control reactions were performed to assess the background level of response (‘-’ column) and the response in a complex mixture (‘+’ column, all analytes mixed). These controls indicated that ribozymes in the context of the array behaved similarly to those in solution in that

they were inactive without analyte, and could also sense the presence of their analyte partner while in a reasonably complex mixture.

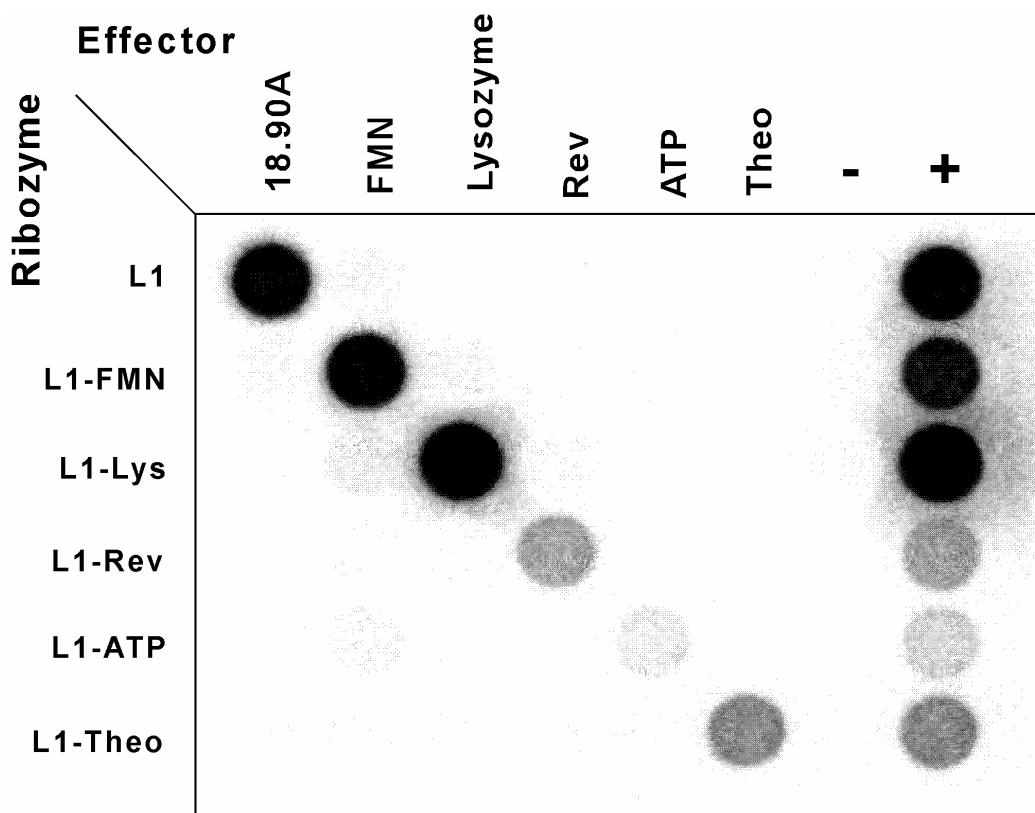


Figure 3.2. Aptazyme ligase array. Aptazymes and analytes were mixed in various combinations in a 96-well plate. The diagonal line of wells represents a positive reaction between an aptazyme and its appropriate analyte partner. The ‘-’ column represents the absence of analytes, and the ‘+’ column includes a mixture of all the analytes. Individual concentrations of analytes used in the array are included in the Experimental Procedures.

Response characteristics of the aptazyme array

The aptazymes we employed for the biosensor have inherent affinities for their corresponding analytes, and thus their activities are directly influenced by the concentration of those analytes. Previously we had shown that L1 aptazymes selected against small-molecules had affinities for their ligands in the low micromolar range (Robertson and Ellington, 2000). The values were also consistent with previously calculated values for the aptamer domains alone (Dieckmann et al., 1996; Jenison et al., 1994; Sassanfar and Szostak, 1993). L1 aptazymes selected against the protein lysozyme

also possessed affinities in the low micromolar range (Robertson and Ellington, 2001) and it was shown that affinity values calculated using either catalytic rate or nitrocellulose membrane capture were coincident, implying that the binding event triggers catalysis. To address whether these numbers could be similarly evaluated using the array, individual aptazymes were incubated with increasing concentrations of their ligand partners, and the levels of reaction observed in individual wells were plotted versus the analyte concentration. Three aptazymes in particular, L1-Rev, L1-lysozyme and L1-FMN demonstrated K_d s for their analytes in the micromolar range; these values are consistent with those previously reported based on catalytic activity (Robertson and Ellington, 2000; Robertson and Ellington, 2001) (**Figure 3.3**).

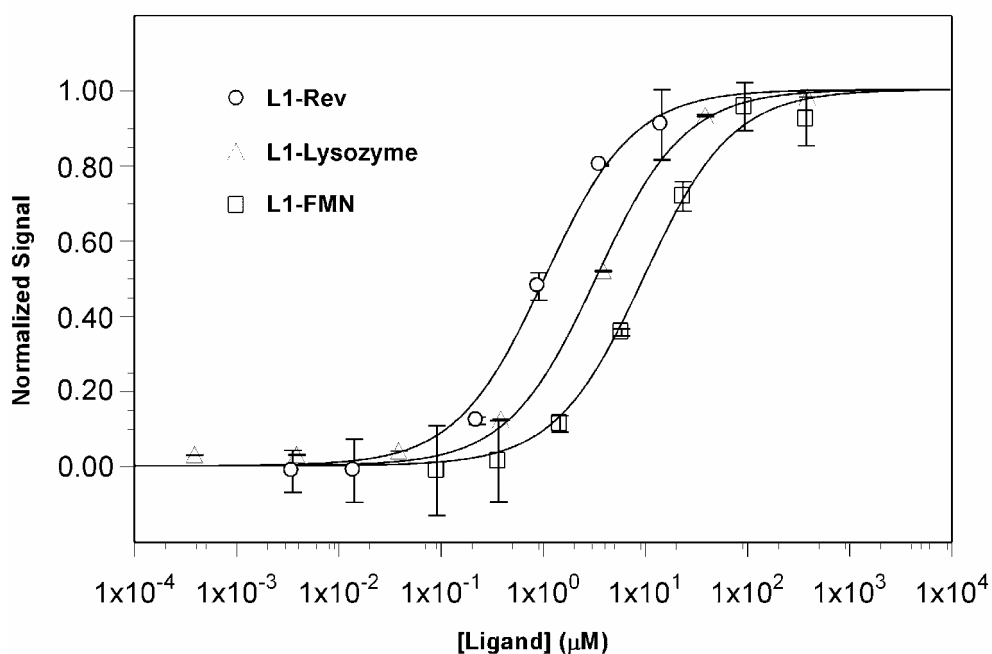


Figure 3.3. Response curves for arrayed aptazyme ligases. Signals from the array were determined as a function of analyte concentration for three different aptazymes (L1-Rev, L1-Lysozyme and L1-FMN). Signals generated by incubation in the absence of analyte were subtracted. The signals obtained from a given reaction were normalized relative to the highest extent observed for a given aptazyme. All values are the average of three determinations, and error bars are shown.

Sensitivity of the aptazyme array

The issue of sensitivity inevitably arises when employing a sensor to detect the presence of an analyte. The reproducible, sensitive detection of molecules is the chief responsibility of a biosensor, and thus is a measure by which different sensing platforms can be compared. We were therefore interested in determining the limits of detection for the aptazyme biosensor, and how those limits compared with previously constructed biosensors designed to detect and quantitate similar ligands. In order to experimentally address the issue of sensitivity, we incubated decreasing levels of analyte with individual aptazymes. In addition, because the array is able to detect several classes of analyte, we wanted to demonstrate that a single platform was capable of addressing all of the classes at once both sensitively and reproducibly. As shown in Figure 3.4, several analytes could be detected in the nanomolar concentration range (femtomoles of analyte). Of particular note, we show that proteins and peptides can be detected in the low nanomolar range; a level at which detection of cellular metabolites becomes feasible.

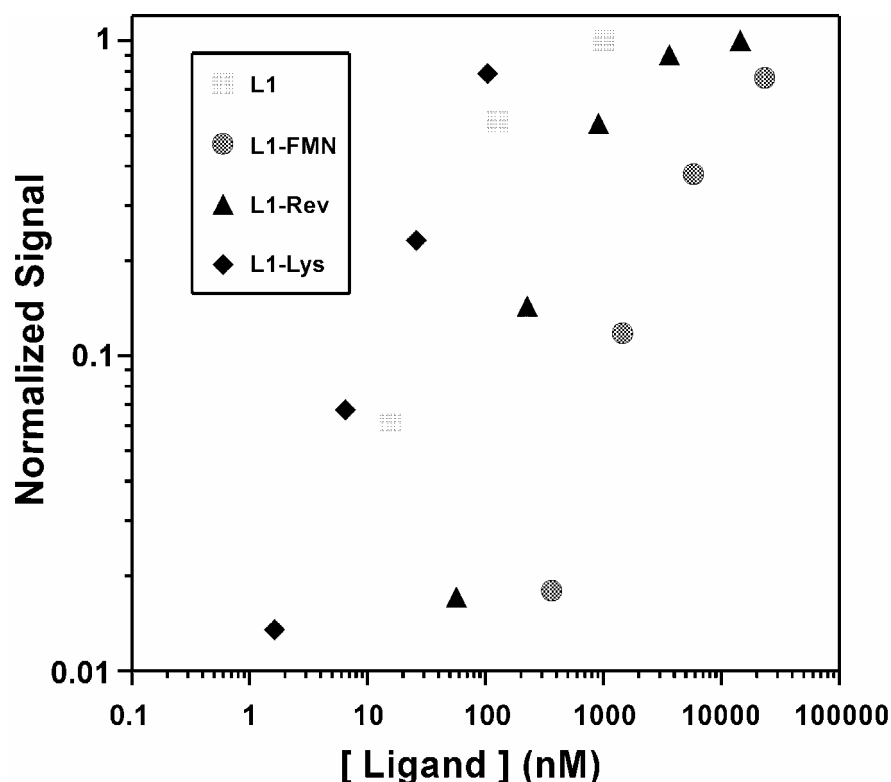


Figure 3.4. Limits of detection determined using the aptazyme array. Signals from the array were determined as a function of analyte concentration. Signals observed in the absence of analyte were subtracted, and all signals were normalized relative to the highest extent of reaction seen for a given aptazyme. A log-log plot is used to illustrate the relationship between the dynamic range of an aptazyme and its detection limit. Values shown are the average of two or three determinations, and errors were typically less than 5%.

Previously, arrays have been constructed with similar goals: quantitative detection of multiple classes of analyte. In particular, Fran Ligler and her colleagues have shown that multiple types of biological warfare agents can be detected in parallel using specific antibodies (Rowe et al., 1999; Rowe-Taitt et al., 2000a; Rowe-Taitt et al., 2000b). For example, the protein ricin was detected at a level of 40 ng / mL (~ 1 nM) (Rowe-Taitt et al., 2000a), a finding consistent with results obtained quantitating other proteins using the ligase array. A detection system was also developed utilizing rolling-circle amplification in order to detect a set of 75 cytokines using an antibody microarray (Schweitzer et al., 2002).

Improvement of the signal-to-noise ratio

Another important characteristic of a biosensor is its ability to distinguish noise from signal. Because our biosensor incorporated the formation of a covalent bond between our product and a solid-phase, we hypothesized that an increase in washing stringency would lead to an improvement in the signal-to-noise ratio afforded by the array. Experimentally we showed that after washing with ~6 milliliters of buffer (~20 buffer exchanges) resulted in a two fold increase in the signal observed (**Figure 3.5**).

Previously a similar system employing hammerhead aptazymes was used to sense the presence of analytes by monitoring the self-catalyzed release of aptazymes from a surface. The resulting signal was effectively the loss of radiolabel from a given ‘pixel’ on the array. By exploiting the covalent attachment of an aptazyme in response to an analyte, we have circumvented the problems and artifacts that could potentially arise during stringent washing of such a ‘loss of signal’ based platform. Although the increase in signal-to-noise ratio we obtained was modest, the ability to perform thorough washing to reduce background signals is an advantage not available to a platform using radiolabeled hammerhead aptazymes (or cleaving ribozymes in general) as any washing inevitably reduces the signal observed. That said, schemes could easily be imagined in which fluorophore-quencher pairs or FRET dyes are appropriately placed on cleaving ribozymes in order to produce the desired ‘signal-from-loss’ scenario, and has been put into practice using protein- and peptide-responsive hammerhead aptazymes (Hartig et al., 2002; Jenne et al., 2001).

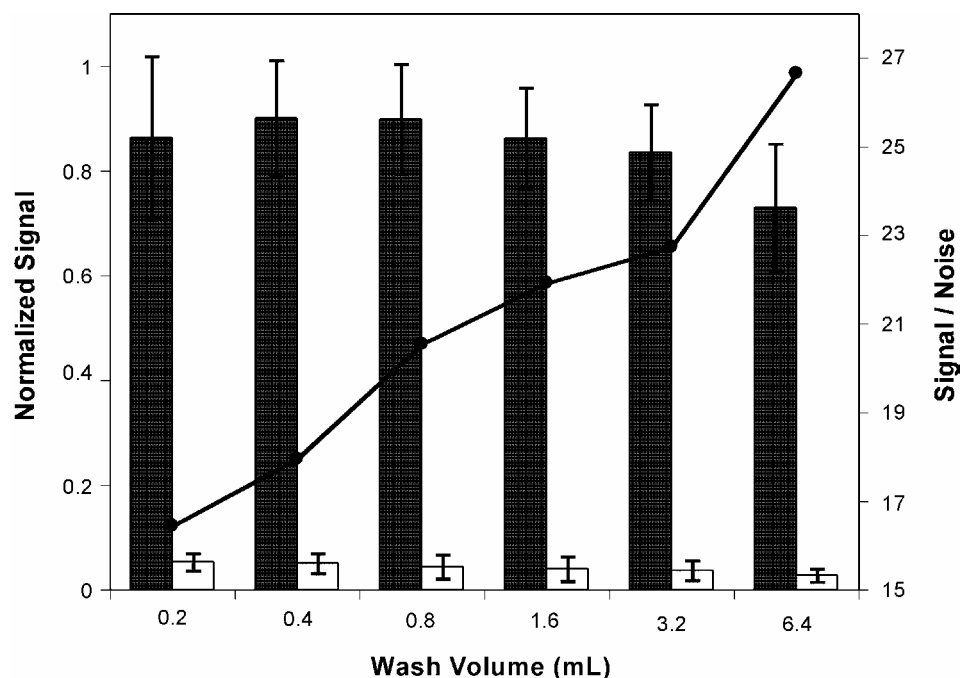


Figure 3.5. Thorough washing leads to an increase in signal-to-noise ratio. A standard assay was carried out using the L1-lysozyme aptazyme. The volume of wash buffer was progressively incremented for each reaction. Samples were initially washed with 200 μ L of buffer, and additional aliquots were added to other reactions up to a total of 6.4 mL. Dark gray bars represent reactions which contained analyte (lysozyme at 5 μ M) and light gray bars represent reactions carried out in the absence of analyte. Signal generated in the absence of analyte was subtracted. The extent of ligation reactions (signal) were normalized relative to the highest signal observed in reactions treated with a given wash volume. Values are the average of three determinations, and error bars are shown. The black line represents the ratio of signal obtained in the presence of analyte to signal in the absence of analyte (signal-to-noise ratio).

Alternative detection strategies

The detection system used in the array (radioactivity) was chosen because of the facility with which radioactively labeled RNA can be made, and the availability and sensitivity of radiation detection systems. Several types of detection systems could however have been easily adapted to the array. For example, the use of an appropriately labeled oligonucleotide (e.g. in place of the oligonucleotide acceptor) could just as easily have been incorporated into the system. Prior to its final incarnation, several options were explored in order to determine the best method of attachment and detection. Of particular note, two strategies involving the incorporation of modified nucleotides into

the ribozyme were attempted. The use of biotinylated and fluorophore-labeled thymidine nucleotides has been previously shown to be useful for the stochastic incorporation of functional groups into RNA and DNA (Qin and Pyle, 1999). Two strategies similar to the one depicted in Figure 1 could also be envisioned: array elements could consist of biotinylated ribozyme with a labeled substrate, labeling of the ribozyme in the original scheme could be accomplished with fluorophores, as opposed to radiation. In our case however, transcription with small amounts of labeled (biotinylated and fluoresceinated) thymidine ribonucleotides resulted in L1 ribozyme molecules that were completely inactive (data not shown), rendering them useless in the array.

One obvious disadvantage of a biosensor platform utilizing aptazymes (or aptamers) is that the reagents incorporated into the array must already be available, or must be selected from pools. However, with the advent of high-throughput techniques for the selection of functional nucleic acids, (Cox and Ellington, 2001; Cox et al., 2002; Sooter et al., 2001) this problem may be easily overcome.

Aptamer derivatives as biosensor components

RNA and DNA aptamers recognize their ligand partners with high affinity and specificity, and therefore could also be thought of as ideal biosensor components. However, because they do not possess catalytic activity, other strategies must be devised in order to translate target binding into signal generation. One of the simplest approaches is the immobilization of a given aptamer on a surface (e.g. a polystyrene bead), and then flowing labeled analyte over that bead. Signal results from the specific interaction between the labeled analyte and the corresponding bead, relative to some control reaction. A strategy similar to this has been incorporated into an imaging system termed the 'electronic tongue' for the detection of a wide variety of analytes using protein antibodies to capture labeled analytes (Goodey et al., 2001; Goodey and McDevitt, 2003).

This platform also could also detect multiple analytes in parallel for the estimation of cardiac risk (Christodoulides et al., 2002).

An alternate strategy for aptamer-based signal generation is to couple the structural rearrangements undergone by an aptamer during interacting with its ligand into an optical signal. To that end, we have developed two strategies for the incorporation of fluorescent dyes into aptamers that allow for the direct readout of a binding event. The first strategy for creating these 'signaling aptamers' involves the careful placement of dye molecules into the structure of an existing aptamer. Variants derived from ATP-binding RNA and DNA aptamers were previously designed using such a strategy (Jhaveri et al., 2000b). Both variants were able to discriminate between ATP and three other ribonucleotides, while the RNA aptamer had greater affinity for ATP than its DNA counterpart (μM vs. mM K_d s). An important disadvantage of this scheme is that while the creation of these molecules is straightforward, detailed structural information, which may not always be available, is needed for appropriate placement of the dye molecule.

An alternative scheme based on the explicit design of signaling aptamers used *in vitro* selection in order to identify molecules that specifically underwent structural rearrangements in the presence of a ligand (Jhaveri et al., 2000a). By incorporating dye molecules into every member of a nucleic acid pool, variants were identified that underwent ligand specific release from a immobilized state, and thus were presumed to undergo structural rearrangements. Indeed, a variant was identified that could detect ATP at low micromolar concentrations, while maintaining molecular discrimination over several related molecules. Importantly, a single uridine residue was identified as being involved in the 'switch' responsible for signal generation.

Other schemes for the adaptation of nucleic acids to signal generation have been employed with great success. These schemes inevitably exploit structural rearrangements undergone by an aptamer in order to produce the capacity for signal generation. Variants on the commonly used molecular beacons (Hamaguchi et al., 2001; Li et al., 2002;

Yamamoto et al., 2000) have been designed in order to signal the presence of proteins, and DNA aptamers specific to cocaine have been adapted to a signaling platform (Stojanovic et al.).

Here we wish to put forth an additional, perhaps general design strategy that might be useful for the design of signaling aptamers. By exploiting the structural stability of pseudoknots (Du and Hoffman, 1997), one can envision appending sequence elements to an existing aptamer in order for it to assume an additional, pseudo-knotted conformation (**Figure 3.6**). The resulting molecule would have the potential to adopt two alternate folds, and the presence of the cognate ligand could stabilize the molecule in one conformation. The appropriate placement of a fluorophore and quenching molecule could thus potentially lead to signaling capability.

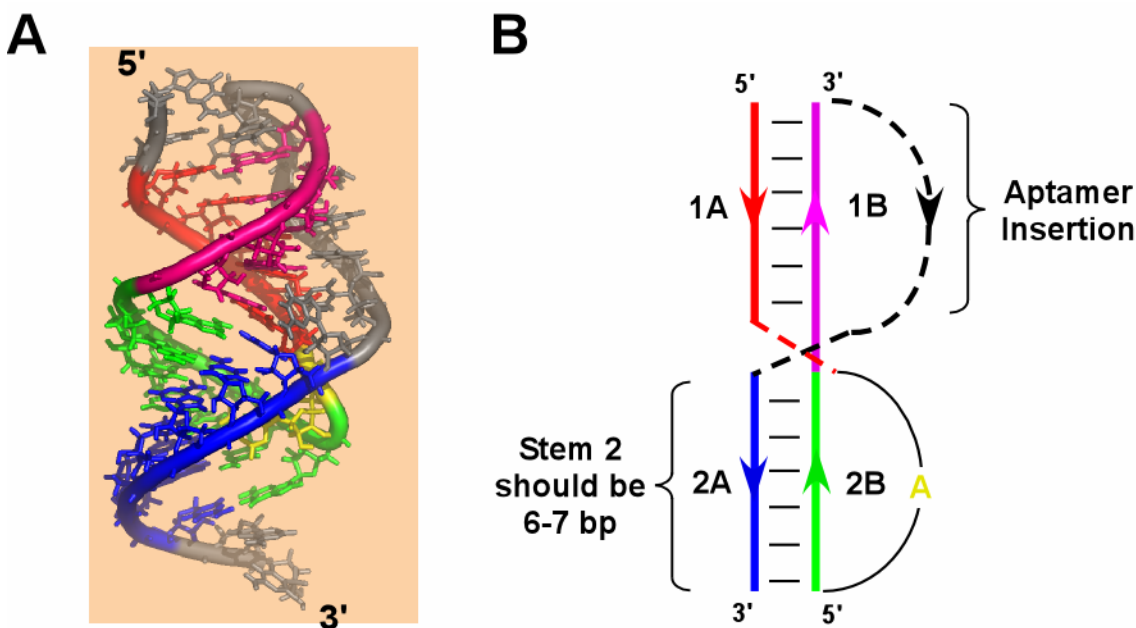


Figure 3.6. Aptamer beacon design strategy utilizing a pseudoknot to induce a conformation change. Incorporation of sequence elements 1A,B and 2A,B could force the molecule to assume an alternate conformation. Insertion of an aptamer sequence and appropriate placement of dye molecules (e.g. fluorophore and quencher) could produce an aptamer beacon.

Conclusion

The adaptation of aptazyme ligases to an array format is an important first step towards the large scale acquisition of biological data using catalytic RNAs. As it stands, the array is comparable to many available technologies in terms of its limits of detection, although several facets of the approach could be improved upon. As more allosteric variants of the L1 ligase are generated, an array that is able to detect additional analytes could be used to address more types of analyte on a single platform, and the development of alternative detection strategies could lead to lower limits of detection as well as improving the time for signal acquisition. The biosensor field in general is rife with examples for using nucleic acids as both acquisition and signal generation components, but improvement of these molecule's capabilities is of the utmost importance for the design of sensors to come.

EXPERIMENTAL PROTOCOLS

Aptazymes and analytes

Sequences for L1, L1-ATP, L1-FMN, L1-theophylline and L1-lysozyme have been previously published (Robertson and Ellington, 1999; Robertson and Ellington, 2000; Robertson and Ellington, 2001). The sequence of the L1-Rev aptazyme was: 5'-GGACCUCGGCGAAAGCCGGUAACGCCACAAGUCGGA
GGGUAAGAUCUGACAGGAACUGGGUGACGGAGGUUAGGUGC (Robertson, 2001). Adenosine 5'-triphosphate (ATP), flavin mononucleotide (FMN), theophylline, and hen egg-white lysozyme were purchased from Sigma-Aldrich (St. Louis, MO). The Rev peptide (NH₂-acyl-TRQARRNRRRRWRERQR-COOH) was obtained from Genemed (San Francisco, CA).

RNA preparation

Templates for individual aptazymes were PCR-amplified using a pair of common, flanking primers (20.T7: 5' TTCTAATACGACTCACTATA, and 18.90A: 5' GCGACTGGACATCACGAG). RNA was generated by *in vitro* transcription. Transcription reactions contained 500 ng of PCR-amplified template, 50 mM Tris-HCl, pH 8.0, 100 mM DTT, 7.5 mM of each of the four ribonucleoside triphosphates, 10 μ Ci [α -³²P] UTP (3000 Ci/mmol, Perkin-Elmer, Boston, MA), and 50 U T7 RNA polymerase (Epicentre, Madison, WI) in total volume of 200 μ l. Following purification on denaturing polyacrylamide gels, RNA molecules were eluted in water, precipitated with ethanol, and resuspended in water. Aptazyme concentrations were determined via absorption measurements at 260 nm.

Aptazyme arrays

Aptazymes (100 pmol per reaction) were annealed with 18.90A (120 pmol / reaction) at 70 °C for 3 minutes in 50 μ l of water, and allowed to cool to room temperature over five minutes. Reaction buffer (30 mM Tris-HCl, pH 7.5, 50 mM NaCl, 60 mM MgCl₂) and substrate (S28A-biotin, 5' biotin-A22ugcacu 3', ribonucleotides in lowercase, 120 pmol) were added, and 50 μ l of this mix was aliquotted into multiple wells of a 96-well PCR plate (MJ Research, Waltham, MA). Analytes in reaction buffer (50 μ l) or reaction buffer alone were added to each well, bringing the final reaction volume to 100 μ l. Reactions were incubated at 25 °C for four hours, and then stopped by the addition of 20 μ l of 0.5 M EDTA. Reactions were transferred to Hi-Bind streptavidin-coated polystyrene plates (Pierce, Rockford, IL) and incubated at room temperature for one hour. The supernatant was removed and wells in the Hi-bind plates were rinsed three times with 200 μ l wash buffer (30 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1% SDS, 7 M urea) and once with 200 μ l TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA). The concentration of aptazyme in figure 3.2 was always 1 μ M, and the

concentrations of analytes were: 1 μ M 18.90A (oligonucleotide), 0.5 mM flavin mononucleotide (FMN), 5 μ M lysozyme, 1 μ M Rev peptide, 1 mM ATP, and 1 mM theophylline. In some reactions in figure 3.6, the number and total volume of wash steps with buffer was increased. Reactions in Figure 3.6 contained lysozyme at a final concentration of 5 μ M. The extents of ligation reactions (signal) were determined by exposing the microtitre plates to PhosphorImager screens and then quantitating bound radioactivity using a PhosphorImager and ImageQuant software (Molecular Dynamics, Sunnyvale, CA). For reactions carried out as a function of ligand concentration, apparent K_d values were calculated by fitting the data to a simple saturation equation, $y = Ax / (B+x)$; A = maximum fraction bound, B = K_d . Limits of detection were defined as the lowest analyte concentrations that gave reproducible signals above background.

REFERENCES

- Brody, E.N., Willis, M.C., Smith, J.D., Jayasena, S., Zichi, D. and Gold, L. (1999) The use of aptamers in large arrays for molecular diagnostics. *Mol Diagn*, **4**, 381-388.
- Christodoulides, N., Tran, M., Floriano, P.N., Rodriguez, M., Goodey, A., Ali, M., Neikirk, D. and McDevitt, J.T. (2002) A microchip-based multianalyte assay system for the assessment of cardiac risk. *Anal Chem*, **74**, 3030-3036.
- Cox, J.C. and Ellington, A.D. (2001) Automated selection of anti-protein aptamers. *Bioorg Med Chem*, **9**, 2525-2531.
- Cox, J.C., Hayhurst, A., Hesselberth, J., Bayer, T.S., Georgiou, G. and Ellington, A.D. (2002) Automated selection of aptamers against protein targets translated *in vitro*: from gene to aptamer. *Nucleic Acids Res*, **30**, e108.
- de Wildt, R.M., Mundy, C.R., Gorick, B.D. and Tomlinson, I.M. (2000) Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat Biotechnol*, **18**, 989-994.
- DeRisi, J.L., Iyer, V.R. and Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680-686.
- Dieckmann, T., Suzuki, E., Nakamura, G.K. and Feigon, J. (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *Rna*, **2**, 628-640.

- Drolet, D.W., Moon-McDermott, L. and Romig, T.S. (1996) An enzyme-linked oligonucleotide assay. *Nat Biotechnol*, **14**, 1021-1025.
- Du, Z. and Hoffman, D.W. (1997) An NMR and mutational study of the pseudoknot within the gene 32 mRNA of bacteriophage T2: insights into a family of structurally related RNA pseudoknots. *Nucleic Acids Res*, **25**, 1130-1135.
- Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gustafsdottir, S.M., Ostman, A. and Landegren, U. (2002) Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol*, **20**, 473-477.
- Fritz, J., Baller, M.K., Lang, H.P., Rothuizen, H., Vettiger, P., Meyer, E., Guntherodt, H., Gerber, C. and Gimzewski, J.K. (2000) Translating biomolecular recognition into nanomechanics. *Science*, **288**, 316-318.
- Fritz, J., Cooper, E.B., Gaudet, S., Sorger, P.K. and Manalis, S.R. (2002) Electronic detection of DNA by its intrinsic molecular charge. *Proc Natl Acad Sci U S A*, **99**, 14142-14146.
- Goodey, A., Lavigne, J.J., Savoy, S.M., Rodriguez, M.D., Curey, T., Tsao, A., Simmons, G., Wright, J., Yoo, S.J., Sohn, Y., Anslyn, E.V., Shear, J.B., Neikirk, D.P. and McDevitt, J.T. (2001) Development of multianalyte sensor arrays composed of chemically derivatized polymeric microspheres localized in micromachined cavities. *Journal of the American Chemical Society*, **123**, 2559-2570.
- Goodey, A.P. and McDevitt, J.T. (2003) Multishell microspheres with integrated chromatographic and detection layers for use in array sensors. *J Am Chem Soc*, **125**, 2870-2871.
- Hamaguchi, N., Ellington, A. and Stanton, M. (2001) Aptamer beacons for the direct detection of proteins. *Anal Biochem*, **294**, 126-131.
- Hartig, J.S., Najafi-Shoushtari, S.H., Grune, I., Yan, A., Ellington, A.D. and Famulok, M. (2002) Protein-dependent ribozymes report molecular interactions in real time. *Nat Biotechnol*, **20**, 717-722.
- Hesselberth, J.R., Robertson, M.P., Knudsen, S.M. and Ellington, A.D. (2003) Simultaneous detection of diverse analytes with an aptazyme ligase array. *Anal Biochem*, **312**, 106-112.
- Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425-1429.
- Jenne, A., Hartig, J.S., Piganeau, N., Tauer, A., Samarsky, D.A., Green, M.R., Davies, J. and Famulok, M. (2001) Rapid identification and characterization of hammerhead-ribozyme inhibitors using fluorescence-based technology. *Nat Biotechnol*, **19**, 56-61.

- Jhaveri, S., Rajendran, M. and Ellington, A.D. (2000a) *In vitro* selection of signaling aptamers. *Nat Biotechnol*, **18**, 1293-1297.
- Jhaveri, S.D., Kirby, R., Conrad, R., Maglott, E.J., Bowser, M., Kennedy, R.T., Glick, G. and Ellington, A.D. (2000b) Designed Signaling Aptamers that Transduce Molecular Recognition to Changes in Fluorescence Intensity. *Journal of the American Chemical Society*, **122**, 2469-2473.
- Koizumi, M., Soukup, G.A., Kerr, J.N. and Breaker, R.R. (1999) Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nat Struct Biol*, **6**, 1062-1071.
- Li, J.J., Fang, X. and Tan, W. (2002) Molecular aptamer beacons for real-time protein recognition. *Biochem Biophys Res Commun*, **292**, 31-40.
- McDonnell, J.M. (2001) Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. *Curr Opin Chem Biol*, **5**, 572-577.
- Porta, H. and Lizardi, P.M. (1995) An allosteric hammerhead ribozyme. *Biotechnology (N Y)*, **13**, 161-164.
- Qin, P.Z. and Pyle, A.M. (1999) Site-specific labeling of RNA with fluorophores and other structural probes. *Methods*, **18**, 60-70.
- Robertson, M.P. (2001) Engineered regulation of an RNA ligase ribozyme. *Molecular and Cell Biology*. University of Texas, Austin, TX, p. 197.
- Robertson, M.P. and Ellington, A.D. (1999) *In vitro* selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat Biotechnol*, **17**, 62-66.
- Robertson, M.P. and Ellington, A.D. (2000) Design and optimization of effector-activated ribozyme ligases. *Nucleic Acids Res*, **28**, 1751-1759.
- Robertson, M.P. and Ellington, A.D. (2001) *In vitro* selection of nucleoprotein enzymes. *Nat Biotechnol*, **19**, 650-655.
- Rowe, C.A., Tender, L.M., Feldstein, M.J., Golden, J.P., Scruggs, S.B., MacCraith, B.D., Cras, J.J. and Ligler, F.S. (1999) Array biosensor for simultaneous identification of bacterial, viral, and protein analytes. *Anal Chem*, **71**, 3846-3852.
- Rowe-Taitt, C.A., Golden, J.P., Feldstein, M.J., Cras, J.J., Hoffman, K.E. and Ligler, F.S. (2000a) Array biosensor for detection of biohazards. *Biosens Bioelectron*, **14**, 785-794.
- Rowe-Taitt, C.A., Hazzard, J.W., Hoffman, K.E., Cras, J.J., Golden, J.P. and Ligler, F.S. (2000b) Simultaneous detection of six biohazardous agents using a planar waveguide array biosensor. *Biosens Bioelectron*, **15**, 579-589.

- Sassanfar, M. and Szostak, J.W. (1993) An RNA motif that binds ATP. *Nature*, **364**, 550-553.
- Schweitzer, B., Roberts, S., Grimwade, B., Shao, W., Wang, M., Fu, Q., Shu, Q., Laroche, I., Zhou, Z., Tchernev, V.T., Christiansen, J., Velleca, M. and Kingsmore, S.F. (2002) Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nat Biotechnol*, **20**, 359-365.
- Seetharaman, S., Zivarts, M., Sudarsan, N. and Breaker, R.R. (2001) Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nat Biotechnol*, **19**, 336-341.
- Sooter, L.J., Riedel, T., Davidson, E.A., Levy, M., Cox, J.C. and Ellington, A.D. (2001) Toward automated nucleic acid enzyme selection. *Biol Chem*, **382**, 1327-1334.
- Soukup, G.A. and Breaker, R.R. (1999) Engineering precision RNA molecular switches. *Proc Natl Acad Sci U S A*, **96**, 3584-3589.
- Stojanovic, M.N., de Prada, P. and Landry, D.W. (2001) Aptamer-based folding fluorescent sensor for cocaine. *J Am Chem Soc*, **123**, 4928-4931.
- Tang, J. and Breaker, R.R. (1997) Rational design of allosteric ribozymes. *Chem Biol*, **4**, 453-459.
- Vaish, N.K., Dong, F., Andrews, L., Schweppe, R.E., Ahn, N.G., Blatt, L. and Seiwert, S.D. (2002) Monitoring post-translational modification of proteins with allosteric ribozymes. *Nat Biotechnol*, **20**, 810-815.
- Wang, D.Y. and Sen, D. (2002) Rationally designed allosteric variants of hammerhead ribozymes responsive to the HIV-1 Tat protein. *Comb Chem High Throughput Screen*, **5**, 301-312.
- Winssinger, N., Ficarro, S., Schultz, P.G. and Harris, J.L. (2002) Profiling protein function with small molecule microarrays. *Proc Natl Acad Sci U S A*, **99**, 11139-11144.
- Yamamoto, R., Baba, T. and Kumar, P.K. (2000) Molecular beacon aptamer fluoresces in the presence of Tat protein of HIV-1. *Genes Cells*, **5**, 389-396.

Chapter 4: Computational engineering of allosteric ribozymes

INTRODUCTION

Computational design of biopolymers

Computational design strategies have been extremely successful in the design of several classes of biopolymers. One of the most striking examples is the *in silico* design of proteins. Since inaugural studies on the relationship between protein folding and activity (Anfinsen, 1973), researchers have used computational means in order to describe and predict the folds and activities of proteins. Levinthal put forth his paradox to illustrate the huge number of possible conformations a protein can assume while going from an unfolded to folded state (Zwanzig et al., 1992). While Levinthal's paradox is a theoretical maximum, simple assumptions made from studying the native folds of proteins have also proved true, leading to the discovery that only a very small subset of those possible conformations need be considered, and ultimately ending in the ability to design and test functional protein variants. Over the last twenty years, researchers have become almost comfortable with the design of novel proteins and protein folds, allowing them to produce proteins of novel function in addition to the ability to predict three-dimensional fold given a protein's primary sequence. Homme Hellinga and his colleagues have pioneered the field of computational design of functional proteins. By superimposing geometrically positioned amino acid side chains onto existing protein structures, they have succeeded in incorporating redox-active Fe-S clusters into proteins which previously had none (Benson et al., 2000; Hellinga, 1998). Stephen Mayo provided some of the first examples of computational protein folding in the *de novo* design of novel proteins that adopt a predefined fold (Dahiyat and Mayo, 1997). David Baker and his lab have shown that a reductionist approach to the protein folding problem

(the 'Rosetta algorithm') can be applied in order to predict the three-dimensional folds of proteins from primary sequence alone with high accuracy (Bonneau and Baker, 2001; Bonneau et al., 2001).

The RNA folding problem

While researchers appear to have a firm handle on the design of novel proteins and protein folds, the design of RNA folds remains an unsolved problem. Natural RNAs are known to adopt incredibly complex folds. The crystal structure of the ribosome has illustrated a myriad of novel conformations RNA is able to assume (Ban et al., 2000; Yusupov et al., 2001). The crystal structures of several ribozymes have provided insight into how RNA utilizes its unique structural and chemical properties to enable catalysis (Cate et al., 1996; Ferre-D'Amare et al., 1998; Rupert and Ferre-D'Amare, 2001). Researchers have also devised methods for the prediction of novel RNA folds and motifs from phylogenetic comparisons of RNA sequences using covariation that can be powerful in predicting novel types of RNA folds (Gutell et al., 2000a; Gutell et al., 2000b; Lee et al., 2003). The difficulty in predicting how a given RNA sequence will fold in three-dimensions is rooted in the conformational flexibility of RNA: peptide chains are limited to two torsional degrees of freedom in their backbones (per amino acid monomer) and typically assume a finite number of structures (e.g. helices, sheets and turns), while the phosphodiester backbone of RNA has seven degrees of torsional freedom, allowing it to fold in complex manners. As the structures of different RNAs are solved, its ability to adopt strange and unpredicted folds becomes more apparent (Tereshko et al., 2003).

Algorithms for the prediction of nucleic acid secondary structure have advanced to the point where nucleic acid secondary structures can be rapidly enumerated based on primary sequence (Hofacker et al., 1994; Zuker, 2000). Because the algorithms are relatively fast and accurate, they have been employed widely by researchers as 'first-pass'

glimpses into how a particular RNA sequence might fold. Variants of the folding algorithm have also been developed which aligned sequences are compared in a pairwise fashion in order to identify residue pairs that differ in the alignment (e.g. A vs. G) but may contribute identically at the level of secondary structure (Mathews and Turner, 2002).

Aptazyme engineering

Allosteric ribozymes (aptazymes) have been generated using both selection and rational design methods (Soukup and Breaker, 2000). Both methods have frequently relied on conformational changes that involve alternative secondary structures or pairings. For example, flavin-binding induces a conformational change in the structure of a selected, anti-FMN aptamer (Fan et al., 1996; Patel et al., 1997). When the anti-FMN aptamer is conjoined to a hammerhead ribozyme via a randomized stem, variants can be selected in which this conformational change is transduced into a change in ribozyme conformation and activity, yielding effector-modulated aptazymes (Soukup and Breaker, 1999a). At least one class of the flavin-dependent hammerhead aptazymes underwent a conformational change in which the register of the joining region between aptamer and ribozyme changed upon addition of flavin (Soukup and Breaker, 1999a; Soukup and Breaker, 2000). Other hammerhead aptazymes have been generated in which the ligand-dependent formation of an aptamer resulted in the disruption of a programmed, inhibitory interaction and the concomitant activation of the ribozyme (Hartig et al., 2002; Vaish et al., 2002).

One mechanism of activation for allosteric ribozymes was found to be a ‘slip structure’ in which FMN-binding facilitated the formation of alternative base-pairs in the joining region. Because RNA secondary structures can be addressed computationally, we have attempted to determine if a model could be developed to describe the energetic

changes a ribozyme undergoes upon a structural rearrangement, and whether that model could in turn be used for the identification of novel allosteric ribozymes.

RESULTS AND DISCUSSION

Computational assessment of slip structures

The ‘slip structure’ model assumes that there are two conformational states for flavin-activated hammerhead aptazymes: unliganded, inactive conformers in which the joining regions are ‘out-of-register,’ and liganded, active conformers in which the joining regions are ‘in register.’ In our model, a change in register is simply the stepwise movement of the top strand of a joining region relative to the bottom strand (**Figure 4.1a**).

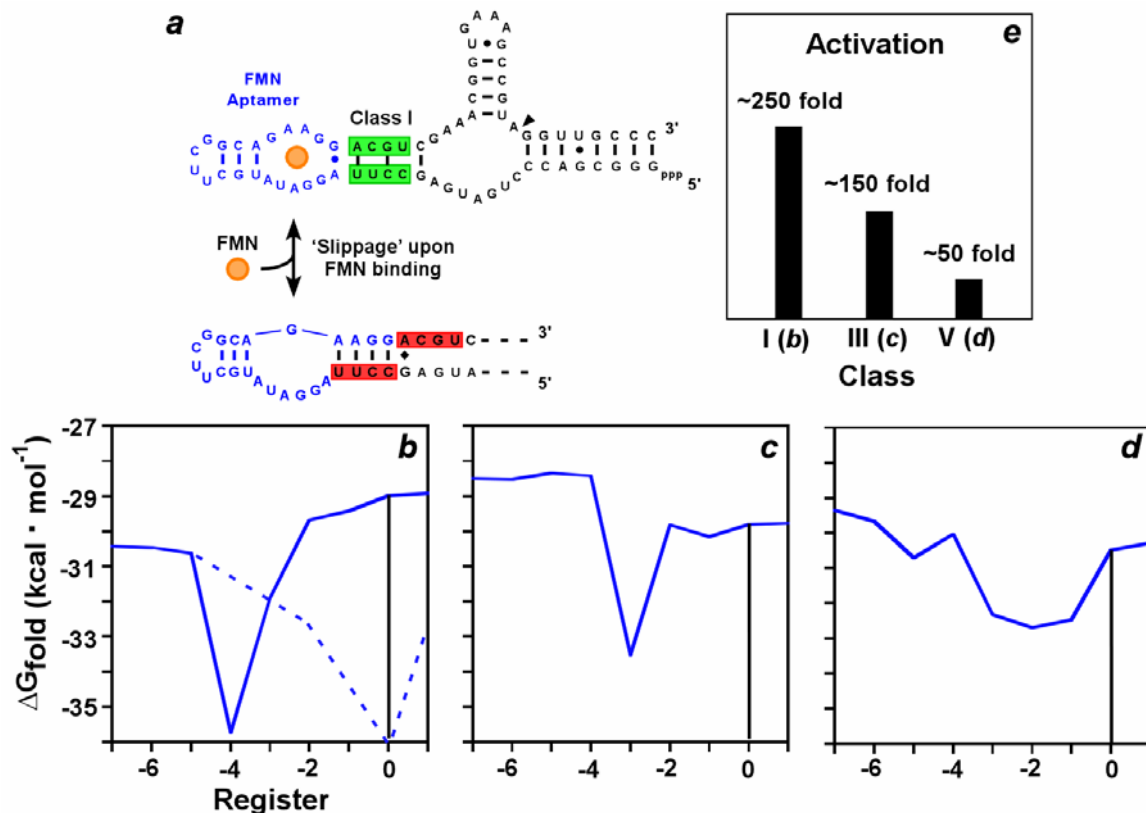


Figure 4.1. Energetic profile model. **a**. Schematic of a 'slip-structure'. In the absence of ligand (FMN), the joining region between an aptamer and a ribozyme is trapped in an inactive conformation (red). A structural rearrangement ensues upon ligand binding, shifting the joining region to the 'active register' (green) and enabling ribozyme catalysis. **b,c,d**. Energetic profiles of three previously selected, FMN-responsive hammerhead aptazymes. The free energies of folding (ΔG_{fold}) are plotted versus the register of the joining region. Position '0' (vertical line) denotes the 'active' register of the ribozyme. The dashed line in b is a hypothetical energetic profile following ligand-binding. b represents the Class I hammerhead aptazyme, c represents the Class III, and d represents the Class V. **e**. The experimental levels of flavin-activation seen *in vitro* for each of the hammerhead aptazymes depicted in b,c and d.

In order to assess whether there might be one or more stable, out-of-register conformers for flavin-activated hammerhead aptazymes, we constructed a plot of the free energy of folding (ΔG_{fold}) for each possible base-pair register within the joining regions of several different aptazymes (Figure 4.1b,c,d). An 'active register' (denoted '0') was first defined based on the known secondary structure of the hammerhead ribozyme. All conformations near the active register were then computationally generated by simply shifting the lower (5') strand of the joining region relative to the upper strand, and

determining the folding energy at each register (see also **Figure 4.2**). Conceptually, the energetic profiles that were generated describe the changes in free energy that the aptazymes undergo as they ‘slip’ from one register to the next.

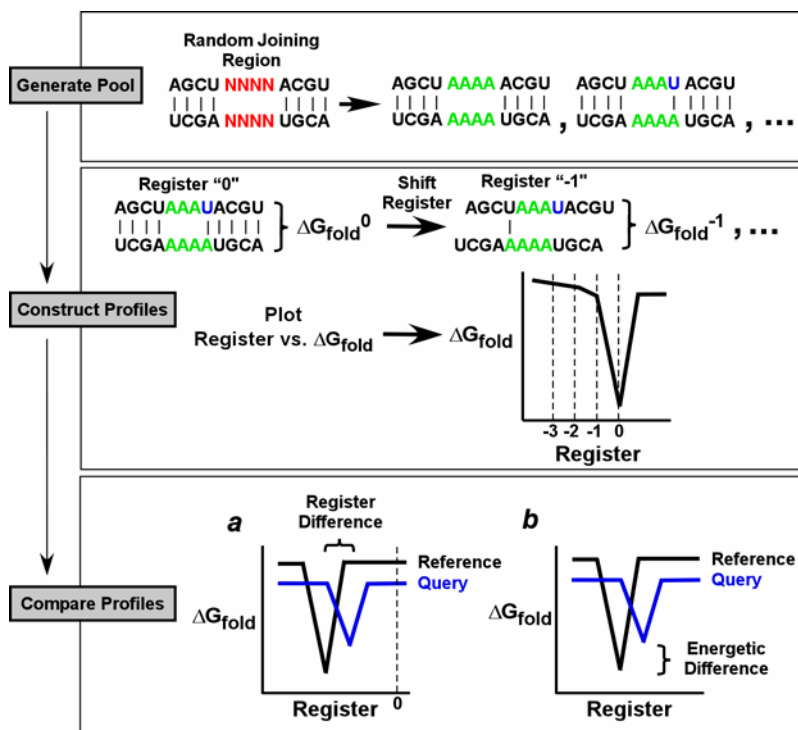


Figure 4.2. Computational selection of aptazymes. Initially, the sequences of all members of a joining region pool were generated. To construct profiles, the ‘register’ of the joining region was fixed and the ΔG_{fold} was calculated. The register was then moved and ΔG_{fold} was again calculated. A profile of ΔG_{fold} versus register was created and compared with a reference profile from an aptazyme that had been generated experimentally. In order to be identified by the algorithm (‘Identified,’ **Table 4.1**), (a) the stable inactive energetic minimum of a given sequence variant (‘Query’) had to be less than or equal to two registers away from the stable ‘inactive’ energetic minimum in the reference profile (e.g., if the difference in register between an inactive conformation and the active register was 4 in the reference profile, then for a computed variant it had to be within 2 to 4). In addition, (b) the difference in ΔG_{fold} at the inactive registers of the reference and query profiles had to be within 3 kcal mol⁻¹. Except for the N5-N5 selection, typically relatively few sequences were computationally sieved from the populations based on these parameters.

Based on this analysis, we hypothesize that in one of the selected, flavin-activated hammerhead aptazymes (Class I) a register exists in which residues in the joining region (CCUU) are predicted to form base-pairs with a complementary portion of the anti-FMN

aptamer (Figure 4.1a). This pairing should effectively disrupt the structure and function of both the anti-FMN aptamer (Fan et al., 1996) and the hammerhead. The free energy of flavin-binding presumably induces the formation of a different stem containing only two Watson-Crick base-pairs between the strands of the joining region. In this ‘active register’, aptamer residues important for flavin-binding (such as the terminal A-G pairing, see Fan et al., 1996) are no longer sequestered and the hammerhead ribozyme can in turn assume an active conformation.

The energetic profiles of two other flavin-activated hammerhead aptazymes with different joining regions (Class III and Class V) were also constructed. As was the case with the Class I aptazyme, out-of-register free energy minima were observed (Figures 4.1c, d). Interestingly, the relative, experimental levels of activation observed for each of the selected aptazymes could be explained based on these profiles. Class I was predicted to have the greatest energetic separation between the stable inactive and active registers (Figure 4.1b), and also showed the greatest effector-activation (Figure 4.1e). Class III and Class V have smaller predicted energy barriers, and exhibited smaller activations. Both the height of the energy barrier and the distance between the stable inactive register and the active register may affect the level of aptazyme activation.

The free energy model is not only consistent with the available kinetic data, but also with available structural data. A hammerhead ribozyme that contained an anti-FMN aptamer and a Class I communication module was probed using an ‘in-line assay’ that exploited conformation-dependent differences in hydrolysis (Jose et al., 2001; Soukup and Breaker, 1999b). In the assay, flexible residues (e.g. a helix-capping tetraloop) show greater cleavage activity than residues that retain a fixed conformation (e.g. canonically base-paired helices). The residues that we predict to be unpaired in the unliganded structure are in fact cleaved (see Figure 4 in Jose et al., 2001). These residues should assume a more stable conformation in the liganded structure, and the cleavage patterns are found to be correspondingly suppressed. A model was proposed previously

suggested that movements of only single base-pairs were responsible for the activation of individual aptazymes (Soukup and Breaker, 1999a). While single base-pair changes in register might be responsible for the activation of other aptazymes, the model originally proposed for the 'class I induction element' is neither consistent with our energetic model, nor from subsequent data from the same researchers.

As final support for our model, energetic profiles for previously selected hammerhead aptazymes (Soukup and Breaker, 1999a) that were inhibited rather than activated by FMN were constructed (data not shown). As expected, there were energetic minima at the active register and higher energy peaks adjacent to these minima; in other words, the converse of the energetic profiles observed for flavin-activated aptazymes. Presumably, FMN-binding leads to a reorganization of the joining region such that an inactive register becomes favored relative to the active register.

Computational selection of allosteric ribozymes

To the extent that the energetic model can explain the activation parameters of hammerhead aptazymes, it should also prove possible to predict the sequences and structures of new hammerhead aptazymes. To this end, we developed a simple algorithm to select aptazymes from computationally-generated random sequence populations (**Figure 4.2**). Energetic profiles similar to those shown in Figure 4.1 were computed for each member of a pool in which the hammerhead ribozyme was joined to the anti-FMN aptamer via a symmetric N3-N3 joining region ($4^6 = 4,096$ variants; **Figure 4.3a**). The black line in Figure 4.3a corresponds to the average energetic composition of the unselected pool. The algorithm compared the energetic profiles of each candidate in the pool to a reference energetic profile, that of the Class VI hammerhead aptazyme (Figure 4.3a, blue line; this aptazyme is as activated as Class I, **Figure 4.1e**). Aptazyme variants that had matching profiles were selected for further analysis. In order to be selected, variants first had to contain a single energetic minimum between 2 to 4 registers of

position ‘0’ in the reference profile (‘inactive register,’ **Table 4.1**). In addition, the ΔG_{fold} at the inactive register of both the query and reference profiles was compared, and the difference between those energies was minimized. Based on these criteria a total of 23 candidate aptazymes were chosen; the energetic profiles of two of these aptazymes are shown in **Figure 4.3a**.

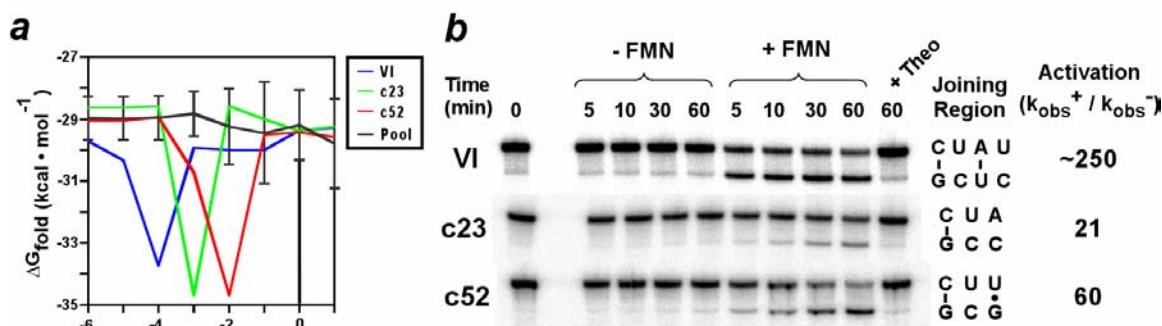


Figure 4.3. Application of the energetic profile model. a. The energetic profiles of every member of a pool in which the hammerhead ribozyme was joined to an anti-flavin aptamer via a symmetric, N3-N3 stem were compared to a reference energy profile (Class VI hammerhead aptazyme, blue line). The average energetic profile and distribution (error bars) of the pool are shown in black. The energetic profiles of two candidate aptazymes are shown in green (c23) and red (c52). b. Activity assays for selected variants. The sequences of the joining regions for the aptazymes are shown, in addition to their levels of activation in the presence of FMN.

All 23 designed aptazymes were synthesized and the aptazyme variants were assayed *in vitro*. Three variants exhibited greater than 4-fold activation in the presence of FMN (Table 4.1), and two of these, c23 and c52, exhibited activations ($k_{\text{obs}}^+ / k_{\text{obs}}^-$) of 21- and 60-fold, respectively (Figure 4.3b). Of the remaining aptazymes eight were constitutively inactive and twelve were constitutively active.

A control was carried out in which 23 aptazyme variants were selected at random from the population of 4,096 variants. Three of these clones were found to be activated by FMN, although not at the same level as with energetic profiling; the best variant was activated 5-fold by FMN (Table 4.1). In addition, five of the random selections were found to be inactivated by FMN, whereas no profiled variants were inactivated.

Some joining regions identified by previous selection experiments were shown to serve as ‘communication modules’ that allowed different aptamers to activate the hammerhead ribozyme (Soukup and Breaker, 1999a). We therefore replaced the flavin-binding domain with an anti-theophylline aptamer that had previously been shown to be useful for the selection of theophylline-activated hammerhead aptazymes (Jenison et al., 1994; Soukup et al., 2000). One of the computationally-selected joining regions (c23) was able to confer 17-fold activation when the anti-theophylline aptamer was swapped for the anti-FMN aptamer; the other joining region (c52) did not.

As a comparison, the anti-theophylline aptamer was appended to the N3-N3 pool, and the pool was computationally sieved using parameters identical to those that had identified flavin-dependent aptazymes. Four sequences were identified as candidate aptazymes, and were assayed for theophylline responsivity (**Table 4.1**). A single variant from the four exhibited a 14-fold increase in rate upon the addition of theophylline. At least in this example, both adaptation of a pre-existing communication module and the de novo prediction of aptazyme joining regions yielded aptazymes with similar activation parameters.

Varying the length of the joining region

Anecdotal evidence indicated that there might be a relationship between the length of the joining region between an aptamer and a ribozyme and the levels of activation exhibited by selected aptazymes. For example, when an anti-FMN aptamer was joined to the hammerhead ribozyme by a N3/4-N3/4 stem, selected aptazymes were activated from 50- to 270-fold in the presence FMN (Soukup and Breaker, 1999a). When an anti-theophylline aptamer was joined to the hammerhead ribozyme via a N5-N5 stem, selected aptazymes were activated between 1,300- to 4,700-fold by theophylline (Soukup et al., 2000).

To more directly address the question of whether stem length directly influences activation parameters, the anti-theophylline aptamer and hammerhead ribozyme were joined by randomized, symmetric stems of three different lengths (N3-N3, N4-N4 or N5-N5) and the resultant pools were computationally sieved. The register used for sieving was also changed to mimic the best-activated variants from the experimental N5-N5 selection (**Table 4.1**). From the pool with a N3-N3 joining region 44 sequences were identified and 24 were assayed, while from the pool with a N4-N4 joining region 48 candidates were identified and 27 were experimentally assayed. From the pool containing a N5-N5 joining region, 11,637 sequences (representing 1.1% of the original pool) were identified, despite the fact that the same parameter sets were used for evaluation. The complexity of this population was further reduced by eliminating all but one variant from amongst those whose 5'-most N5 regions were identical. This left 68 candidate sequences, and 25 were experimentally assayed. Following computational selection, one clone from the N3-N3 pool was activated 14-fold by theophylline. Two variants from the N4-N4 pool were found to be activated by theophylline (9- and 20-fold), and one clone from the N5-N5 pool was theophylline-responsive (16-fold). At least for computationally-selected variants, the length of the pool does not appear to increase the likelihood of obtaining aptazymes with better activation parameters.

To again ensure that we had not identified aptazymes by chance alone, ten designs from the N4-N4 and N5-N5 pools were picked at random from the pool, synthesized, and assayed for activation in the presence of theophylline. Not only were no effector-dependent variants were found, but the distribution of active and inactive variants was very different between the randomly selected and energetically sieved sequences (**Table 4.1**). The randomly selected sequences favored constitutively active variants, while the energetically sieved sequences contained many fewer constitutively active variants than inactive variants.

Table 4.1. Summary of computationally selected aptazymes.

Pool	Ribozyme	Aptamer	State	Register ^a	Off ^b	On ^c	Activated	Inhibited	Identified ^d	Tested
N3-N3	HH	FMN	Random	-	3	12	3	5	-	23
N3-N3	HH	FMN	Selected	-4	8	12	3	0	23	23
N3-N3	HH	Theo	Selected (1)	-4	0	3	1	0	4	4
N3-N3	HH	Theo	Selected (2)	-8	4	19	1	0	44	24
N4-N4	HH	Theo	Random	-	3	7	0	0	-	10
N4-N4	HH	Theo	Selected	-8	17	8	2	0	48	27
N5-N5	HH	Theo	Random	-	4	6	0	0	-	10
N5-N5	HH	Theo	Selected	-8	23	1	1	0	11,637 (68)	25
N3-N3	Hairpin	FMN	Selected	-4	12	0	1	0	13	13
N5-N5	HH	FMN	Selected	-	13	1	1	0	8,630	15

HH = Hammerhead

a 'Register' indicates the required distance between the stable inactive and active registers.

This distance could vary up to 2 positions in the algorithm, as stated in Figure 2.

b 'Off' indicates ribozymes that were found to be constitutively inactive.

c 'On' indicates ribozymes that were found to be constitutively active.

d 'Identified' gives the number of ribozymes picked by the algorithm.

e 'Tested' gives the number of ribozymes synthesized and assayed.

Generalization of the method

We attempted to determine whether our computational methods could be generalized to different catalytic platforms. The hairpin ribozyme was chosen as a candidate platform because of its structure, which consists of stacked helices (Rupert and Ferre-D'Amare, 2001). In addition, hairpin aptazymes have been recently described which were selected *in vitro* for their ability to undergo cleavage in the presence of a short oligonucleotide (Komatsu et al., 2002).

The anti-FMN aptamer was appended to the end of loop B of the hairpin ribozyme via a symmetric N3-N3 random region, and thirteen candidate aptazymes were identified using the same sieving algorithm that had proven successful for hammerhead aptazymes. All 13 sequences were then assayed *in vitro* for their ability to undergo cleavage in the presence of FMN. A single sequence was found to be activated 10-fold in the presence of FMN (**Figure 4.4**).

given the individual base-pairing probabilities. In addition, another structure approximation is obtained by finding the portion of the Boltzmann ensemble occupied by the minimum free energy structure (f_{mfe}) (Wuchty et al., 1999). It was previously shown the sequences can adopt several suboptimal folds, and their stability can be addressed by comparison with the minimum free energy fold. Either measure effectively quantifies the stability of a given RNA structure relative to alternate folding conformations. Herein I call individual folded states 'occupancies': since we are interested in the stability of a small portion of the RNA molecule (the joining region and surrounding areas) we can assess a sequence's ability to lie in alternate conformations using the above approximations.

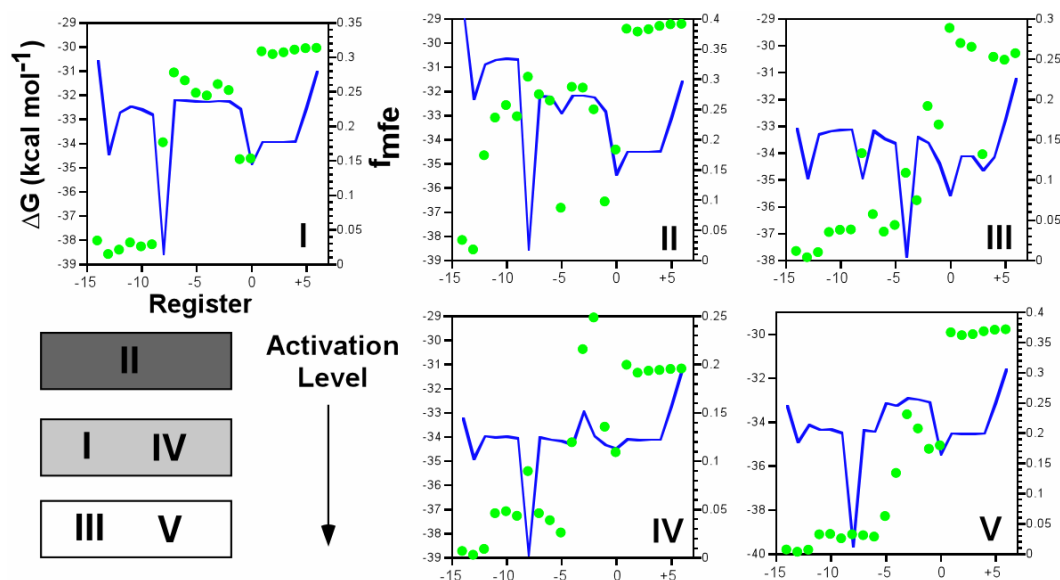


Figure 4.5. Incorporation of occupancy measures in the design algorithm. Theophylline-dependent hammerhead aptazymes were previously identified from a pool in which a symmetric N5-N5 stem joined the aptamer and ribozyme domains (Soukup et al., 2000). Free energy (blue lines) and occupancy measures (green dots) were calculated for each of the variants and are plotted above. The most active variant, class II, exhibits several features: the presence of a highly-occupied free energy minimum (the inactive register), a relatively poorly occupied active register, and an alternate, highly occupied window (registers 1-6). Several other features may explain the reduced levels of activation seen for some classes. The class III aptazyme contains a poorly occupied register in the alternate window (position +3), while the class V ribozyme's inactive register is poorly occupied (position -8).

This information was incorporated in a second iteration of the design algorithm. By plotting the fraction of molecules involved in the minimum free energy structure in a particular conformation (f_{mfe}), a distinctive pattern arises for previously selected theophylline hammerhead aptazymes (**Figure 4.5**). In our algorithm, we assume that two registers of the ribozyme structure are important for its activity, the active and inactive registers. For example, in Figure 4.5 the class I aptazyme contains an inactive register at position -8 while its inactive

register lies at position 0. The occupancy values at these positions are consistent with their activity: the inactive register is highly occupied while the active register has a reduced occupancy value, indicating that the molecule is well-folded in an inactive state compared to the state in which it would be catalytically active. In addition, the occupancy states between these two registers have high values, indicating that this region is well folded and represents a large fraction of the ensemble of possible structures. This stretch of occupancy values (an 'occupancy window'), can thus be thought of as the ability of a molecule to move from one structural state (i.e. folded in a particular register) to another. By looking through an occupancy window from the current register to a register in question, we can assess the ability of the molecule to undergo a structural rearrangement.

The second iteration of the algorithm allowed for the rank ordering of clones by propensity for structural change. By first identifying clones with defined energy minima, we could then find the occupancy windows between the inactive register and that energy minimum, and maximize the ability of the RNA to move through that window to an alternate conformation. In practice, we maximize the average value of a given occupancy window and minimize its standard deviation. This process was empirically optimized on two test sets of data (sequences previously selected from an N4-N4 FMN and N5-N5 theophylline pools in Soukup and Breaker, 1999a; Soukup et al., 2000). The position and depth of energetic minima and the average and standard deviations of occupancy windows surrounding the active register were used as input data for a

series of sorting steps; ultimately the sorting procedure that yielded the highest ranking of previously identified aptazymes was used in subsequent selections.

In order to test the second iteration of the algorithm, a pool containing and N5-N5 stem joining the FMN aptamer and hammerhead ribozyme was sieved using both the free energy and occupancy parameters. Following this selection, 8,630 clones were identified as candidates, and the top 15 were tested for FMN responsivity. One of the clones exhibited an activation of ~20 fold in the presence of FMN.

A second test of this improved algorithm was attempted in order to test its ability to accommodate different aptamer structures. An RNA aptamer that recognizes coenzyme A was previously selected, and its secondary structure, an extended, bulged helix, is similar in structure to aptamers that had previously been used successfully for the selection of aptazymes, and thus seemed appropriate for a design attempt (Burke and Hoffman, 1998). The aptamer was appended via an N4-N4 stem to the hammerhead ribozyme and a selection for candidate aptazymes was performed. After selection, 202 aptazymes were identified and the top 10 were tested *in vitro* for coenzyme A responsivity. None of the clones were activated in the presence of coenzyme A at both neutral and slightly acidic (6.0) pH values. In addition, an aptamer that recognizes a peptide derived from the Rev protein of HIV (Giver et al., 1993) was incorporated into the design algorithm. This aptamer was also shown to undergo a structural rearrangement upon peptide-binding, and so was also used in an additional design attempt. Following application of our model, 6,231 candidates were identified, and the top

ten 10 were tested *in vitro*. None of the tested candidates exhibited peptide-dependence.

Aptamer structures could dictate design success

Several aptazyme constructs were used in the design algorithm, but only a few yielded active aptazymes. One theme that was noted was the structure of the aptamers that yielded active aptazymes. In all cases, the design algorithm was capable of success when the aptamers employed possessed simple secondary structures with an apparent single helical axis. For example, the FMN and theophylline aptamers are small (~40 nucleotide) motifs that are comprised of a single helix capped with a tetraloop containing bulged nucleotides on either side of the structure. The structures of these aptamers have previously studied using NMR spectroscopy (Fan et al., 1996; Zimmermann et al., 2000), and in both cases the ligand binding site located in the bulged area in the middle of the helix.

On the other hand, several of the constructs that failed to yield active aptazymes were more complicated aptamer motifs. Aptazyme constructs incorporating aptamer regions specific for ATP (Dieckmann et al., 1996; Sassanfar and Szostak, 1993) and cGMP (Koizumi et al., 1999) were also sieved for active molecules, and none were identified. Considerably fewer candidate profiles were identified using the aptamers due to the lack of defined energetic minima in the majority of the profiles. These two empirical observations suggest a limitation in the folding algorithm itself, which may be unable to describe RNA structures that are not 'well-folded'. In other words the sequence of a given RNA

may not adequately reflect the structure it adopts, and thus cannot be appropriately described by our model. Taken further, this information may be incorporated as an additional parameter in the design algorithm in order to eliminate candidate clones, as it indicates that the secondary structures of the aptamers are poorly defined by the existing RNA folding software.

Other applications of the design algorithm

Recently, an algorithm for the design of molecular beacons was described. The researchers propose that beacons can be designed using free energy values calculated using mfold (Monroe and Haselton, 2003). Similarly, because our design algorithm describes and predicts the local conformational changes an RNA molecule can undergo, we reasoned that the process might be applicable to the design of aptamer beacons. In the simplest case, the ribozyme domain from a previously selected aptazyme (e.g. **Figure 4.1a**) can be replaced with a fluorophore on the 5' terminus and a quenching molecule on the 3' terminus. The structural change undergone by the molecule upon FMN binding could then bring the fluorophore out of proximity from the quencher resulting in an optical signal.

The appendage of random sequence elements to the termini of existing aptamer structures followed by application of the design algorithm could identify molecules that are able to undergo the requisite structural changes, and thus be suitable for signal generation with appropriately placed dye molecules. This strategy has previously been used to generate aptamer beacons that respond to

VegF, Angiogenin and ricin (Bradley Hall, Ellington Lab, personal communication).

Conclusion

The levels of activation obtained via computation are disproportionate to those obtained from *in vitro* selection. This may be because the algorithms we have developed do not take into account all of the energetic factors involved in aptazyme conformational changes; for example, we constrain conformational change to slip structures and there may be other accessible structures. Alternatively, it may be that the computed energetic profiles of the experimental standards used for comparison are not representative of the true energetics of a conformational change. For example, if the inactive conformers of the experimentally selected aptazymes are less stable than calculated, then these aptazymes may be better poised to undergo conformational changes than their computed counterparts. Thus, it is possible that our initial algorithm was simply too restrictive in choosing aptazymes, and that relaxation of the energetic parameters may lead to the identification of more active aptazymes. Even being able to consider such hypotheses is a large step forward in understanding aptazyme mechanisms, and this is one of the primary benefits of the computational method we have developed. It seems likely that as the algorithms and energetic models progress that predictive success and activation parameters will also improve.

While *in vitro* selection can yield an optimal solution to a binding problem, the method has several drawbacks. For example, frequently *in vitro* selections yield answers which might be optimal, but might not be the desired solution. When selecting nucleic acid aptamers against protein targets using nitrocellulose as a partitioning medium, aptamers are frequently generated that preferentially bind the filter over the target. Unfortunately, any selection process could potentially be exploited by these parasitic molecules, and eventually selections can be overrun with unwanted species. One impetus, then, for developing a purely computational approach is that the desired properties of an RNA molecule can be addressed and optimized *in silico* with no need for *in vitro* selection. Even now the computational sieving algorithm can be employed to quickly generate a set of aptazymes that would correspond to the large number of aptamers that have already been found (see, for example, the aptamer database at <http://aptamer.icmb.utexas.edu>).

EXPERIMENTAL METHODS

Computational methods

The free energies of RNA secondary structures were calculated using the ViennaRNA package (Hofacker et al., 1994), and scripts to automate the evaluation and comparison of RNA secondary structure profiles were coded in Python. These programs first generated nucleic acid sequences corresponding to clones from a pool in which an aptamer and a ribozyme were connected by a randomized joining region. For each sequence the free energy of folding was

calculated for a series of registers that could be assumed by its joining region, and those folding energies were plotted versus the register. The constructed energetic profiles were then compared to the energetic profile of an experimentally-verified, reference aptazyme. Variants were chosen that most closely matched the reference profile in terms of both the change between a stable inactive register and the active register, and the energetic difference between the inactive registers of compared profiles. The flowchart for this process is presented in greater detail in Figure 4.2. In the second iteration of the design algorithm, f_{mfe} values were calculated using the equation:

$$f_{mfe} = e^{\left(\frac{\Delta G_{pfe} - \Delta G_{mfe}}{kT}\right)}$$

where ΔG_{pfe} and ΔG_{mfe} are the partition and minimum free energies respectively, T is the temperature (300.15 K) and k is the Boltzmann constant.

***In vitro* transcription**

Double-stranded DNA templates used for transcription were made by primer extension of complementary synthetic DNA oligonucleotides (IDT, Coralville, IA). Radiolabeled RNA was generated by *in vitro* transcription using T7 RNA polymerase with trace amounts of [α - ^{32} P] UTP (Perkin-Elmer, Boston, MA), followed by purification on 10% denaturing polyacrylamide gels.

Ribozyme assays

Twenty picomoles of radiolabeled RNA were thermally equilibrated by denaturing at 70 °C in water and then cooling to room temperature. Buffer (50 mM Tris-HCl, pH 7.5, 20 mM MgCl₂, final concentration) and the appropriate

ligand (cGMP, Rev Peptide, Coenzyme A and FMN at 200 μ M, theophylline and ATP at 1 mM) were then added to a final volume of 20 μ L. Following incubation at room temperature, aliquots were removed at time points and quenched in stop dye (85% formamide, 0.01% bromophenol blue, 0.01% xylene cyanol and 60 mM EDTA, pH 8.0). Uncleaved and cleaved ribozymes were separated on 8% denaturing acrylamide gels, and bands were quantitated using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). Rate constants were calculated by plotting the negative natural logarithm of the fraction of uncleaved ribozyme versus time, and determining the slope of a line fitted through three time points. Assays were repeated twice or more and the observed levels of activation differed by less than two-fold.

REFERENCES

- Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science*, **181**, 223-230.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905-920.
- Benson, D.E., Wisz, M.S. and Hellinga, H.W. (2000) Rational design of nascent metalloenzymes. *Proc Natl Acad Sci U S A*, **97**, 6292-6297.
- Bonneau, R. and Baker, D. (2001) Ab initio protein structure prediction: progress and prospects. *Annu Rev Biophys Biomol Struct*, **30**, 173-189.
- Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C.E. and Baker, D. (2001) Rosetta in CASP4: progress in ab initio protein structure prediction. *Proteins*, **Suppl 5**, 119-126.
- Burke, D.H. and Hoffman, D.C. (1998) A novel acidophilic RNA motif that recognizes coenzyme A. *Biochemistry*, **37**, 4653-4663.

- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678-1685.
- Dahiyat, B.I. and Mayo, S.L. (1997) De novo protein design: fully automated sequence selection. *Science*, **278**, 82-87.
- Dieckmann, T., Suzuki, E., Nakamura, G.K. and Feigon, J. (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *Rna*, **2**, 628-640.
- Fan, P., Suri, A.K., Fiala, R., Live, D. and Patel, D.J. (1996) Molecular recognition in the FMN-RNA aptamer complex. *J Mol Biol*, **258**, 480-500.
- Ferre-D'Amare, A.R., Zhou, K. and Doudna, J.A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature*, **395**, 567-574.
- Giver, L., Bartel, D.P., Zapp, M.L., Green, M.R. and Ellington, A.D. (1993) Selection and design of high-affinity RNA ligands for HIV-1 Rev. *Gene*, **137**, 19-24.
- Gutell, R.R., Cannone, J.J., Konings, D. and Gautheret, D. (2000a) Predicting U-turns in ribosomal RNA with comparative sequence analysis. *J Mol Biol*, **300**, 791-803.
- Gutell, R.R., Cannone, J.J., Shang, Z., Du, Y. and Serra, M.J. (2000b) A story: unpaired adenosine bases in ribosomal RNAs. *J Mol Biol*, **304**, 335-354.
- Hartig, J.S., Najafi-Shoushtari, S.H., Grune, I., Yan, A., Ellington, A.D. and Famulok, M. (2002) Protein-dependent ribozymes report molecular interactions in real time. *Nat Biotechnol*, **20**, 717-722.
- Hellinga, H.W. (1998) The construction of metal centers in proteins by rational design. *Fold Des*, **3**, R1-8.
- Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M. and Schuster, P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, **125**, 167-188.
- Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425-1429.
- Jose, A.M., Soukup, G.A. and Breaker, R.R. (2001) Cooperative binding of effectors by an allosteric ribozyme. *Nucleic Acids Res*, **29**, 1631-1637.
- Koizumi, M., Soukup, G.A., Kerr, J.N. and Breaker, R.R. (1999) Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nat Struct Biol*, **6**, 1062-1071.

- Komatsu, Y., Nobuoka, K., Karino-Abe, N., Matsuda, A. and Ohtsuka, E. (2002) *In vitro* selection of hairpin ribozymes activated with short oligonucleotides. *Biochemistry*, **41**, 9090-9098.
- Lee, J.C., Cannone, J.J. and Gutell, R.R. (2003) The lonepair triloop: a new motif in RNA structure. *J Mol Biol*, **325**, 65-83.
- Mathews, D.H. and Turner, D.H. (2002) Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J Mol Biol*, **317**, 191-203.
- McCaskill, J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105-1119.
- Monroe, W.T. and Haselton, F.R. (2003) Molecular beacon sequence design algorithm. *Biotechniques*, **34**, 68-70, 72-63.
- Patel, D.J., Suri, A.K., Jiang, F., Jiang, L., Fan, P., Kumar, R.A. and Nonin, S. (1997) Structure, recognition and adaptive binding in RNA aptamer complexes. *J Mol Biol*, **272**, 645-664.
- Rupert, P.B. and Ferre-D'Amare, A.R. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, **410**, 780-786.
- Sassanfar, M. and Szostak, J.W. (1993) An RNA motif that binds ATP. *Nature*, **364**, 550-553.
- Soukup, G.A. and Breaker, R.R. (1999a) Engineering precision RNA molecular switches. *Proc Natl Acad Sci U S A*, **96**, 3584-3589.
- Soukup, G.A. and Breaker, R.R. (1999b) Relationship between internucleotide linkage geometry and the stability of RNA. *Rna*, **5**, 1308-1325.
- Soukup, G.A. and Breaker, R.R. (2000) Allosteric nucleic acid catalysts. *Curr Opin Struct Biol*, **10**, 318-325.
- Soukup, G.A., Emilsson, G.A. and Breaker, R.R. (2000) Altering molecular recognition of RNA aptamers by allosteric selection. *J Mol Biol*, **298**, 623-632.
- Tereshko, V., Skripkin, E. and Patel, D.J. (2003) Encapsulating Streptomycin within a Small 40-mer RNA. *Chem Biol*, **10**, 175-187.
- Vaish, N.K., Dong, F., Andrews, L., Schweppe, R.E., Ahn, N.G., Blatt, L. and Seiwert, S.D. (2002) Monitoring post-translational modification of proteins with allosteric ribozymes. *Nat Biotechnol*, **20**, 810-815.

- Wuchty, S., Fontana, W., Hofacker, I.L. and Schuster, P. (1999) Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**, 145-165.
- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. and Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883-896.
- Zimmermann, G.R., Wick, C.L., Shields, T.P., Jenison, R.D. and Pardi, A. (2000) Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *Rna*, **6**, 659-667.
- Zuker, M. (2000) Calculating nucleic acid secondary structure. *Curr Opin Struct Biol*, **10**, 303-310.
- Zwanzig, R., Szabo, A. and Bagchi, B. (1992) Levinthal's paradox. *Proc Natl Acad Sci U S A*, **89**, 20-22.

Chapter 5: Computational methods for the identification of riboswitches

INTRODUCTION

Riboswitches are a novel class of genetic regulatory element: they are RNA structures able to bind directly to small-molecule metabolites, causing changes in the genetic program. They are typically located in the 5' untranslated region (UTR) of an mRNA and bind the products of the proteins the mRNA encodes. For example, in the 5' UTR of the *thiCEFSGH* mRNA in *Escherichia coli* a riboswitch (the THI element) controls the fidelity with which the first gene is transcribed by participating in a terminator-antiterminator structural mechanism that is regulated by the concentrations of thiamine pyrophosphate (Mironov et al., 2002; Winkler et al., 2002a). Other examples of riboswitches include those that regulate the production of riboflavin (Vitreschak et al., 2002; Winkler et al., 2002b) and cobabamin (Vitamin B₁₂) (Nahvi et al., 2002), as well as the biosynthesis of S-adenosyl methionine (McDaniel et al., 2003).

Riboswitch elements have several diagnostic characteristics. Because they are functional RNAs, they preserve both sequence and structural elements. For example, the THI element has a characteristic structure that is well-conserved across several lineages (Rodionov et al., 2002). In addition, stretches of its primary sequence are conserved, most corresponding to the region of the element responsible for thiamine binding. Riboswitch elements are typically 80-150 nucleotides in length and this size is roughly characteristic for members of a given

riboswitch class. Finally, riboswitches have a defined phylogenetic distribution. The THI and RFN (riboflavin-specific) elements have been well characterized in terms of their conservation across organismal lineages. The THI element has been found in both the bacteria and Archaea (Rodionov et al., 2002), while the RFN element appears to be localized to the bacterial kingdom (Vitreschak et al., 2002).

All of the existing examples of riboswitch elements existed previously in the literature prior to their confirmation as bona fide riboswitches. For example, speculation previously arose that the THI element, which was initially identified upstream of thiamine biosynthesis genes in *Escherichia coli*, was solely responsible for the regulation of thiamine production in that bacteria (Miranda-Rios et al., 2001). Because no protein-based regulatory components were known, researchers had suggested that the mRNA itself was responsible for self-regulation by an unknown mechanism (Stormo and Ji, 2001). Biosynthesis of Vitamin B12 was apparently regulated by some unknown RNA-based mechanism (Nou and Kadner, 2000), and the RFN element was discovered following the study of bacterial riboflavin synthesis (Gelfand et al., 1999). Finally, Gram-positive bacteria have two classes of conserved sequences upstream of amino acid biosynthesis genes. The first class, the 'T-boxes' are regulated by an interaction with a tRNA molecule, and the charging state of the tRNA dictates the transcriptional fidelity of the downstream genes (Grundy et al., 2000). The S-box class appears very similar to the T-boxes, but no regulation was found in the presence of tRNA molecules (Grundy and Henkin, 1998). Recently it was

confirmed that the S-box sequences are riboswitches whose activity is regulated by the metabolite S-adenosyl methionine (McDaniel et al., 2003).

Here we describe several approaches taken in an effort to identify novel examples of riboswitch elements. We have taken advantage of this structural conservation in evaluating a search strategy that relies solely on the structural context of a sequence without regard to primary sequence conservation.

RESULTS AND DISCUSSION

Functional RNAs can be identified by structural signatures

Recently, a novel method for identifying functional RNAs has been described (Le et al., 2003). The method utilizes the RNA folding algorithm of Zuker (Zuker, 2000) in order to identify sequence elements that are 'well-folded', and thus have structural a signature. The first step of the algorithm is identical to the free energy minimizing fold in the Zuker algorithm. Importantly, all of the base-pairs involved in this minimum free-energy fold are noted. A second folding step is then done in which all of the base-pairs in the first set are disallowed. This step addresses what the authors term the 'optimally restrained structure': by requiring the sequence to fold into an alternate structure in which none of its native base pairs are formed, an assessment of a sequence's 'uniqueness' is made. Because each of the folding steps yields a ΔG_{fold} , the degree to which an RNA sequence is well-folded can be expressed as the difference between its ΔG_{fold} in both a restrained and unrestrained conformation ($\Delta G_{\text{fold}}^{\text{diff}} = \Delta G_{\text{fold}}^{\text{restrained}} - \Delta G_{\text{fold}}^{\text{unrestrained}}$). A quantitative score can then be derived for a given RNA

sequence by calculating a $\Delta G_{\text{fold}}^{\text{diff}}$ as a fixed-length window is slid across the sequence and then normalizing individual window scores by the average and standard deviation of the set of scores. Each score is thus expressed as $(\Delta G_{\text{fold}}^{\text{diff}} - \Delta G_{\text{fold}}^{\text{ave}}) / \Delta G_{\text{fold}}^{\text{stdev}}$, where $\Delta G_{\text{fold}}^{\text{ave}}$ and $\Delta G_{\text{fold}}^{\text{stdev}}$ are the mean and standard deviations of the individually calculated free energy differences. The authors implement their algorithm in a software package called 'ed-scan'.

The original report identifies several classes of functional RNA elements within the context of a larger RNA sequence. For example, the authors are able to distinguish structurally unique regions of the HIV genome that correspond to the RRE, TAR and Gag/Pol frameshifting elements, the iron-responsive element in the context of individual mRNAs and the binding sites for let-7 in the context of their mRNAs. Because this method appeared useful for the identification of functional RNAs at the structural level, we were interested in whether the algorithm could be applied to the identification of riboswitches.

Structural identification of *B. subtilis* riboswitches

The Gram-positive bacteria *Bacillus subtilis* possesses an abundance of structural RNA elements in the untranslated leaders of operonic transcripts. Some of these elements are *bona fide* riboswitches (McDaniel et al., 2003; Rodionov et al., 2002; Vitreschak et al., 2002; Winkler et al., 2002b), and others are very similar in spirit to known riboswitch elements. These additional elements, the T-boxes, lie upstream of amino acid biosynthetic and transport genes (Grundy and Henkin, 2003). They have well-conserved secondary structures but each uniquely

regulates a given set of genes. For example, the T-box element is frequently found upstream of tRNA synthetase genes, and the transcription termination mechanism used by the T-box is regulated by the charging state of the cognate tRNA (Grundy et al., 2000). In addition, two as-yet published but confirmed riboswitch elements are specific for lysine and the nucleobase guanine; these riboswitches are positioned in front of biosynthesis and transport genes (J. Barrick, Ronald Breaker lab, Yale University, personal communication).

Because this bacterial species contains a wealth of known RNA regulatory elements, we assessed the performance of ed-scan for the identification of these elements. Table 5.1 describes the chromosomal positions and annotations of the known structural RNA elements in *B. subtilis* (J. Barrick, Ronald Breaker Lab, Yale University, personal communication).

Description	Annotation^a	Strand	Ig Start^b	Ig End^c	Length^d
T Boxes (20)					
T Box (Ser)	serS	+	20557	20877	321
T Box (Met)	metS	+	45137	45630	494
T Box (Cys)	cysE	+	112496	112796	301
T Box (Unkown)	yczA	+	276751	277151	401
T Box (Trp)	trpS	-	1218578	1219320	743
T Box (Unkown)	proB	+	1377610	1377959	350
T Box (Ile)	ileS	+	1612417	1612757	341
T Box (Unkown)	yqjO	+	2471976	2472379	404
T Box (Gly)	glyQ	-	2607926	2608221	296
T Box (Ala)	alaS	-	2800081	2800410	330
T Box (His)	hisS	-	2817075	2817455	381
T Box (Val)	valS	-	2868568	2869063	496
T Box (Unkown)	ilvB	-	2896072	2896886	815
T Box (Phe)	pheS	-	2929575	2929928	354
T Box (Thr)	thrS	-	2960264	2960659	396
T Box (Tyr)	tyrS	-	3036947	3037287	341

T Box (Leu)	leuS	-	3104125	3104550	426
T Box (Unkown)	yvbW	-	3490317	3490717	401
T Box (Thr)	thrZ	-	3855225	3855815	591
T Box (Tyr)	tyrZ	+	3945704	3945951	248
Riboswitches (24)					
Lysine Box	yclN	-	431568	431951	384
N Box	ydhM	-	626120	626337	218
N Box	yebB	+	693801	694179	379
N Box	purE	+	697799	698120	322
Thi Box	thiA	+	955153	955402	250
S Box	yitJ	-	1180070	1180382	313
Thi Box	tenA	+	1241672	1241919	248
S Box	yjcl	+	1257608	1257962	355
S Box	metC	-	1385073	1385487	415
Thi Box	ykoF	-	1391108	1391417	310
S Box	ykrT	-	1423907	1424232	326
S Box	ykrW	+	1426304	1426499	196
Thi Box	ylmB	+	1606750	1606950	201
S Box	cysH	+	1629370	1629780	411
S Box	yoaD	-	2024332	2024654	323
N Box	xpt	-	2319267	2319596	330
RFN Element	ypaA	-	2409824	2410319	496
RFN Element	ypuE	-	2430712	2430969	258
S Box	metK	-	3128109	3128610	502
S Box	yusC	-	3363351	3363676	326
B12 Box	yvrC	-	3403496	3403892	397
S Box	yxjH	+	3996502	3996755	254
S Box	yxjG	+	3997779	3998140	362
N Box	yxjA	+	4003926	4004540	615

Table 5.1 Summary of RNA structural elements found in *Bacillus subtilis* (source: Jeffrey Barrick, Ronald Breaker Lab, Yale University).

- Annotated gene name 3' of the element
- Start coordinate of the intergenic region
- Stop coordinate of the intergenic region
- Length of the intergenic region

The ed-scan software was used in order to evaluate sequences corresponding to the mRNAs derived from **Table 5.1**. Using the previously reported window size of 50 (Le et al., 2003), several of the sequences were found to contain structurally unique regions in their leaders, and the regions were confirmed as being coincident with known riboswitch elements. An example scan is depicted in **Figure 5.1**. The peak at position ~400 corresponds to a structural element in the leader sequence that binds the nucleobases guanine and xanthine bases in order to regulate the translation of the subsequent genes (J. Barrick, Ronald Breaker Lab, Yale University, personal communication). Subsequent sequences were evaluated based on the presence of peaks in their upstream leaders that met a requisite score cutoff, and were positioned upstream of the annotated start of the coding region.

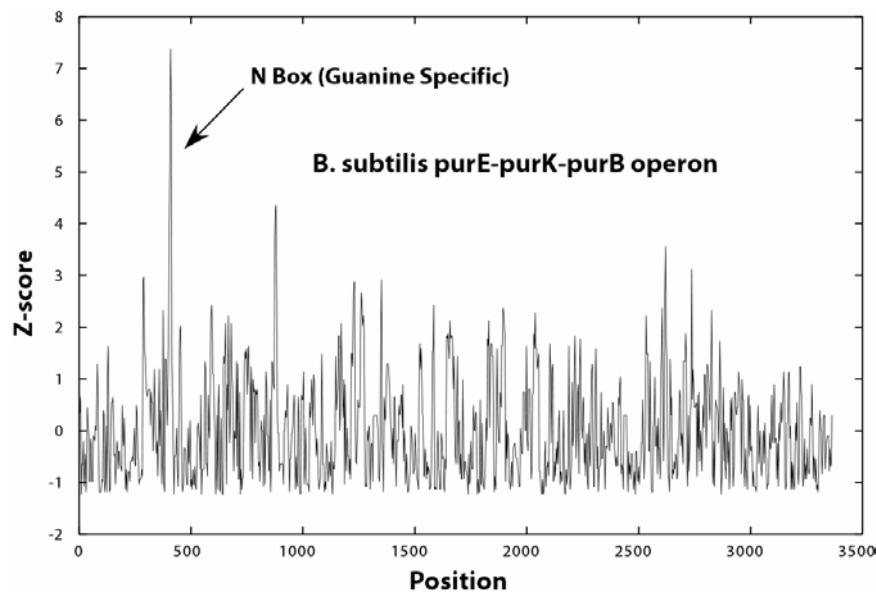


Figure 5.1. Ed-scan trace for the purE-purK-purB operon of *Bacillus subtilis*. Using a window size of 50, the ed-scan software was run across the the sequence

of the purE-purK-purB operon in addition to a 500 base leader sequence. The start codon of the operon lies at position 500.

Optimization of the ed-scan window size for functional RNA elements

Previously the size of the window used for ed-scan evaluations was empirically optimized for various structural RNA elements by systematically varying its length and finding the window size that yielded the highest score for a given sequence element (Le et al., 2003). We wished to perform the same type of analysis in order to tune this parameter for riboswitch elements, so we varied the window size used to scan a set of previously identified riboswitch-containing operons of *B. subtilis* from 30 to 240 nucleotides. Window sizes of 30-60 nucleotides consistently yielded the highest scores (Z-scores of 6-8) for peaks corresponding to RNA structural elements. One of the optimizations for the hisS-aspS operon of *B. subtilis* is depicted in **Figure 5.2**. Subsequent ed-scans were performed using a defined window size of 50 nucleotides.

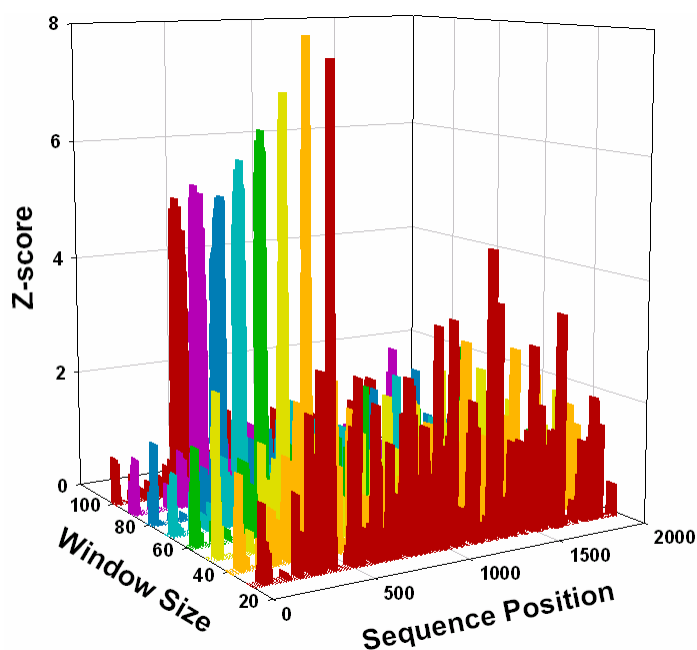


Figure 5.2. Optimization of the ed-scan window size for the *hisS-aspS* operon of *Bacillus subtilis*. The Z-scores for different window sizes are plotted versus sequence position. The start codon for the operon lies at 500 nt, and the T-box region is ~100 nucleotides upstream.

The scores for a set of sequences corresponding to mRNAs derived from *Bacillus subtilis* were evaluated (**Figure 5.3a**). mRNA sequences were extracted using a minimum cutoff between genes of 70 nucleotides, and individual sequences contained 600 bases of additional upstream sequence. In order to provide a reference, the sequences were also randomized and the scores for the shuffled set were collected. The distribution of scores for structurally conserved RNA elements is different from the distribution of the random sequences. After visual examination of this distribution, a minimum score of 6.0 was assigned. The ability of different scores to identify the positive controls was evaluated

(**Figure 5.3b**). Using a score of 4.0, nearly all of the controls were identified and using our cutoff of 6.0, ~30 % of the controls were identified.

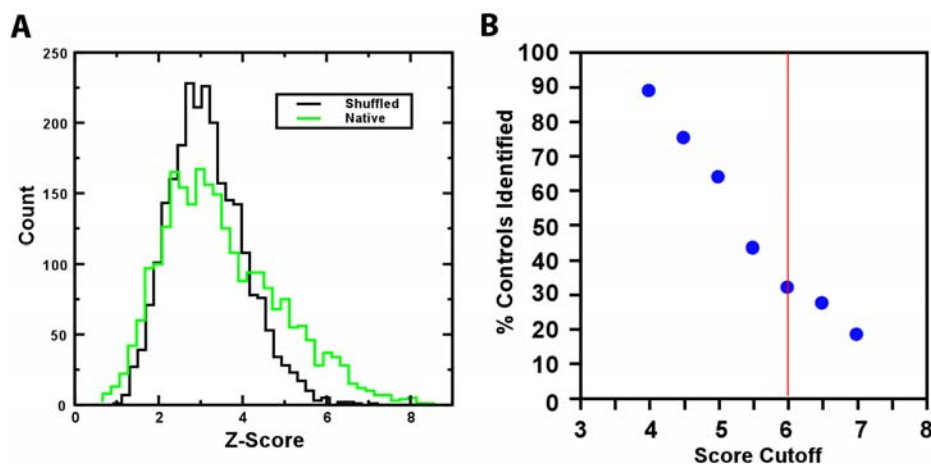


Figure 5.3. Comparison of score cutoff and control recovery. A. Distributions of *B. subtilis* ed-scan scores from shuffled and native sequences. The scores from each of 2487 shuffled and native operonic sequences from *Bacillus subtilis* were tallied. A minimum discriminatory score of 6 was assigned and used for subsequent calculations. B. Fraction of the 44 known *Bacillus subtilis* structural elements recovered as a function of score cutoff. The red line indicates the score cutoff that was used for evaluating unknown sequences.

Discovering structurally conserved RNA elements using orthology

Riboswitches are known to regulate genes that can be grouped together into 'clusters of orthology' (Rodionov et al., 2002; Tatusov et al., 2000). Using these groupings, we hypothesized that we might find novel riboswitch-regulated genes by looking for structurally conserved regions in front of genes found in orthologous groups. We used the NCBI COG groupings, which are a set of pre-defined orthologous proteins that are generated by all-against-all BLAST searches. In the method, proteins are assigned to groups (COGs) using the

assumption that any two orthologous proteins must be more similar to each other than to any proteins from other species (Tatusov et al., 2000). A cluster of orthology is then defined as a set of orthologous proteins with members from at least three different species.

A probabilistic model was developed in order to examine the distribution of confirmed structural RNAs and search for novel examples of these elements. The following steps were taken to identify novel elements from two classes of organisms, Gram positive (e.g. *Bacillus subtilis*) and Gram negative bacteria (e.g. *Escherichia coli*):

- I. Assign genes from each class of organism into COGs.
- II. Assemble genes into operons using a maximum intergenic distance of 70 nucleotides. Analyses of intergenic regions between genes in confirmed operons indicated that intergenic regions are on average 40 nucleotides (Salgado et al., 2000), so our groupings may be more promiscuous than what would be expected naturally.
- III. Use ed-scan to find structurally unique regions in front of operons derived from each organism. Identified regions must satisfy two requirements that are optimized empirically using known riboswitch elements:
 - a. The element lies a pre-defined distance from the start codon of the operon
 - b. The element achieves a requisite Z-score.

- IV. For each operon with a structurally unique region, determine the COG membership for the first gene in the operon.
- V. After tallying the scores, calculate the probability of finding this number of genes (COGs) containing structurally unique regions from a binomial distribution. The probability of finding a positive COG at random is the number of COGs determined to have structurally unique regions divided by the total number of COGs evaluated.
- VI. Rank the COGs by their calculated probabilities.

This method was used to evaluate 34,904 operons from 22 Gram-positive bacteria and 55,113 operons from 41 Gram-negative bacteria (NCBI RefSeq Genomes, <ftp.ncbi.nih.gov>). After application, 855 COGs in Gram positive and 532 COGs in Gram negative bacteria were found unlikely to occur by chance. **Table 5.2** summarizes some of the highest scoring COGs and their associated annotations.

Several of the identified COGs have previously described functions. For the Gram positive bacteria, 14 of the identified COGs correspond to previously confirmed structural RNA elements (the RFN, T-box and N-boxes). In addition, the method identifies regions associated with the regulation of glycerol-3-phosphate dehydrogenase (glpD, glpK) (Glatz et al., 1998), histidine utilization (Hut) (Delorme et al., 1999) and tryptophan biosynthesis (TRAP) (Babitzke and Gollnick, 2001). Several unexpected and potentially interesting genes were also present. For example, the *ilvGMEDA* gene in *E.coli* encodes the genes

responsible branched-chain amino acids. Studies of its transcript indicated that the leucine-responsive regulatory protein represses transcription of the operon, but sequences in the RNA are responsible for its premature transcriptional termination by an ill-defined mechanism (Rhee et al., 1996). Examples such as these are interesting leads which will be evaluated in light of the discovery they possess structurally conserved RNA elements.

Prob ^a	# COG Positive	# COG Seen	Annotation ^b	Note ^c	COG ID	COG Description
<i>Gram Positive</i>						
1.493E-06	7	8	glpD	GlpR	COG0578	Glycerol-3-phosphate dehydrogenase
5.971E-06	7	9	glyS	T-box	COG0752	Glycyl-tRNA synthetase, alpha subunit
0.0010263	8	21	No Annot.		COG3464	Transposase and inactivated derivatives
0.0013358	5	9	gatC		COG3775	Phosphotransferase system, galactitol-specific IIC component
0.0013358	5	9	miaA		COG0324	tRNA delta(2)-isopentenylpyrophosphate transferase
0.0013358	5	9	leuA		COG0119	Isopropylmalate/homocitrate/citramalate synthases
0.0014202	6	13	rncS		COG0571	dsRNA-specific ribonuclease
0.0014202	6	13	No Annot.		COG4886	Leucine-rich repeat (LRR) protein
0.0018105	4	6	No Annot.		COG4122	Predicted O-methyltransferase
0.0032723	6	15	pheS	T-box	COG0016	Phenylalanyl-tRNA synthetase alpha subunit
0.0037548	4	7	No Annot.		COG4932	Predicted outer membrane protein
0.0048858	3	4	mtlA		COG4668	Mannitol/fructose-specific phosphotransferase system, IIA domain
0.0058956	5	12	trpE	TRAP	COG0144	tRNA and rRNA cytosine-C5-methylases
0.0066748	4	8	addB		COG3857	ATP-dependent nuclease, subunit B
0.0071499	9	34	ydaA	RFN	COG3711	Transcriptional antiterminator
0.0085152	5	13	ytrA		COG1725	Predicted transcriptional regulators
0.0106788	4	9	yukE		COG4842	Uncharacterized protein conserved in bacteria
0.0106788	4	9	argC		COG0002	Acetylglutamate semialdehyde dehydrogenase
0.0106788	4	9	purC	N box	COG0152	Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase
0.0106788	4	9	glpK	GlpR	COG0554	Glycerol kinase
0.0108565	3	5	hisC	Hut	COG0079	Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase
0.0108565	3	5	nusB		COG0781	Transcription termination factor
0.0108565	3	5	rbsA		COG1129	ABC-type sugar transport system, ATPase component
0.0117732	5	14	proB	T-box	COG0263	Glutamate 5-kinase
0.0117732	5	14	uvrC		COG0322	Nuclease subunit of the excinuclease complex
<i>Gram Negative</i>						
2.629E-14	13	22	rpoB		COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit
6.804E-10	16	68	ilvG		COG0027	Formate-dependent phosphoribosylglycinamide formyltransferase (GAR)

						transformylase)
1.655E-09	7	9	yihI		COG3078	Uncharacterized protein conserved in bacteria
2.826E-08	9	23	purH	N-box	COG0138	AICAR transformylase/IMP cyclohydrolase PurH (only IMP cyclohydrolase domain in Aful)
1.737E-07	12	54	aceE		COG2609	Pyruvate dehydrogenase complex, dehydrogenase (E1) component
2.676E-07	6	10	yaeM		COG0743	1-deoxy-D-xylulose 5-phosphate reductoisomerase
4.138E-07	7	16	yihK		COG1217	Predicted membrane GTPase involved in stress response
4.138E-07	7	16	hisG	Hut	COG0040	ATP phosphoribosyltransferase
1.288E-06	4	4	No Annot.		COG3758	Uncharacterized protein conserved in bacteria
3.337E-06	6	14	murB		COG0812	UDP-N-acetylmuramate dehydrogenase
6.222E-06	4	5	yhbE		COG1664	Integral membrane protein CcmA involved in cell shape determination
8.308E-06	6	16	ebgA		COG3250	Beta-galactosidase/beta-glucuronidase
9.211E-06	5	10	apbA		COG1893	Ketopantoate reductase
9.211E-06	5	10	luxS / ygaG		COG1854	LuxS protein involved in autoinducer AI2 synthesis
4.743E-05	6	21	thrA		COG0460	Homoserine dehydrogenase
7.86E-05	4	8	hypF		COG0068	Hydrogenase maturation factor
7.86E-05	4	8	rfbB		COG1088	dTDP-D-glucose 4,6-dehydratase
8.629E-05	7	33	aldA		COG1012	NAD-dependent aldehyde dehydrogenases
9.248E-05	5	15	ptrB		COG1770	Protease II

Table 5.2. Summary of COGs containing structurally unique leader regions.

- Probability calculated from a binomial distribution (see experimental methods). The probability of encountering a match at random is the total number of positive COGs divided by the total number of COGs evaluated. For the Gram negative set, 28438 COGs were seen of which 958 were positive, and in the Gram positive set, 13240 COGs were seen of which 1472 were positive.
- First gene of the operon analyzed. Genes with no confirmed functions are assigned no annotations.
- Established regulatory mechanisms for the genes in the COGs.

Comparing local versus genomic scans

Because annotated sequence data is not always available, ORF boundaries and thus operon architecture may not be easily estimated. While the ed-scan performs well for sequences of defined length and modest size (the size of an operonic transcript, typically 1000-5000 nt), we wished to determine if ed-scan could also identify structurally unique regions in the context of a larger sequence region (i.e. an entire chromosome). To address this issue, the ed-scan program was used to scan a region of the *B. subtilis* chromosome (total length ~4.1 Mb) in order to identify well-folded sequences. Because of sequence length limitations of the software, the chromosome was divided into 300,000 nucleotide segments prior to scanning. A 300,000 nt region containing six known T-box elements is shown in **Figure 5.4**. These T-box elements are known to lie upstream of amino acid biosynthetic operons (Grundy et al., 2000), and we have previously confirmed that ed-scan identifies them unambiguously in the context of their operonic sequence. As can be seen in **Figure 5.4**, none of the T-box elements are identified as the most highly structured sequence regions. While they are easily distinguished in the context of local sequence (e.g. the valS T-box), they could not be unambiguously assigned as unique regions. Closer inspection of the alternate high-scoring regions indicates that the highest scoring regions are in sequences that would correspond to mRNA transcripts. It is possible that these particular mRNAs have optimized their structure for some other purpose such as transcriptional or translational fidelity. This method may be useful as an initial

scan for structurally unique regions of chromosome, which can further be refined by other means.

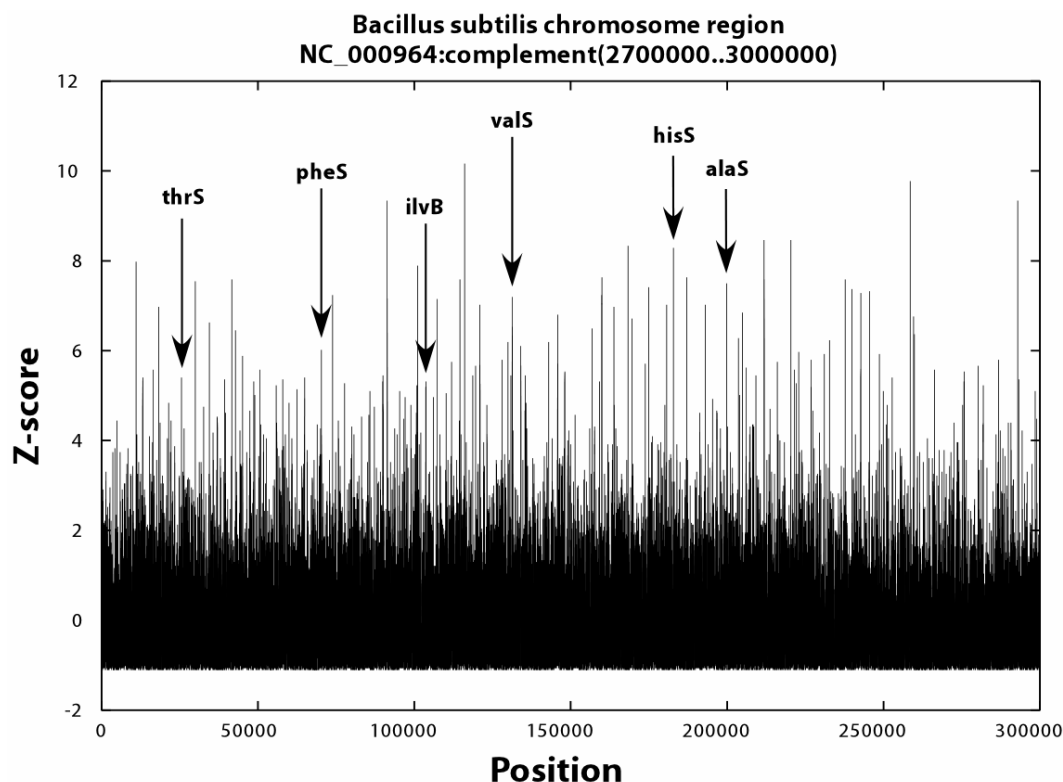


Figure 5.4. RNA structural elements are not distinguished in the context of unrelated sequence. A region of the *Bacillus subtilis* chromosome containing six T-box motifs was analyzed using ed-scan. The positions of each element are shown in the context of the sequence that was scanned.

Using conservation of sequence and position for identifying riboswitch elements

Each class of riboswitch (e.g. the THI box specific for thiamine-related genes) has strong primary sequence homology among its members. In addition, the position of known riboswitch elements is on the 5' side of the gene or operon it is regulating. Therefore we wished to determine if an approach utilizing these

shared features, sequence homology and positional conservation could be used in order to identify novel functional RNA elements. **Figure 5.5** summarizes approach we conceived in order to identify homologous primary sequences that were positionally conserved with respect to their downstream, homologous genes.

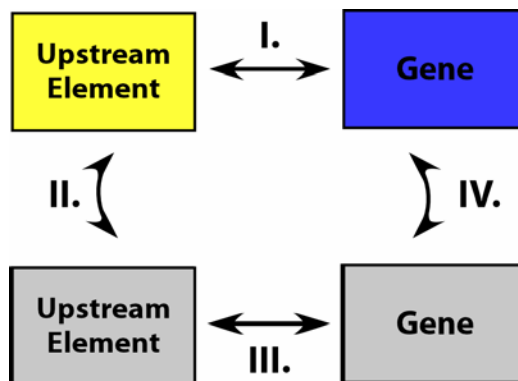


Figure 5.5. Algorithm for identifying homologous positionally conserved sequence elements. I. Intergenic regions of a defined length are extracted from a set of genes. II. The set of intergenic regions are compared using BLAST. III. Genes the lie next to homologous intergenic regions are identified. IV. The genes are compared with the original gene to see if they are either homologous or lie in an operon with a homologous gene.

This method was employed to search 113 organismal genomes (NCBI RefSeq Genomes). Following application, we counted the number of times a given protein and its homologs were identified as possessing conserved sequence elements near them. This method was determined to be unsuitable early in its incarnation for several reasons, the most severe of which is the inability of the BLAST algorithm to appropriately handle short stretches of conserved nucleic acid sequence that are separated by large, variable length gaps. This is a basic shortcoming of the BLAST algorithm, as it begins with seed alignments from which to build up a larger alignment. Presumably, because riboswitch elements

tend to have only short stretches (4-5 nucleotides) of primary sequence homology, elements do not score well against each other. As a test, we assembled the sequences corresponding to the THI box elements (Rodionov et al., 2002) into a database. Searching using any one of the sequences as a query identified a maximum of 8 of a total of ~50 related sequences in the database.

Because of the inability of BLAST to recognize riboswitch elements, a previously developed, powerful method for finding functional RNAs fails from the start. Sean Eddy and his colleagues previously developed an algorithm in which homologous RNA sequences are identified using BLAST and then applying a covariation algorithm in order to distinguish between three classifications of homologous sequence from different organisms: coding RNA, which preserves codon usage, structural RNA, which preserves base pairing, and a null model in which no obvious pattern is preserved (Rivas et al., 2001). This method, implemented in a software package called QRNA, was previously used to identify non-coding, functional RNAs conserved between *Escherichia coli* and three related organisms. In all, 275 sequences were identified, of which 11 were confirmed to be expressed as small RNAs of unknown function. The method itself is powerful, but fails for the identification of riboswitches because of the shortcomings of BLAST. In addition, other algorithms have been developed for the identification of short RNAs (sRNA) in bacteria, and several have converged on a subset of sRNA genes in *Escherichia coli* whose expression has been confirmed (Hershberg et al., 2003).

Other sequence motif discovery algorithms have been employed for the identification of conserved sequence elements upstream of homologous genes from a number of organisms. George Church and his colleagues have pioneered the use of Gibbs sampling in order to identify highly conserved sequences from several bacteria that lie upstream of genes grouped by different methods (i.e. phylogenetic profiles and orthology) (McGuire et al., 2000). These methods are able to identify previously characterized operator sequences, for example, in addition to several novel classes of DNA regulatory sequences. One important filter they used to filter their data sets was the presence of palindromicity in highly conserved sequences (McGuire et al., 2000), which they use to bias their discovery toward symmetric DNA operator sequences. In addition, they limit their search window to a few (~50) nucleotides. While this may be reasonable for the identification of transcription factor binding sites, it may miss larger sequence motifs like those found in riboswitches. Other statistical approaches have been used in order to identify conserved sequence motifs. Eric Siggia and his colleagues have been successful identifying statistically over-represented sequences across a set of bacterial genomes that conformed to a particular pattern (Li et al., 2002). Again, this method is powerful for the identification of transcription factor binding sites that conform to a pattern, but may fail for larger sequence motifs having unknown patterns. Interestingly, one of their highest scoring sequence motifs is the THI box motif, illustrating that this approach may be a useful first step in identifying small sequence motifs that can be subsequently expanded into larger matches.

Structure versus sequence based identification

The ed-scan algorithm correctly identifies a large number of known riboswitch elements and has identified additional candidates. The power of this method stems from its separation from the constraints of a sequence-based search. The primary sequence is a transparent factor in the algorithm, used only to describe a given structure, which is then evaluated. This has several advantages, the most important being it eliminates the dependence on sequence comparison (e.g. BLAST) which we have already determined to be unsuitable for the identification of large, sparsely homologous nucleic acid elements. In addition, the structure based search can be tuned to allow for the discrimination of elements of a given size. For example, the highest scores were obtained for T-box elements were obtained when the algorithm's window size was set at 50 nucleotides (**Figure 5.2**). As additional riboswitch elements are identified, their identification by the ed-scan algorithm can be optimized by adjusting this parameter.

Phylogenetic distribution of riboswitches

It is interesting to consider how riboswitch elements may have evolved. One mechanism in which riboswitches may have been distributed is by horizontal transfer. Riboswitch elements are extremely well-conserved: specific examples are found across several genetic boundaries up to and including the kingdom classification. If horizontal transfer were the mechanism behind riboswitch evolution, the elements must have existed very early on because they are well

conserved. If, for example, examples of the same riboswitch could be found in all organisms, horizontal transfer must have occurred very early, perhaps back to the earliest organism. However, few examples of existing RNA regulatory elements argue against this hypothesis. In Gram-positive bacteria, there are a set of RNA elements (the T-boxes, see Grundy et al., 2000) that regulate the expression of amino acid biosynthesis and transport genes. The distribution of T-boxes is very specific; no bacteria outside of the Firmicutes phylum possess these elements. On the other hand, some riboswitch elements are absent from organisms in which they might be expected. As an example, the THI element is conspicuously absent from *Lactococcus lactis*, while it is possessed by nearly all the other Gram-positive organisms (Rodionov et al., 2002). Therefore, it might be unreasonable to suspect that all riboswitches and riboswitch-like elements originated early in an evolutionary time scale.

As an alternative, riboswitches may be able to arise *de novo*, through a process which could be described as 'genomic selection': much like the *in vitro* selection of aptamers and ribozymes, one can envision that an organism has the ability to identify sequences capable of binding small metabolites based on, for example, a growth advantage conferred by the sequence. This hypothesis would suggest that riboswitches would be widespread, and indeed in several cases they are profligate (e.g. the THI element is found in nearly all prokaryotes). Several empirical examples are relevant to this hypothesis. One important finding from the *in vitro* selection of aptamers has come to be called the 'tyranny of small motifs' (Jhaveri et al., 1997). Simply stated, this means that the smallest sequence

capable of performing a function will be selected preferentially over larger sequences, mainly due to the fidelity with which the sequence can be replicated. It is unreasonable to suspect that small sequences are not capable of binding metabolites, and in fact RNA and DNA aptamers have been raised against scores of small molecules (Famulok, 1999). On the other hand, riboswitch elements tend to be large (hundreds of nucleotides in length), much larger than sequences that are perfectly capable of binding small molecules. The regulation mechanisms employed for regulation by riboswitches typically involve a conformational equilibrium between the riboswitch element and a terminator / anti-terminator structure. These additional sequence requirements exaggerate the size of riboswitch elements, but critical residues involved in the conformational changes associated with ligand binding can be over 100 nucleotides apart in primary sequence (see Figure 1 in Nahvi et al., 2002). Perhaps the discrepancy between size and function can be resolved by considering that the large size of riboswitches might be necessary to confer genetic regulatory capacity, enabling their propagation as an evolutionarily conserved entity.

The phylogenetic distribution of riboswitches is not well-characterized. Other than two reports describing the conservation of the THI (Rodionov et al., 2002) and RFN (Vitreschak et al., 2002) elements, little is known with regard to the global conservation of individual classes of riboswitches. Recently, the COG database was used to illustrate that particular classes of proteins have phylogenetic distributions similar to ribosomal RNA, implying that the proteins evolved identically to ribosomal RNA and thus may have been present in the last

universal common ancestor (Harris et al., 2003). A similar analysis using riboswitch elements specific for particular COG groupings may yield insight into how riboswitches arose and evolved with respect to the genes they regulate or compared with ribosomal RNAs.

CONCLUSIONS

The development of computational tools for the discovery of novel RNA regulatory elements is imperative as more genomic data is generated. Because these elements remained undiscovered for so long, the aspects of their function and phylogenetic conservation remain mostly uncharacterized. By employing a structure-based search strategy, we were able to successfully identify previously confirmed riboswitch elements and estimate the accuracy of the method. In addition, several examples of other structured mRNA leaders were identified, in addition to a set of previously undescribed conserved structural regions in orthologous mRNAs.

EXPERIMENTAL METHODS

Computer software

The ed-scan software was obtained from the author's FTP site (Le et al., 2003). A few minor modifications were made to the source code for purposes of platform compatibility. These changes had no effect on the outcome of the test case included in the distribution. NCBI BLAST (version 2.2.4) was obtained at

from the NCBI ftp site (<ftp://ftp.ncbi.nih.gov>). All scripts for analysis were coded in Python (www.python.org).

Comparing ed-scan scores with COGs

In order to develop a scoring system for the evaluation of sequences identified by the ed-scan software, gene groupings according to the COG database were employed to identify COG groupings that possessed an over-representation of high scoring sequences. Sequences corresponding to mRNAs derived from Gram-positive, Gram-negative and Archaea were assembled assuming a maximum distance of 70 nucleotides between adjacent genes. In addition to coding sequence, upstream UTRs were appended by including a fixed number (600) of bases upstream of the start codon of the first gene in the operon. Sequences were then evaluated using ed-scan with a window of 50 nucleotides. Sequences were then identified that possessed a defined set of characteristics. Sequences that scored above a prerequisite threshold and were positioned a predetermined distance from the start codon were identified. An additional constraint was applied in which the upstream regions had to meet a minimum length, determined by the placement of the upstream gene on the chromosome. Once sequence regions that met the above criteria were identified, the gene downstream was noted and its COG membership determined. The probability of getting a certain number of high-scoring regions from a given COG was then calculated from a binomial distribution, in which the probability of encountering a high-scoring COG was the number of observed high-scoring COGs divided by the

total number of COGs noted. Probabilities were then calculated from a binomial distribution using the equation:

$$\binom{N}{k} p^k (1-p)^{(N-k)}$$

where N is the total number of an individual COG observed, k is the number of individual, positive COGs observed and p is the probability of encountering a positive COG at random (total positive COGs / total COGs observed). COGs were then ranked by their calculated probabilities and candidates were subsequently classified based on previous studies.

REFERENCES

- Babitzke, P. and Gollnick, P. (2001) Posttranscription initiation control of tryptophan metabolism in *Bacillus subtilis* by the trp RNA-binding attenuation protein (TRAP), anti-TRAP, and RNA structure. *J Bacteriol*, **183**, 5795-5802.
- Delorme, C., Ehrlich, S.D. and Renault, P. (1999) Regulation of expression of the *Lactococcus lactis* histidine operon. *J Bacteriol*, **181**, 2026-2037.
- Famulok, M. (1999) Oligonucleotide aptamers that recognize small molecules. *Curr Opin Struct Biol*, **9**, 324-329.
- Gelfand, M.S., Mironov, A.A., Jomantas, J., Kozlov, Y.I. and Perumov, D.A. (1999) A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet*, **15**, 439-442.
- Glatz, E., Farewell, A. and Rutberg, B. (1998) The *Bacillus subtilis* glpD leader and antiterminator protein GlpP provide a target for glucose repression in *Escherichia coli*. *FEMS Microbiol Lett*, **162**, 93-96.
- Grundy, F.J., Collins, J.A., Rollins, S.M. and Henkin, T.M. (2000) tRNA determinants for transcription antitermination of the *Bacillus subtilis* tyrS gene. *Rna*, **6**, 1131-1141.
- Grundy, F.J. and Henkin, T.M. (1998) The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Mol Microbiol*, **30**, 737-749.

- Grundy, F.J. and Henkin, T.M. (2003) The T box and s box transcription termination control systems. *Front Biosci*, **8**, D20-31.
- Harris, J.K., Kelley, S.T., Spiegelman, G.B. and Pace, N.R. (2003) The genetic core of the universal ancestor. *Genome Res*, **13**, 407-412.
- Hershberg, R., Altuvia, S. and Margalit, H. (2003) A survey of small RNA-encoding genes in Escherichia coli. *Nucleic Acids Res*, **31**, 1813-1820.
- Jhaveri, S.D., Hirao, I., Bell, S., Uphoff, K.W. and Ellington, A.D. (1997) Landscapes for molecular evolution: lessons from *in vitro* selection experiments with nucleic acids. *Annual Reports in Combinatorial Chemistry and Molecular Diversity*, **1**, 169-191.
- Le, S.Y., Chen, J.H., Konings, D. and Maizel, J.V., Jr. (2003) Discovering well-ordered folding patterns in nucleotide sequences. *Bioinformatics*, **19**, 354-361.
- Li, H., Rhodius, V., Gross, C. and Siggia, E.D. (2002) Identification of the binding sites of regulatory proteins in bacterial genomes. *Proc Natl Acad Sci U S A*, **99**, 11772-11777.
- McDaniel, B.A., Grundy, F.J., Artsimovitch, I. and Henkin, T.M. (2003) Transcription termination control of the S box system: Direct measurement of S-adenosylmethionine by the leader RNA. *Proc Natl Acad Sci U S A*, **100**, 3083-3088.
- McGuire, A.M., Hughes, J.D. and Church, G.M. (2000) Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome Res*, **10**, 744-757.
- Miranda-Rios, J., Navarro, M. and Soberon, M. (2001) A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proc Natl Acad Sci U S A*, **98**, 9736-9741.
- Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A. and Nudler, E. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, **111**, 747-756.
- Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L. and Breaker, R.R. (2002) Genetic control by a metabolite binding mRNA. *Chem Biol*, **9**, 1043.
- Nou, X. and Kadner, R.J. (2000) Adenosylcobalamin inhibits ribosome binding to *btuB* RNA. *Proc Natl Acad Sci U S A*, **97**, 7190-7195.
- Rhee, K.Y., Parekh, B.S. and Hatfield, G.W. (1996) Leucine-responsive regulatory protein-DNA interactions in the leader region of the *ilvGMEDA* operon of Escherichia coli. *J Biol Chem*, **271**, 26499-26507.

- Rivas, E., Klein, R.J., Jones, T.A. and Eddy, S.R. (2001) Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol*, **11**, 1369-1373.
- Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. and Gelfand, M.S. (2002) Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *J Biol Chem*, **277**, 48949-48959.
- Salgado, H., Moreno-Hagelsieb, G., Smith, T.F. and Collado-Vides, J. (2000) Operons in *Escherichia coli*: genomic analyses and predictions. *Proc Natl Acad Sci U S A*, **97**, 6652-6657.
- Stormo, G.D. and Ji, Y. (2001) Do mRNAs act as direct sensors of small molecules to control their expression? *Proc Natl Acad Sci U S A*, **98**, 9465-9467.
- Sudarsanam, P., Pilpel, Y. and Church, G.M. (2002) Genome-wide co-occurrence of promoter elements reveals a cis-regulatory cassette of rRNA transcription motifs in *Saccharomyces cerevisiae*. *Genome Res*, **12**, 1723-1731.
- Tatusov, R.L., Galperin, M.Y., Natale, D.A. and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*, **28**, 33-36.
- Vitreschak, A.G., Rodionov, D.A., Mironov, A.A. and Gelfand, M.S. (2002) Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res*, **30**, 3141-3151.
- Winkler, W., Nahvi, A. and Breaker, R.R. (2002a) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952-956.
- Winkler, W.C., Cohen-Chalamish, S. and Breaker, R.R. (2002b) An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci U S A*, **99**, 15908-15913.
- Zuker, M. (2000) Calculating nucleic acid secondary structure. *Curr Opin Struct Biol*, **10**, 303-310.

Appendix

The computational RNA folding work described in Chapter 4 was done with the Vienna RNA package. The version of the software as of the time of this writing is 1.4. However, as the energetic parameters used by the package will undoubtedly be updated in the future, the committee overseeing this work expressed some concern that that parameters used in this work be made available. As such, a copy of the current version of the software is available at <http://www.tbi.univie.ac.at/RNA/> or the version used in this work is available at <http://ellingtonlab.org/software>.

Bibliography

- Albert, K.J., Walt, D.R., Gill, D.S. and Pearce, T.C. (2001) Optical multibead arrays for simple and complex odor discrimination. *Anal Chem*, **73**, 2501-2508.
- Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science*, **181**, 223-230.
- Arnon, S.S., Schechter, R., Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Fine, A.D., Hauer, J., Layton, M., Lillibridge, S., Osterholm, M.T., O'Toole, T., Parker, G., Perl, T.M., Russell, P.K., Swerdlow, D.L. and Tonat, K. (2001) Botulinum toxin as a biological weapon: medical and public health management. *JAMA*, **285**, 1059-1070. FIELD Reference Number:1106 FIELD Journal Code:7501160 FIELD Call Number:.
- Babitzke, P. and Gollnick, P. (2001) Posttranscription initiation control of tryptophan metabolism in *Bacillus subtilis* by the trp RNA-binding attenuation protein (TRAP), anti-TRAP, and RNA structure. *J Bacteriol*, **183**, 5795-5802.
- Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R.A., Sweet, R., Frank, J., Moore, P.B. and Steitz, T.A. (1998) A 9 Å resolution X-ray crystallographic map of the large ribosomal subunit. *Cell*, **93**, 1105-1115.
- Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P.B. and Steitz, T.A. (1999) Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature*, **400**, 841-847.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905-920.
- Banerjee, D. and Slack, F. (2002) Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. *Bioessays*, **24**, 119-129.
- Bartel, D.P., Zapp, M.L., Green, M.R. and Szostak, J.W. (1991) HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell*, **67**, 529-536.

- Baskerville, S., Zapp, M. and Ellington, A.D. (1995) High-resolution mapping of the human T-cell leukemia virus type 1 Rex-binding element by *in vitro* selection. *J Virol*, **69**, 7559-7569.
- Baugh, C., Grate, D. and Wilson, C. (2000) 2.8 Å crystal structure of the malachite green aptamer. *J Mol Biol*, **301**, 117-128.
- Bell, S.D., Denu, J.M., Dixon, J.E. and Ellington, A.D. (1998) RNA molecules that bind to and inhibit the active site of a tyrosine phosphatase. *J Biol Chem*, **273**, 14309-14314.
- Benson, D.E., Wisz, M.S. and Hellinga, H.W. (2000) Rational design of nascent metalloenzymes. *Proc Natl Acad Sci U S A*, **97**, 6292-6297.
- Blind, M., Kolanus, W. and Famulok, M. (1999) Cytoplasmic RNA modulators of an inside-out signal-transduction cascade. *Proc Natl Acad Sci U S A*, **96**, 3606-3610.
- Bonneau, R. and Baker, D. (2001) Ab initio protein structure prediction: progress and prospects. *Annu Rev Biophys Biomol Struct*, **30**, 173-189.
- Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C.E. and Baker, D. (2001) Rosetta in CASP4: progress in ab initio protein structure prediction. *Proteins*, **Suppl 5**, 119-126.
- Brody, E.N. and Gold, L. (2000) Aptamers as therapeutic and diagnostic agents. *J Biotechnol*, **74**, 5-13.
- Brody, E.N., Willis, M.C., Smith, J.D., Jayasena, S., Zichi, D. and Gold, L. (1999) The use of aptamers in large arrays for molecular diagnostics. *Mol Diagn*, **4**, 381-388.
- Burgstaller, P. and Famulok, M. (1994) Isolation of RNA aptamers for biological cofactors by *in vitro* selection. *Angewandte Chemie*, **106**, 1163-1166 (See also *Angew. Chem., Int. Ed. Engl.*, 1994, 1133(1110), 1084-1167).
- Burgstaller, P. and Famulok, M. (1996) Structural characterization of a flavin-specific RNA aptamer by chemical probing. *Bioorganic & Medicinal Chemistry Letters*, **6**, 1157-1162.
- Burgstaller, P., Girod, A. and Blind, M. (2002) Aptamers as tools for target prioritization and lead identification. *Drug Discov Today*, **7**, 1221-1228.

- Burke, D.H. and Hoffman, D.C. (1998) A novel acidophilic RNA motif that recognizes coenzyme A. *Biochemistry*, **37**, 4653-4663.
- Carmell, M.A., Xuan, Z., Zhang, M.Q. and Hannon, G.J. (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev*, **16**, 2733-2742.
- Caruthers, M.H., Barone, A.D., Beaucage, S.L., Dodds, D.R., Fisher, E.F., McBride, L.J., Matteucci, M., Stabinsky, Z. and Tang, J.Y. (1987) Chemical synthesis of deoxyoligonucleotides by the phosphoramidite method. *Methods in Enzymology*, **154**, 287-313.
- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678-1685.
- Chen, X.-Y., Berti, P.J. and Schramm, V.L. (2000) Ricin A-Chain: Kinetic Isotope Effects and Transition State Structure with Stem-Loop RNA. *Journal of the American Chemical Society*, **122**, 1609-1617.
- Chen, X.-Y., Berti, P.J. and Schramm, V.L. (2000) Transition-State Analysis for Depurination of DNA by Ricin A-Chain. *Journal of the American Chemical Society*, **122**, 6527-6534.
- Chen, X.Y., Link, T.M. and Schramm, V.L. (1998) Ricin A-chain: kinetics, mechanism, and RNA stem-loop inhibitors. *Biochemistry*, **37**, 11605-11613.
- Chen, X.-Y., Link, T.M. and Schramm, V.L. (1996) Inhibition of Ricin by an RNA Stem-Loop Containing a Ribo-Oxocarbenium Mimic. *Journal of the American Chemical Society*, **118**, 3067-3068.
- Christodoulides, N., Tran, M., Floriano, P.N., Rodriguez, M., Goodey, A., Ali, M., Neikirk, D. and McDevitt, J.T. (2002) A microchip-based multianalyte assay system for the assessment of cardiac risk. *Anal Chem*, **74**, 3030-3036.
- Christopher, G.W., Cieslak, T.J., Pavlin, J.A. and Eitzen, E.M., Jr. (1997) Biological warfare. A historical perspective. *Jama*, **278**, 412-417.
- Cieslak, T.J., Christopher, G.W., Kortepeter, M.G., Rowe, J.R., Pavlin, J.A., Culpepper, R.C. and Eitzen, E.M., Jr. (2000) Immunization against potential biological warfare agents. *Clin Infect Dis*, **30**, 843-850.

- Conn, M.M., Prudent, J.R. and Schultz, P.G. (1996) Porphyrin Metalation Catalyzed by a Small RNA Molecule. *Journal of the American Chemical Society*, **118**, 7012-7013.
- Conrad, R.C., Baskerville, S. and Ellington, A.D. (1995) *In vitro* selection methodologies to probe RNA function and structure. *Mol Divers*, **1**, 69-78.
- Conrad, R.C., Giver, L., Tian, Y. and Ellington, A.D. (1996) *In vitro* selection of nucleic acid aptamers that bind proteins. *Methods Enzymol*, **267**, 336-367.
- Convery, M.A., Rowsell, S., Stonehouse, N.J., Ellington, A.D., Hirao, I., Murray, J.B., Peabody, D.S., Phillips, S.E. and Stockley, P.G. (1998) Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution. *Nat Struct Biol*, **5**, 133-139.
- Coombs, G.S., Rao, M.S., Olson, A.J., Dawson, P.E. and Madison, E.L. (1999) Revisiting catalysis by chymotrypsin family serine proteases using peptide substrates and inhibitors with unnatural main chains. *J Biol Chem*, **274**, 24074-24079.
- Correll, C.C., Munishkin, A., Chan, Y.L., Ren, Z., Wool, I.G. and Steitz, T.A. (1998) Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc Natl Acad Sci U S A*, **95**, 13436-13441.
- Cox, J.C. and Ellington, A.D. (2001) Automated selection of anti-protein aptamers. *Bioorg Med Chem*, **9**, 2525-2531.
- Cox, J.C., Hayhurst, A., Hesselberth, J., Bayer, T.S., Georgiou, G. and Ellington, A.D. (2002) Automated selection of aptamers against protein targets translated *in vitro*: from gene to aptamer. *Nucleic Acids Res*, **30**, e108.
- Cox, J.C., Rudolph, P. and Ellington, A.D. (1998) Automated RNA selection. *Biotechnol Prog*, **14**, 845-850.
- Curey, T.E., Goodey, A., Tsao, A., Lavigne, J., Sohn, Y., McDevitt, J.T., Anslyn, E.V., Neikirk, D. and Shear, J.B. (2001) Characterization of multicomponent monosaccharide solutions using an enzyme-based sensor array. *Anal Biochem*, **293**, 178-184.
- Dahiyat, B.I. and Mayo, S.L. (1997) De novo protein design: fully automated sequence selection. *Science*, **278**, 82-87.

- Day, P.J., Ernst, S.R., Frankel, A.E., Monzingo, A.F., Pascal, J.M., Molina-Svinth, M.C. and Robertus, J.D. (1996) Structure and activity of an active site substitution of ricin A chain. *Biochemistry*, **35**, 11098-11103.
- de Wildt, R.M., Mundy, C.R., Gorick, B.D. and Tomlinson, I.M. (2000) Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat Biotechnol*, **18**, 989-994.
- Delehanty, J.B. and Ligler, F.S. (2002) A microarray immunoassay for simultaneous detection of proteins and bacteria. *Anal Chem*, **74**, 5681-5687.
- Delorme, C., Ehrlich, S.D. and Renault, P. (1999) Regulation of expression of the *Lactococcus lactis* histidine operon. *J Bacteriol*, **181**, 2026-2037.
- Dennis, D.T., Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Fine, A.D., Friedlander, A.M., Hauer, J., Layton, M., Lillibridge, S.R., McDade, J.E., Osterholm, M.T., O'Toole, T., Parker, G., Perl, T.M., Russell, P.K. and Tonat, K. (2001) Tularemia as a biological weapon: medical and public health management. *JAMA*, **285**, 2763-2773.
- DeRisi, J.L., Iyer, V.R. and Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680-686.
- Diamond, J.M., Turner, D.H. and Mathews, D.H. (2001) Thermodynamics of three-way multibranch loops in RNA. *Biochemistry*, **40**, 6971-6981.
- Dickinson, T.A., Michael, K.L., Kauer, J.S. and Walt, D.R. (1999) Convergent, self-encoded bead sensor arrays in the design of an artificial nose. *Anal Chem*, **71**, 2192-2198.
- Dieckmann, T., Suzuki, E., Nakamura, G.K. and Feigon, J. (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *Rna*, **2**, 628-640.
- Drolet, D.W., Moon-McDermott, L. and Romig, T.S. (1996) An enzyme-linked oligonucleotide assay. *Nat Biotechnol*, **14**, 1021-1025.
- Du, Z. and Hoffman, D.W. (1997) An NMR and mutational study of the pseudoknot within the gene 32 mRNA of bacteriophage T2: insights into a family of structurally related RNA pseudoknots. *Nucleic Acids Res*, **25**, 1130-1135.

- Ekland, E.H. and Bartel, D.P. (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res*, **23**, 3231-3238.
- Ekland, E.H., Szostak, J.W. and Bartel, D.P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science*, **269**, 364-370.
- Ellington, A.D. and Szostak, J.W. (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818-822.
- Ellington, A.D. and Szostak, J.W. (1992) Selection *in vitro* of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature*, **355**, 850-852.
- Endo, Y., Gluck, A. and Wool, I.G. (1991) Ribosomal RNA identity elements for ricin A-chain recognition and catalysis. *J Mol Biol*, **221**, 193-207.
- Endo, Y., Mitsui, K., Motizuki, M. and Tsurugi, K. (1987) The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28 S ribosomal RNA caused by the toxins. *J Biol Chem*, **262**, 5908-5912.
- Endo, Y. and Tsurugi, K. (1987) RNA N-glycosidase activity of ricin A-chain. Mechanism of action of the toxic lectin ricin on eukaryotic ribosomes. *J Biol Chem*, **262**, 8128-8130.
- Endo, Y. and Tsurugi, K. (1988) The RNA N-glycosidase activity of ricin A-chain. *Nucleic Acids Symp Ser*, 139-142.
- Ermolaeva, M.D., Khalak, H.G., White, O., Smith, H.O. and Salzberg, S.L. (2000) Prediction of transcription terminators in bacterial genomes. *J Mol Biol*, **301**, 27-33.
- Famulok, M. (1999) Oligonucleotide aptamers that recognize small molecules. *Curr Opin Struct Biol*, **9**, 324-329.
- Famulok, M. and Jenne, A. (1998) Oligonucleotide libraries--variatio delectat. *Curr Opin Chem Biol*, **2**, 320-327.
- Fan, P., Suri, A.K., Fiala, R., Live, D. and Patel, D.J. (1996) Molecular recognition in the FMN-RNA aptamer complex. *J Mol Biol*, **258**, 480-500.

- Ferre-D'Amare, A.R., Zhou, K. and Doudna, J.A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature*, **395**, 567-574.
- Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gustafsdottir, S.M., Ostman, A. and Landegren, U. (2002) Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol*, **20**, 473-477.
- Fritz, J., Baller, M.K., Lang, H.P., Rothuizen, H., Vettiger, P., Meyer, E., Guntherodt, H., Gerber, C. and Gimzewski, J.K. (2000) Translating biomolecular recognition into nanomechanics. *Science*, **288**, 316-318.
- Fritz, J., Cooper, E.B., Gaudet, S., Sorger, P.K. and Manalis, S.R. (2002) Electronic detection of DNA by its intrinsic molecular charge. *Proc Natl Acad Sci U S A*, **99**, 14142-14146.
- Gelfand, M.S., Mironov, A.A., Jomantas, J., Kozlov, Y.I. and Perumov, D.A. (1999) A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet*, **15**, 439-442.
- Gesteland, R.F. and Atkins, J.F. (eds.). (1993) *The RNA World*. Cold Spring Harbor Laboratory Press.
- Gilbert, W. (1986) The RNA world. *Nature*, **319**, 618.
- Giver, L., Bartel, D., Zapp, M., Pawul, A., Green, M. and Ellington, A.D. (1993) Selective optimization of the Rev-binding element of HIV-1. *Nucleic Acids Res*, **21**, 5509-5516.
- Giver, L., Bartel, D.P., Zapp, M.L., Green, M.R. and Ellington, A.D. (1993) Selection and design of high-affinity RNA ligands for HIV-1 Rev. *Gene*, **137**, 19-24.
- Glatz, E., Farewell, A. and Rutberg, B. (1998) The *Bacillus subtilis* glpD leader and antiterminator protein GlpP provide a target for glucose repression in *Escherichia coli*. *FEMS Microbiol Lett*, **162**, 93-96.
- Gold, L., Polisky, B., Uhlenbeck, O. and Yarus, M. (1995) Diversity of oligonucleotide functions. *Annu Rev Biochem*, **64**, 763-797.
- Good, P.D., Krikos, A.J., Li, S.X., Bertrand, E., Lee, N.S., Giver, L., Ellington, A., Zaia, J.A., Rossi, J.J. and Engelke, D.R. (1997) Expression of small, therapeutic RNAs in human cell nuclei. *Gene Ther*, **4**, 45-54.

- Goodey, A., Lavigne, J.J., Savoy, S.M., Rodriguez, M.D., Curey, T., Tsao, A., Simmons, G., Wright, J., Yoo, S.J., Sohn, Y., Anslyn, E.V., Shear, J.B., Neikirk, D.P. and McDevitt, J.T. (2001) Development of multianalyte sensor arrays composed of chemically derivatized polymeric microspheres localized in micromachined cavities. *Journal of the American Chemical Society*, **123**, 2559-2570.
- Goodey, A.P. and McDevitt, J.T. (2003) Multishell microspheres with integrated chromatographic and detection layers for use in array sensors. *J Am Chem Soc*, **125**, 2870-2871.
- Green, L.S., Jellinek, D., Bell, C., Beebe, L.A., Feistner, B.D., Gill, S.C., Jucker, F.M. and Janjic, N. (1995) Nuclease-resistant nucleic acid ligands to vascular permeability factor/vascular endothelial growth factor. *Chem Biol*, **2**, 683-695.
- Grundy, F.J., Collins, J.A., Rollins, S.M. and Henkin, T.M. (2000) tRNA determinants for transcription antitermination of the *Bacillus subtilis* tyrS gene. *Rna*, **6**, 1131-1141.
- Grundy, F.J. and Henkin, T.M. (1998) The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Mol Microbiol*, **30**, 737-749.
- Grundy, F.J. and Henkin, T.M. (2003) The T box and s box transcription termination control systems. *Front Biosci*, **8**, D20-31.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849-857.
- Guo, F. and Cech, T.R. (2002) Evolution of *Tetrahymena* ribozyme mutants with increased structural stability. *Nat Struct Biol*, **9**, 855-861.
- Gutell, R.R., Cannone, J.J., Konings, D. and Gautheret, D. (2000) Predicting U-turns in ribosomal RNA with comparative sequence analysis. *J Mol Biol*, **300**, 791-803.
- Gutell, R.R., Cannone, J.J., Shang, Z., Du, Y. and Serra, M.J. (2000) A story: unpaired adenosine bases in ribosomal RNAs. *J Mol Biol*, **304**, 335-354.
- Hamaguchi, N., Ellington, A. and Stanton, M. (2001) Aptamer beacons for the direct detection of proteins. *Anal Biochem*, **294**, 126-131.

- Harris, J.K., Kelley, S.T., Spiegelman, G.B. and Pace, N.R. (2003) The genetic core of the universal ancestor. *Genome Res*, **13**, 407-412.
- Hartig, J.S., Najafi-Shoushtari, S.H., Grune, I., Yan, A., Ellington, A.D. and Famulok, M. (2002) Protein-dependent ribozymes report molecular interactions in real time. *Nat Biotechnol*, **20**, 717-722.
- Hellings, H.W. (1998) The construction of metal centers in proteins by rational design. *Fold Des*, **3**, R1-8.
- Henderson, D.A., Inglesby, T.V., Bartlett, J.G., Ascher, M.S., Eitzen, E., Jahrling, P.B., Hauer, J., Layton, M., McDade, J., Osterholm, M.T., O'Toole, T., Parker, G., Perl, T., Russell, P.K. and Tonat, K. (1999) Smallpox as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. *JAMA*, **281**, 2127-2137.
- Henkin, T.M. (2000) Transcription termination control in bacteria. *Curr Opin Microbiol*, **3**, 149-153.
- Henkin, T.M. and Yanofsky, C. (2002) Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *Bioessays*, **24**, 700-707.
- Hermann, T. and Patel, D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820-825.
- Hershberg, R., Altuvia, S. and Margalit, H. (2003) A survey of small RNA-encoding genes in Escherichia coli. *Nucleic Acids Res*, **31**, 1813-1820.
- Hesselberth, J., Robertson, M.P., Jhaveri, S. and Ellington, A.D. (2000) *In vitro* selection of nucleic acids for diagnostic applications. *J Biotechnol*, **74**, 15-25.
- Hesselberth, J.R. and Ellington, A.D. (2002) A (ribo) switch in the paradigms of genetic regulation. *Nat Struct Biol*, **9**, 891-893.
- Hesselberth, J.R., Miller, D., Robertus, J. and Ellington, A.D. (2000) *In vitro* selection of RNA molecules that inhibit the activity of ricin A-chain. *J Biol Chem*, **275**, 4937-4942.
- Hesselberth, J.R., Robertson, M.P., Knudsen, S.M. and Ellington, A.D. (2003) Simultaneous detection of diverse analytes with an aptazyme ligase array. *Anal Biochem*, **312**, 106-112.

- Hirao, I., Madin, K., Endo, Y., Yokoyama, S. and Ellington, A.D. (2000) RNA aptamers that bind to and inhibit the ribosome-inactivating protein, pepocin. *J Biol Chem*, **275**, 4943-4948.
- Hirao, I., Spingola, M., Peabody, D. and Ellington, A.D. (1998) The limits of specificity: an experimental analysis with RNA aptamers to MS2 coat protein variants. *Mol Divers*, **4**, 75-89.
- Hirao, I., Yoshinari, S., Yokoyama, S., Endo, Y. and Ellington, A.D. (1997) *In vitro* selection of aptamers that bind to ribosome-inactivating toxins. *Nucleic Acids Symp Ser*, 283-284.
- Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M. and Schuster, P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, **125**, 167-188.
- Homann, M. and Goring, H.U. (1999) Combinatorial selection of high affinity RNA ligands to live African trypanosomes. *Nucleic Acids Res*, **27**, 2006-2014.
- <http://www.bt.cdc.gov/Agent/agentlist.asp>. (2003) Bioterrorism Preparedness and Response Planning. Centers for Disease Control and Prevention.
- Inglesby, T.V., Dennis, D.T., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Fine, A.D., Friedlander, A.M., Hauer, J., Koerner, J.F., Layton, M., McDade, J., Osterholm, M.T., O'Toole, T., Parker, G., Perl, T.M., Russell, P.K., Schoch-Spana, M. and Tonat, K. (2000) Plague as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. *JAMA*, **283**, 2281-2290.
- Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Friedlander, A.M., Hauer, J., McDade, J., Osterholm, M.T., O'Toole, T., Parker, G., Perl, T.M., Russell, P.K. and Tonat, K. (1999) Anthrax as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. *JAMA*, **281**, 1735-1745.
- Jellinek, D., Green, L.S., Bell, C. and Janjic, N. (1994) Inhibition of receptor binding by high-affinity RNA ligands to vascular endothelial growth factor. *Biochemistry*, **33**, 10450-10456.
- Jellinek, D., Lynott, C.K., Rifkin, D.B. and Janjic, N. (1993) High-affinity RNA ligands to basic fibroblast growth factor inhibit receptor binding. *Proc Natl Acad Sci U S A*, **90**, 11227-11231.

- Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425-1429.
- Jenne, A., Hartig, J.S., Piganeau, N., Tauer, A., Samarsky, D.A., Green, M.R., Davies, J. and Famulok, M. (2001) Rapid identification and characterization of hammerhead-ribozyme inhibitors using fluorescence-based technology. *Nat Biotechnol*, **19**, 56-61.
- Jhaveri, S., Olwin, B. and Ellington, A.D. (1998) *In vitro* selection of phosphorothiolated aptamers. *Bioorganic & Medicinal Chemistry Letters*, **8**, 2285-2290.
- Jhaveri, S., Rajendran, M. and Ellington, A.D. (2000) *In vitro* selection of signaling aptamers. *Nat Biotechnol*, **18**, 1293-1297.
- Jhaveri, S.D., Hirao, I., Bell, S., Uphoff, K.W. and Ellington, A.D. (1997) Landscapes for molecular evolution: lessons from *in vitro* selection experiments with nucleic acids. *Annual Reports in Combinatorial Chemistry and Molecular Diversity*, **1**, 169-191.
- Jhaveri, S.D., Kirby, R., Conrad, R., Maglott, E.J., Bowser, M., Kennedy, R.T., Glick, G. and Ellington, A.D. (2000) Designed Signaling Aptamers that Transduce Molecular Recognition to Changes in Fluorescence Intensity. *Journal of the American Chemical Society*, **122**, 2469-2473.
- Jose, A.M., Soukup, G.A. and Breaker, R.R. (2001) Cooperative binding of effectors by an allosteric ribozyme. *Nucleic Acids Res*, **29**, 1631-1637.
- Kauffman, S. and Levin, S. (1987) Towards a general theory of adaptive walks on rugged landscapes. *J Theor Biol*, **128**, 11-45.
- Kim, Y., Mlsna, D., Monzingo, A.F., Ready, M.P., Frankel, A. and Robertus, J.D. (1992) Structure of a ricin mutant showing rescue of activity by a noncatalytic residue. *Biochemistry*, **31**, 3294-3296.
- Kjems, J., Frankel, A.D. and Sharp, P.A. (1991) Specific regulation of mRNA splicing *in vitro* by a peptide from HIV-1 Rev. *Cell*, **67**, 169-178.
- Knight, R. and Yarus, M. (2003) Finding specific RNA motifs: Function in a zeptomole world? *Rna*, **9**, 218-230.
- Koizumi, M. and Breaker, R.R. (2000) Molecular recognition of cAMP by an RNA aptamer. *Biochemistry*, **39**, 8983-8992.

- Koizumi, M., Kerr, J.N., Soukup, G.A. and Breaker, R.R. (1999) Allosteric ribozymes sensitive to the second messengers cAMP and cGMP. *Nucleic Acids Symp Ser*, **42**, 275-276.
- Koizumi, M., Soukup, G.A., Kerr, J.N. and Breaker, R.R. (1999) Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nat Struct Biol*, **6**, 1062-1071.
- Komatsu, Y., Nobuoka, K., Karino-Abe, N., Matsuda, A. and Ohtsuka, E. (2002) *In vitro* selection of hairpin ribozymes activated with short oligonucleotides. *Biochemistry*, **41**, 9090-9098.
- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147-157.
- Larsson, S.L., Sloma, M.S. and Nygard, O. (2002) Conformational changes in the structure of domains II and V of 28S rRNA in ribosomes treated with the translational inhibitors ricin or alpha-sarcin. *Biochim Biophys Acta*, **1577**, 53-62.
- Laskowski, M., Jr. and Kato, I. (1980) Protein inhibitors of proteinases. *Annu Rev Biochem*, **49**, 593-626.
- Le, S.Y., Chen, J.H., Konings, D. and Maizel, J.V., Jr. (2003) Discovering well-ordered folding patterns in nucleotide sequences. *Bioinformatics*, **19**, 354-361.
- Lee, J.C., Cannone, J.J. and Gutell, R.R. (2003) The lonepair triloop: a new motif in RNA structure. *J Mol Biol*, **325**, 65-83.
- Lee, M. and Walt, D.R. (2000) A fiber-optic microarray biosensor using aptamers as receptors. *Anal Biochem*, **282**, 142-146.
- Lee, N., Bessho, Y., Wei, K., Szostak, J.W. and Suga, H. (2000) Ribozyme-catalyzed tRNA aminoacylation. *Nat Struct Biol*, **7**, 28-33.
- Levy, M. and Ellington, A.D. (2002) ATP-dependent allosteric DNA enzymes. *Chem Biol*, **9**, 417-426.
- Li, H., Rhodius, V., Gross, C. and Siggia, E.D. (2002) Identification of the binding sites of regulatory proteins in bacterial genomes. *Proc Natl Acad Sci U S A*, **99**, 11772-11777.

- Li, J.J., Fang, X. and Tan, W. (2002) Molecular aptamer beacons for real-time protein recognition. *Biochem Biophys Res Commun*, **292**, 31-40.
- Li, Y. and Sen, D. (1996) A catalytic DNA for porphyrin metallation. *Nat Struct Biol*, **3**, 743-747.
- Lord, J.M., Hartley, M.R. and Roberts, L.M. (1991) Ribosome inactivating proteins of plants. *Semin Cell Biol*, **2**, 15-22.
- Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, **288**, 911-940.
- Mathews, D.H. and Turner, D.H. (2002) Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J Mol Biol*, **317**, 191-203.
- Mathews, D.H. and Turner, D.H. (2002) Experimentally derived nearest-neighbor parameters for the stability of RNA three- and four-way multibranch loops. *Biochemistry*, **41**, 869-880.
- McCaskill, J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105-1119.
- McDaniel, B.A., Grundy, F.J., Artsimovitch, I. and Henkin, T.M. (2003) Transcription termination control of the S box system: Direct measurement of S-adenosylmethionine by the leader RNA. *Proc Natl Acad Sci U S A*.
- McDonnell, J.M. (2001) Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. *Curr Opin Chem Biol*, **5**, 572-577.
- McGall, G.H. and Fidanza, J.A. (2001) Photolithographic synthesis of high-density oligonucleotide arrays. *Methods Mol Biol*, **170**, 71-101.
- McGuire, A.M., Hughes, J.D. and Church, G.M. (2000) Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome Res*, **10**, 744-757.
- Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res*, **15**, 8783-8798.

- Miranda-Rios, J., Navarro, M. and Soberon, M. (2001) A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proc Natl Acad Sci U S A*, **98**, 9736-9741.
- Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A. and Nudler, E. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, **111**, 747-756.
- Mlsna, D., Monzingo, A.F., Katzin, B.J., Ernst, S. and Robertus, J.D. (1993) Structure of recombinant ricin A chain at 2.3 Å. *Protein Sci*, **2**, 429-435.
- Monroe, W.T. and Haselton, F.R. (2003) Molecular beacon sequence design algorithm. *Biotechniques*, **34**, 68-70, 72-63.
- Moore, P.B. and Steitz, T.A. (2003) After the ribosome structures: How does peptidyl transferase work? *Rna*, **9**, 155-159.
- Munishkin, A. and Wool, I.G. (1997) The ribosome-in-pieces: binding of elongation factor EF-G to oligoribonucleotides that mimic the sarcin/ricin and thiostrepton domains of 23S ribosomal RNA. *Proc Natl Acad Sci U S A*, **94**, 12280-12284.
- Muth, G.W., Ortoleva-Donnelly, L. and Strobel, S.A. (2000) A single adenosine with a neutral pKa in the ribosomal peptidyl transferase center. *Science*, **289**, 947-950.
- Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L. and Breaker, R.R. (2002) Genetic control by a metabolite binding mRNA. *Chem Biol*, **9**, 1043.
- Nou, X. and Kadner, R.J. (2000) Adenosylcobalamin inhibits ribosome binding to btuB RNA. *Proc Natl Acad Sci U S A*, **97**, 7190-7195.
- Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., McCormick, M., Norton, J., Pollock, T., Sumwalt, T., Butcher, L., Porter, D., Molla, M., Hall, C., Blattner, F., Sussman, M.R., Wallace, R.L., Cerrina, F. and Green, R.D. (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res*, **12**, 1749-1755.
- Osborne, S.E. and Ellington, A.D. (1997) Nucleic Acid Selection and the Challenge of Combinatorial Chemistry. *Chem Rev*, **97**, 349-370.

- Osborne, S.E., Matsumura, I. and Ellington, A.D. (1997) Aptamers as therapeutic and diagnostic reagents: problems and prospects. *Curr Opin Chem Biol*, **1**, 5-9.
- Patel, D.J., Suri, A.K., Jiang, F., Jiang, L., Fan, P., Kumar, R.A. and Nonin, S. (1997) Structure, recognition and adaptive binding in RNA aptamer complexes. *J Mol Biol*, **272**, 645-664.
- Porta, H. and Lizardi, P.M. (1995) An allosteric hammerhead ribozyme. *Biotechnology (N Y)*, **13**, 161-164.
- Qin, P.Z. and Pyle, A.M. (1999) Site-specific labeling of RNA with fluorophores and other structural probes. *Methods*, **18**, 60-70.
- Ready, M.P., Kim, Y. and Robertus, J.D. (1991) Site-directed mutagenesis of ricin A-chain and implications for the mechanism of action. *Proteins*, **10**, 270-278.
- Reinhart, B.J. and Bartel, D.P. (2002) Small RNAs correspond to centromere heterochromatic repeats. *Science*, **297**, 1831.
- Rezaie, A.R. (1998) Elucidation of the structural basis for the slow reactivity of thrombin with plasminogen activator inhibitor-1. *Biochemistry*, **37**, 13138-13142.
- Rhee, K.Y., Parekh, B.S. and Hatfield, G.W. (1996) Leucine-responsive regulatory protein-DNA interactions in the leader region of the *ilvGMEDA* operon of *Escherichia coli*. *J Biol Chem*, **271**, 26499-26507.
- Rivas, E., Klein, R.J., Jones, T.A. and Eddy, S.R. (2001) Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol*, **11**, 1369-1373.
- Robertson, D.L. and Joyce, G.F. (1990) Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, **344**, 467-468.
- Robertson, M.P. (2001) Engineered regulation of an RNA ligase ribozyme. *Molecular and Cell Biology*. University of Texas, Austin, TX, p. 197.
- Robertson, M.P. and Ellington, A.D. (1999) *In vitro* selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat Biotechnol*, **17**, 62-66.

- Robertson, M.P. and Ellington, A.D. (2000) Design and optimization of effector-activated ribozyme ligases. *Nucleic Acids Res*, **28**, 1751-1759.
- Robertson, M.P. and Ellington, A.D. (2001) *In vitro* selection of nucleoprotein enzymes. *Nat Biotechnol*, **19**, 650-655.
- Robertson, M.P., Hesselberth, J.R. and Ellington, A.D. (2001) Optimization and optimality of a short ribozyme ligase that joins non-Watson-Crick base pairings. *Rna*, **7**, 513-523.
- Robertus, J.D., Yan, X., Ernst, S., Monzingo, A., Worley, S., Day, P., Hollis, T. and Svinth, M. (1996) Structural analysis of ricin and implications for inhibitor design. *Toxicon*, **34**, 1325-1334.
- Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. and Gelfand, M.S. (2002) Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *J Biol Chem*, **277**, 48949-48959.
- Rowe, C.A., Tender, L.M., Feldstein, M.J., Golden, J.P., Scruggs, S.B., MacCraith, B.D., Cras, J.J. and Ligler, F.S. (1999) Array biosensor for simultaneous identification of bacterial, viral, and protein analytes. *Anal Chem*, **71**, 3846-3852.
- Rowe-Taitt, C.A., Golden, J.P., Feldstein, M.J., Cras, J.J., Hoffman, K.E. and Ligler, F.S. (2000) Array biosensor for detection of biohazards. *Biosens Bioelectron*, **14**, 785-794.
- Rowe-Taitt, C.A., Hazzard, J.W., Hoffman, K.E., Cras, J.J., Golden, J.P. and Ligler, F.S. (2000) Simultaneous detection of six biohazardous agents using a planar waveguide array biosensor. *Biosens Bioelectron*, **15**, 579-589.
- Ruffner, D.E., Stormo, G.D. and Uhlenbeck, O.C. (1990) Sequence requirements of the hammerhead RNA self-cleavage reaction. *Biochemistry*, **29**, 10695-10702.
- Rupert, P.B. and Ferre-D'Amare, A.R. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, **410**, 780-786.
- Rusconi, C.P., Scardino, E., Layzer, J., Pitoc, G.A., Ortel, T.L., Monroe, D. and Sullenger, B.A. (2002) RNA aptamers as reversible antagonists of coagulation factor IXa. *Nature*, **419**, 90-94.

- Russell, R. and Herschlag, D. (1999) New pathways in folding of the Tetrahymena group I RNA enzyme. *J Mol Biol*, **291**, 1155-1167.
- Rutenber, E., Katzin, B.J., Ernst, S., Collins, E.J., Mlsna, D., Ready, M.P. and Robertus, J.D. (1991) Crystallographic refinement of ricin to 2.5 Å. *Proteins*, **10**, 240-250.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491.
- Salgado, H., Moreno-Hagelsieb, G., Smith, T.F. and Collado-Vides, J. (2000) Operons in Escherichia coli: genomic analyses and predictions. *Proc Natl Acad Sci U S A*, **97**, 6652-6657.
- Santoro, S.W. and Joyce, G.F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proc Natl Acad Sci U S A*, **94**, 4262-4266.
- Santoro, S.W. and Joyce, G.F. (1998) Mechanism and utility of an RNA-cleaving DNA enzyme. *Biochemistry*, **37**, 13330-13342.
- Sassanfar, M. and Szostak, J.W. (1993) An RNA motif that binds ATP. *Nature*, **364**, 550-553.
- Schlutzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. and Yonath, A. (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, **102**, 615-623.
- Schwarz, D.S. and Zamore, P.D. (2002) Why do miRNAs live in the miRNP? *Genes Dev*, **16**, 1025-1031.
- Schweitzer, B., Roberts, S., Grimwade, B., Shao, W., Wang, M., Fu, Q., Shu, Q., Laroche, I., Zhou, Z., Tchernev, V.T., Christiansen, J., Velleca, M. and Kingsmore, S.F. (2002) Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nat Biotechnol*, **20**, 359-365.
- Seetharaman, S., Zivarts, M., Sudarsan, N. and Breaker, R.R. (2001) Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nat Biotechnol*, **19**, 336-341.

- Sharp, T.W., Brennan, R.J., Keim, M., Williams, R.J., Eitzen, E. and Lillibridge, S. (1998) Medical preparedness for a terrorist incident involving chemical or biological agents during the 1996 Atlanta Olympic Games. *Ann of Emerg Med*, **32**, 214-223.
- Sooter, L.J., Riedel, T., Davidson, E.A., Levy, M., Cox, J.C. and Ellington, A.D. (2001) Toward automated nucleic acid enzyme selection. *Biol Chem*, **382**, 1327-1334.
- Soukup, G.A. and Breaker, R.R. (1999) Design of allosteric hammerhead ribozymes activated by ligand-induced structure stabilization. *Structure Fold Des*, **7**, 783-791.
- Soukup, G.A. and Breaker, R.R. (1999) Engineering precision RNA molecular switches. *Proc Natl Acad Sci U S A*, **96**, 3584-3589.
- Soukup, G.A. and Breaker, R.R. (1999) Nucleic acid molecular switches. *Trends Biotechnol*, **17**, 469-476.
- Soukup, G.A. and Breaker, R.R. (1999) Relationship between internucleotide linkage geometry and the stability of RNA. *Rna*, **5**, 1308-1325.
- Soukup, G.A. and Breaker, R.R. (2000) Allosteric nucleic acid catalysts. *Curr Opin Struct Biol*, **10**, 318-325.
- Soukup, G.A., DeRose, E.C., Koizumi, M. and Breaker, R.R. (2001) Generating new ligand-binding RNAs by affinity maturation and disintegration of allosteric ribozymes. *Rna*, **7**, 524-536.
- Soukup, G.A., Emilsson, G.A. and Breaker, R.R. (2000) Altering molecular recognition of RNA aptamers by allosteric selection. *J Mol Biol*, **298**, 623-632.
- Stojanovic, M.N., de Prada, P. and Landry, D.W. (2001) Aptamer-based folding fluorescent sensor for cocaine. *J Am Chem Soc*, **123**, 4928-4931.
- Stormo, G.D. and Ji, Y. (2001) Do mRNAs act as direct sensors of small molecules to control their expression? *Proc Natl Acad Sci U S A*, **98**, 9465-9467.
- Sudarsanam, P., Pilpel, Y. and Church, G.M. (2002) Genome-wide co-occurrence of promoter elements reveals a cis-regulatory cassette of rRNA

- transcription motifs in *Saccharomyces cerevisiae*. *Genome Res*, **12**, 1723-1731.
- Sun, L., Cui, Z., Gottlieb, R.L. and Zhang, B. (2002) A selected ribozyme catalyzing diverse dipeptide synthesis. *Chem Biol*, **9**, 619-628.
- Sussman, D., Nix, J.C. and Wilson, C. (2000) The structural basis for molecular recognition by the vitamin B 12 RNA aptamer. *Nat Struct Biol*, **7**, 53-57.
- Symensma, T.L., Baskerville, S., Yan, A. and Ellington, A.D. (1999) Polyvalent Rev decoys act as artificial Rev-responsive elements. *J Virol*, **73**, 4341-4349.
- Symensma, T.L., Giver, L., Zapp, M., Takle, G.B. and Ellington, A.D. (1996) RNA aptamers selected to bind human immunodeficiency virus type 1 Rev *in vitro* are Rev responsive *in vivo*. *J Virol*, **70**, 179-187.
- Symons, R.H. (1992) Small catalytic RNAs. *Annu Rev Biochem*, **61**, 641-671.
- Szewczak, A.A., Moore, P.B., Chang, Y.L. and Wool, I.G. (1993) The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc Natl Acad Sci U S A*, **90**, 9581-9585.
- Taitt, C.R., Anderson, G.P., Lingerfelt, B.M., Feldstein, M.J. and Ligler, F.S. (2002) Nine-analyte detection using an array-based biosensor. *Anal Chem*, **74**, 6114-6120.
- Tanaka, K.S.E., Chen, X.-Y., Ichikawa, Y., Tyler, P.C., Furneaux, R.H. and Schramm, V.L. (2001) Ricin A-Chain Inhibitors Resembling the Oxacarbenium Ion Transition State. *Biochemistry*, **40**, 6845-6851.
- Tang, J. and Breaker, R.R. (1997) Rational design of allosteric ribozymes. *Chem Biol*, **4**, 453-459.
- Tang, J. and Breaker, R.R. (1998) Mechanism for allosteric inhibition of an ATP-sensitive ribozyme. *Nucleic Acids Res*, **26**, 4214-4221.
- Tarasow, T.M., Tarasow, S.L. and Eaton, B.E. (1997) RNA-catalysed carbon-carbon bond formation. *Nature*, **389**, 54-57.
- Tatusov, R.L., Galperin, M.Y., Natale, D.A. and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*, **28**, 33-36.

- Tereshko, V., Skripkin, E. and Patel, D.J. (2003) Encapsulating Streptomycin within a Small 40-mer RNA. *Chem Biol*, **10**, 175-187.
- Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.
- Tuerk, C., MacDougall, S. and Gold, L. (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci U S A*, **89**, 6988-6992.
- Vaish, N.K., Dong, F., Andrews, L., Schweppe, R.E., Ahn, N.G., Blatt, L. and Seiwert, S.D. (2002) Monitoring post-translational modification of proteins with allosteric ribozymes. *Nat Biotechnol*, **20**, 810-815.
- van de Locht, A., Bode, W., Huber, R., Le Bonniec, B.F., Stone, S.R., Esmon, C.T. and Stubbs, M.T. (1997) The thrombin E192Q-BPTI complex reveals gross structural rearrangements: implications for the interaction with antithrombin and thrombomodulin. *Embo J*, **16**, 2977-2984.
- Vitreschak, A.G., Rodionov, D.A., Mironov, A.A. and Gelfand, M.S. (2002) Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res*, **30**, 3141-3151.
- Walter, A.E., Turner, D.H., Kim, J., Lyttle, M.H., Muller, P., Mathews, D.H. and Zuker, M. (1994) Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc Natl Acad Sci U S A*, **91**, 9218-9222.
- Wang, D.Y., Lai, B.H. and Sen, D. (2002) A general strategy for effector-mediated control of RNA-cleaving ribozymes and DNA enzymes. *J Mol Biol*, **318**, 33-43.
- Wang, D.Y. and Sen, D. (2002) Rationally designed allosteric variants of hammerhead ribozymes responsive to the HIV-1 Tat protein. *Comb Chem High Throughput Screen*, **5**, 301-312.
- Wassarman, K.M. (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell*, **109**, 141-144.
- Weeks, K.M. and Cech, T.R. (1995) Efficient protein-facilitated splicing of the yeast mitochondrial bI5 intron. *Biochemistry*, **34**, 7728-7738.

- Wiener, S.L. (1996) Strategies for the prevention of a successful biological warfare aerosol attack. *Mil Med*, **161**, 251-256.
- Winkler, W., Nahvi, A. and Breaker, R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952-956.
- Winkler, W.C., Cohen-Chalamish, S. and Breaker, R.R. (2002) An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci U S A*, **99**, 15908-15913.
- Winssinger, N., Ficarro, S., Schultz, P.G. and Harris, J.L. (2002) Profiling protein function with small molecule microarrays. *Proc Natl Acad Sci U S A*, **99**, 11139-11144.
- Wuchty, S., Fontana, W., Hofacker, I.L. and Schuster, P. (1999) Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**, 145-165.
- Xu, W. and Ellington, A.D. (1996) Anti-peptide aptamers recognize amino acid sequence and bind a protein epitope. *Proc Natl Acad Sci U S A*, **93**, 7475-7480.
- Yamamoto, R., Baba, T. and Kumar, P.K. (2000) Molecular beacon aptamer fluoresces in the presence of Tat protein of HIV-1. *Genes Cells*, **5**, 389-396.
- Yan, X., Hollis, T., Svinth, M., Day, P., Monzingo, A.F., Milne, G.W. and Robertus, J.D. (1997) Structure-based identification of a ricin inhibitor. *J Mol Biol*, **266**, 1043-1049.
- Yoshinari, S., Koresawa, S., Yokota, S., Sawamoto, H., Tamura, M. and Endo, Y. (1997) Gypsophilin, a new type 1 ribosome-inactivating protein from *Gypsophila elegans*: purification, enzymatic characterization, and subcellular localization. *Biosci Biotechnol Biochem*, **61**, 324-331.
- Yoshinari, S., Yokota, S., Sawamoto, H., Koresawa, S., Tamura, M. and Endo, Y. (1996) Purification, characterization and subcellular localization of a type-1 ribosome-inactivating protein from the sarcocarp of *Cucurbita pepo*. *Eur J Biochem*, **242**, 585-591.

- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. and Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883-896.
- Zamore, P.D. (2002) Ancient pathways programmed by small RNAs. *Science*, **296**, 1265-1269.
- Zapp, M.L. and Green, M.R. (1989) Sequence-specific RNA binding by the HIV-1 Rev protein. *Nature*, **342**, 714-716.
- Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M. and Snyder, M. (2001) Global analysis of protein activities using proteome chips. *Science*, **293**, 2101-2105.
- Zilinskas, R.A. (1997) Iraq's biological weapons. The past as future? *Jama*, **278**, 418-424.
- Zimmermann, G.R., Wick, C.L., Shields, T.P., Jenison, R.D. and Pardi, A. (2000) Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *Rna*, **6**, 659-667.
- Zuker, M. (2000) Calculating nucleic acid secondary structure. *Curr Opin Struct Biol*, **10**, 303-310.
- Zwanzig, R., Szabo, A. and Bagchi, B. (1992) Levinthal's paradox. *Proc Natl Acad Sci U S A*, **89**, 20-22.

Vita

Jay R. Hesselberth was born on July 16, 1975 in Peoria, Illinois to Jerald and Judy Hesselberth. His sister Kelly teaches elementary school outside Chicago, Illinois. He attended The University of Iowa beginning in 1993 and received his Bachelor of Science in Biochemistry in 1997. From 1997 to 1998 he worked as a research assistant in the laboratory of Brian F. Tack at The University of Iowa. In August 1997 he entered graduate school at The University of Texas at Austin under the supervision of Andrew D. Ellington. His scientific publications include:

Hesselberth J.R., Knudsen S, Robertson M.P., Ellington A.D. Simultaneous detection of diverse analytes with an aptazyme ligase array. *Anal Biochem.* 2003 Jan 15; 312(2) 106-112.

Hesselberth J.R., Ellington A.D. A (ribo) switch in the paradigms of genetic regulation. *Nat Struct Biol.* 2002 Dec;9(12):891-893.

Hoffman D, Hesselberth J, Ellington A.D. Switching nucleic acids for antibodies. *Nat Biotechnol.* 2001 Apr;19(4):313-314.

Hesselberth, J.R., Robertson, M.P., Jhaveri, S., Ellington A.D. *In vitro* selection of nucleic acids for diagnostic applications. *Reviews in Molecular Biotechnology*, 2000, 74 (1) 15-25.

Ellington, A.D., Hesselberth, J.R., Jhaveri, S., Robertson, M.P. Combinatorial methods: aptamers and aptazymes. *Proc. SPIE-Int. Soc. Opt. Eng.*, 1999: v. 3858, 126-134

- Hesselberth J.R., Miller D, Robertus J, Ellington A.D. *In vitro* selection of RNA molecules that inhibit the activity of ricin A-chain. J Biol Chem 2000 Feb 18;275(7):4937-4942
- Singh P.K., Jia H.P., Wiles K, Hesselberth J.R., Liu L, Conway B.A., Greenberg E.P., Valore E.V., Welsh M.J., Ganz T, Tack B.F., McCray P.B. Jr. Production of β -defensins by human airway epithelia. Proc. Natl. Acad. Sci. 1998 95: 14961-14966

Permanent address: 601 E. Tyler, Morton, IL 61550

This dissertation was typed by the author.