In Vitro Selection of
RNA Binding Peptides

Thesis by

Terry Torao Takahashi

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2005

(Defended May 11, 2005)

Acknowledgements

It has been said that you learn more from your failures than your successes – if that is true, I must have learned a lot. Scientific research is a lot like gambling: intermittent reward is a powerful motivator. Although the seven years I've spent here have been filled with much frustration (and some successes too), I do feel it has been worth it for I have learned not only about science, but about who I am. There have been many people who have helped me be who I am today, and they deserve my thanks.

First and foremost, I have to thank my wife-to-be, Katie Percell, for her love, patience, and support over the last seven years. She is one of the kindest and optimistic people I know. I know too how much she wants to go back to Hawai'i, yet she has stayed up here for me, and I very much appreciate her sacrifice. I'm just sorry it will be a little longer… hopefully not too much longer, though. She has also read most of my thesis, even though much of it is foreign to her, and has tried to make my thesis writing as bearable as possible.

I have to thank my family – my mom and dad especially – for their love and support. It amazes me how much I can recognize parts of me that can be attributed to them. I appreciate their sacrifices in sending me to 'Iolani and to Claremont McKenna, both of which aren't exactly cheap. My mom used to drive me to 'Iolani, go to work, pick me up, then drive me back home, a >100 mile daily commute, which especially on a small island, is ridiculous. Without you guys, I would not be here today. Tammy and Mike Keller, my sister and brother-in-law, got me through the culture, weather, and all

other types of shock when I first came up to California. I'm sure it would have been more of an adjustment if they weren't here to help.

Thanks go out to my housemates Hugh Kimura, Jon Yee, and Shawna Ling, from the "Real World House," who made life outside of lab fun and always interesting. Hugh would be up for pretty much anything, including Donutman runs in the middle of the night for some strawberry donuts. Stacy Miyake, Michelle Sugihara, and Wendi Nishikawa were always around for laughs and some good Japanese or Hawaiian food. Hugh, Corey Chung and Mike Torres were my golf buddies and sources of endless laughter and entertainment (especially at Scholl Canyon). Alisa Shiraki listened to my grumbling about graduate school, and always had an optimistic outlook about things.

Special thanks goes to my previous teachers. Pamela Fujinaka, my high school chemistry teacher, convinced me to take her AP Chemistry class as well as encouraged me to take the Chemistry Olympiad Exam. Her class was harder and made you think more creatively than some of my college classes, and gave me a solid foundation in chemistry. The labs in her class were not follow-the-direction experiments and made me think outside of the box, a skill extremely useful in research. Todd Richmond was my undergraduate advisor and all-around cool guy. He taught me much of what I know about molecular biology and pushed me hard to get started in research. His biochemistry course got me excited about biological chemistry – otherwise I would have been an organic chemist.

My thanks go to my advisor at Caltech, Rich Roberts, whose research suggestions and insights have always been helpful. I joined his lab partly because I was impressed by his command of the literature and his broad scientific knowledge, partly because I would

have the opportunity to work on a variety of projects, and partly because he's a nice guy. Rather than telling me what experiment to do next, he let me figure things out for myself, with a nudge here or there.

My thesis committee, Peter Dervan, Steve Mayo, and Doug Rees have always been helpful and provided sound advice. I am grateful for their guidance and support. John Love, John Marino, and Scott Ross taught me many gory details about NMR spectroscopy. Scott has also been extremely helpful in helping me run my NMR samples and I wish him a speedy recovery.

I have had the privilege to work in a lab of very nice people. I thank Shuwei Li who helped me a lot when I first started and for his continuing advice and suggestions after I joined the lab. Bill Ja also helped me get up and running, and has always been a critically thinking scientist with good experimental suggestions, and is an expert in finding free stuff. Bill works hard and is always in lab, but plays hard too. Jie Xu and Joan He were not only fun to talk to but also had helpful experimental suggestions. I have to thank Jeff Barrick who was an undergrad in a graduate student's body and worked with me on the N project and always had great scientific ideas. Chris Balmaseda (Balms) helped me a lot with protein purification and was fun to hang out with. Jinsong Ren taught me a lot about biophysical chemistry and was always around to cheer people up. Tianbing Xia taught me about nucleic acids as well as NMR, has a great scientific mind, and has made me see problems from different perspectives. His wife, Cindy (Xin) Qi, has taught me much about organic chemistry and made me think more about life after Caltech. Shelley (Starck) Green gave me many good experimental suggestions, especially controls, but was also a fiercely loyal and good friend who has always been

supportive of me. Her husband Harry Green provided a biologists' perspective to things, often during rounds of golf. Over the past year, I've worked closely with Christine Ueda and Steve Millward on a variety of projects, some of which have finally come to fruition, and their help has been indispensable. Ryan Austin has not only helped me with science, but has made me think more about non-science things, and along with Adam Frankel, Anders Olson, and Premal Shah, have had many interesting discussions over an Ernie's lunch. Anders is also my baymate and always has insightful scientific suggestions.

My thanks go to Margot Hoyt, who has kept things running behind the scenes as well as Alison Ross who always helped me out when I had administrative questions and problems. Po-Ssu Huang deserves thanks for his help with science as well as his friendship. I still remember rooming with him during the Biology retreat my first year here and he helped me a lot in Bch170 and Bch176. Possu is also the original Mac evangelist and guru and helped me whenever I've had computer problems. I would imagine that he would like me to attribute the problems to the Microsoft programs, rather than OSX. Lastly, others that deserve thanks at Caltech are, Chris Otey, Corey Hu, Justin Bois, Greg Drummond, Robert Dirks, Shantanu Sharma, and Geoff Hom.

Abstract

RNA is recognized to play an increasing number of roles in the cell: transcription regulation, translation, and catalysis. Peptides that bind RNA would therefore be useful as biochemical tools and lead compounds for therapeutics. Existing genetic methods of isolating RNA binding peptides are prone to biases and can only search millions of sequences. *In vitro* selections using mRNA display provide an avenue to discover specific, high affinity peptides that bind to any RNA target from libraries composed of trillions of molecules.

Here, we describe initial experiments to optimize the mRNA display selection cycle for the isolation of RNA binding peptides. We use this optimized cycle to show that enrichment of specific sequences is possible using mRNA display, and select mutants of the λ N peptide which bind in a different conformation than wild-type. Characterization of these peptides demonstrates that affinity is not enough for *in vivo* activity; binding in a correct conformation is also important.

Based on these experiments, we designed a strategy to isolate RNA binding peptides to targets for which no natural ligand is known. We test this strategy and isolate peptides that bind to functionally important domains of telomerase RNA with nanomolar affinity and high specificity. Using mutagenic PCR and additional rounds of selection, we increase the specificity of several peptides for telomerase RNA and also isolate other peptides which bind an important domain of the Hepatitis C Virus internal ribosome entry site.

Table of Contents

Chapters

List of Figures

Chapter 5:

Chapter 6:

Chapter 7:

List of Tables

List of Abbreviations

DNA =   Deoxyribonucleic acid
RNA =   Ribonucleic acid
mRNA =   Messenger ribonucleic acid
cDNA =   Complementary deoxyribonucleic acid
RT =   Reverse transcription/reverse transcriptase
Tris =   Tris-(hydroxymethyl)-aminomethane
pmol =   Picomole
DTT =   Dithiothreitol
EDTA =   Ethylenediaminetetraacetate
ddH$_2$O =   Double distilled water
tRNA =   Transfer ribonucleic acid
Ni-NTA =   Nickel-nitriloacetic acid
RNase =   Ribonuclease
TMV =   Tobacco mosaic virus
Bsa =   Bovine serum albumin
HSQC =   Heteronuclear single quantum coherence
NMR   =   Nuclear magnetic resonance
HCV =   Hepatitis C virus
IRES =   Internal ribosome entry site
hTR =   Human telomerase RNA
hTERT =   Human telomerase reverse transcriptase

Chapter 1

Background and Perspective

Two decades ago RNA's role in biology was simple. The central dogma (Figure 1.1) dictated that RNA's major role was an information carrier between DNA and protein. How things have changed. RNA is increasingly recognized to play a role in many cellular processes (1) including catalysis (2), chromosome maintenance (3), transcriptional control (4), and the synthesis of proteins (5).
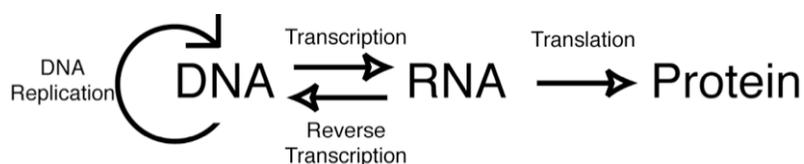
Figure 1.1. The central dogma of biology. DNA is transcribed into RNA, which is then translated into protein. RNA can be reverse transcribed into DNA and DNA can be replicated.

As RNA's role in biology is better understood, it becomes a more attractive target for drug design (6, 7). However, rational drug design is difficult since RNA often changes structure upon ligand binding (8), electrostatics and ion interactions must be accurately modeled (9), and obtaining structures by crystallography is difficult (10).

Genetic approaches (11, 12) and selection methods (13, 14) are powerful methods enabling the isolation of RNA binding proteins and peptides. However, these systems are often subject to biases and libraries sizes of 10^6-10^9 molecules since an *in vivo* step is required. Thus, these methods often target known binding sites on RNA, or limit the amino acid alphabet used in the selection.

The work described in this thesis represents initial efforts to develop methodologies that enable the isolation of RNA binding peptides and proteins to any RNA target. These proteins and peptides would be useful both as tools in biochemistry

and molecular biology and as potential therapeutics. mRNA display is the key experimental technique that enables the isolation of RNA binding peptides. mRNA display is performed entirely *in vitro*, allowing libraries of up to 100 trillion independent sequences to be synthesized (15). Thus, mRNA display will reduce the biases inherent in technologies requiring an *in vivo* step, as well as enable the selection from totally random libraries composed of all twenty amino acids.

The thesis is divided into chapters describing different aspects of the methodology development. Each chapter is described below:

Chapter 2

Chapter 2 is a published review of the mRNA display field, and describes successful selections resulting from mRNA display. Several conclusions can be drawn from mRNA display experiments including the fact that many sequences of nanomolar affinity can be isolated form mRNA display selections, in contrast with the results from phage display selections where micromolar affinities are more typical. Applications unique to mRNA display, including unnatural amino acid incorporation and protein chips are also discussed.

Chapter 3

This published chapter describes efforts to optimize the mRNA display selection cycle. Although we optimized the cycle for the selection of RNA binding peptides, the protocols described in this chapter are generally applicable to any mRNA display selection.

Chapter 4

Part of this chapter has been previously published and describes the initial experiments to use mRNA display for the selection of RNA binding peptides. We began with a control selection based on the λ N peptide system to show that the wild-type sequence could be enriched from a randomized pool, then extended the methodology to select for novel mutants of the λ N peptide. Characterization of one of these mutants implies the mutant binds in a different conformation from the wild-type peptide.

Chapter 5

This published chapter describes the further characterization of the λ N peptide mutants isolated in Chapter 4. The mutants all possess nanomolar affinities and different *in vivo* activites, however, affinity does not correlate with activity. Instead, activity is dependent upon binding in the correct structural conformation.

Chapter 6

Chapter 6 describes the selection of RNA binding peptides from totally random libraries to RNA targets for which no natural ligand is known. A strategy for the isolation of RNA binding peptides is described and used to isolate peptides with nanomolar affinity and high specificity that bind to human telomerase RNA (hTR). Selections targeted two important regions of telomerase RNA. Modification of the selection protocols enabled the selection of D-peptides.

Chapter 7

Chapter 7 discusses the attempts to improve the affinity of some of the peptides isolated in Chapter 6 through the introduction of mutagenic PCR into the selection cycle. Additionally, attempts to improve the specificity of the selected peptides as well as change the specificity toward a similar RNA target from the Hepatitis C Virus internal ribosome entry site are also described.

References

1. Caprara, M.G. and Nilsen, T.W. RNA: versatility in form and function. (2000) *Nat Struct Biol* **7**, 831-833.

2. Derose, V.J. Two decades of RNA catalysis. (2002) *Chem Biol* **9**, 961-969.

3. Blackburn, E.H. The end of the (DNA) line. (2000) *Nat Struct Biol* **7**, 847-850.

4. Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L. and Breaker, R.R. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. (2004) *Science* **306**, 275-279.

5. Nissen, P., Hansen, J., Ban, N., Moore, P.B. and Steitz, T.A. The structural basis of ribosome activity in peptide bond synthesis. (2000) *Science* **289**, 920-930.

6. Dejong, E.S., Luy, B. and Marino, J.P. RNA and RNA-protein complexes as targets for therapeutic intervention. (2002) *Curr Top Med Chem* **2**, 289-302.

7. White, L.K., Wright, W.E. and Shay, J.W. Telomerase inhibitors. (2001) *Trends Biotechnol* **19**, 114-120.

8. Williamson, J.R. Induced fit in RNA-protein recognition. (2000) *Nat Struct Biol* **7**, 834-837.

9. Draper, D.E. A guide to ions and RNA structure. (2004) *Rna* **10**, 335-343.

10. Ferre-D'amare, A.R., Zhou, K. and Doudna, J.A. A general module for RNA crystallization. (1998) *J Mol Biol* **279**, 621-631.

11. Sengupta, D.J., Zhang, B., Kraemer, B., Pochart, P., Fields, S. and Wickens, M. A three-hybrid system to detect RNA-protein interactions *in vivo*. (1996) *Proc Natl Acad Sci U S A* **93**, 8496-8501.

12. Harada, K., Martin, S.S. and Frankel, A.D. Selection of RNA-binding peptides *in vivo*. (1996) *Nature* **380**, 175-179.

13. Danner, S. and Belasco, J.G. T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. (2001) *Proc Natl Acad Sci U S A* **98**, 12954-12959.

14. Friesen, W.J. and Darby, M.K. Phage display of RNA binding zinc fingers from transcription factor IIIA. (1997) *J Biol Chem* **272**, 10994-10997.

15. Takahashi, T.T., Austin, R.J. and Roberts, R.W. mRNA display: ligand discovery, interaction analysis and beyond. (2003) *Trends Biochem Sci* **28**, 159-165.

Chapter 2

mRNA Display:
Ligand Discovery, Interaction Analysis
and Beyond

Abstract

In vitro peptide and protein selection using mRNA display* enables the discovery and directed evolution of new molecules from combinatorial libraries. These selected molecules can serve as tools to control and understand biological processes, enhance our understanding of molecular interactions, and potentially treat disease in therapeutic applications. In mRNA display, mRNA molecules are covalently attached to the peptide or protein they encode. These mRNA-protein fusions enable *in vitro* selection of peptide and protein libraries of more than 10^{13} different sequences. mRNA display has been used to discover novel peptide and protein ligands for RNA, small molecules, and proteins, as well as to define cellular interaction partners of proteins and drugs. In addition, several unique applications are possible with mRNA display, including self-assembling protein chips and library construction with unnatural amino acids, and chemically modified peptides.

*mRNA display has been referred to as mRNA-protein fusions (1), *in vitro* virus and *in vitro* virus virion (2), and PROfusion™ technology (3).

Introduction

Functional approaches, such as *in vitro* selection, currently provide the best means available for isolating peptides and proteins with desired chemical or biochemical properties. Over the last decade, display technologies have been essential tools in the discovery of peptide and protein ligands and in delineating *in vivo* interaction partners. The phage (4) and ribosome display systems (5) have been principally used for discovery, while the yeast two-hybrid method (6) has been used for *in vivo* interaction analysis.

Despite their power, technologies that require *in vivo* step, such as phage display and the yeast two-hybrid system face certain limitations. In phage display, libraries must be transformed into bacteria, limiting the number of possible independent sequences to 10^9–10^{10}. The total number of sequences represented can be further decreased by other issues including: degradation of unfolded molecules, poor expression in the bacterial host, failure in processing to the phage surface, failure to fold in the oxidizing periplasmic space of *Escherichia Coli*, or toxicity of the gene product. Similarly, the two-hybrid system requires that the interaction partners be cloned into yeast, limiting the number of constructs examined to 10^6-10^7. Additionally, in the two-hybrid approach, interactions must occur in the nucleus, limiting control over the binding stringency, appropriate binding partners, and biochemical conditions.

Totally *in vitro* techniques, such as ribosome and mRNA display, overcome many limitations of phage display and the two-hybrid system. These approaches reduce biases due to expression and routinely generate libraries of more than 10^{12} independent

molecules since no transformation step is required. In addition, more control can be exercised over the binding conditions as well as the stringency of selection.

mRNA Display

The mRNA display peptide and protein selection system provides an alternative method that can be applied to both ligand discovery and interaction analysis problems (2, 7). In this approach, encoded peptide and protein libraries are covalently fused to their own mRNA (Figure 2.1). Fusion synthesis is possible because the message can act as
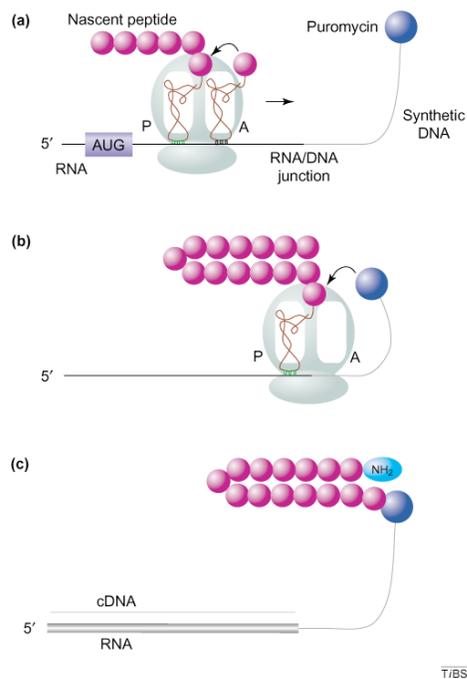
Figure 2.1. Formation of an mRNA-protein fusion. **(a)** mRNA (black) is ligated photochemically (11) or enzymatically (12) to a synthetic oligonucleotide (grey) containing puromycin (blue) at its 3' end. The ribosome (pale green) initiates synthesis of the template and reads in a 5'→3' direction. tRNAs (brown) and amino acids (pink) are shown in the P- and A-site of the ribosome. **(b)** Puromycin enters the ribosome attaching the template to the C-terminus of the nascent peptide. This entry occurs almost exclusively at the last or next to last codon (T. Snyder, A. Balakin, and R. W. Roberts, unpublished observation). **(c)** Reverse transcription generates cDNA (grey) that can be amplified by PCR.
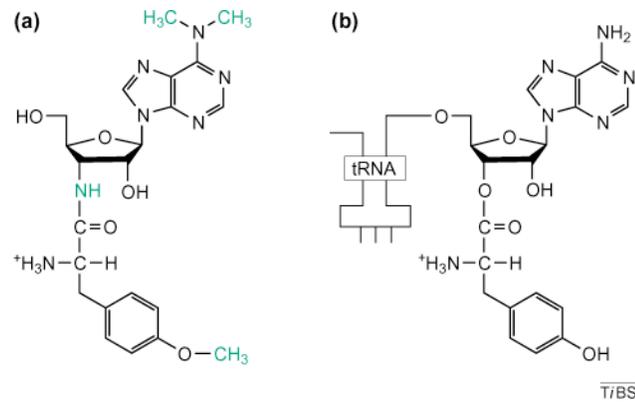
Figure 2.2. Puromycin **(a)** is a small molecule analog of Tyrosyl tRNA **(b)**. Differences between the two molecules are highlighted in green text.
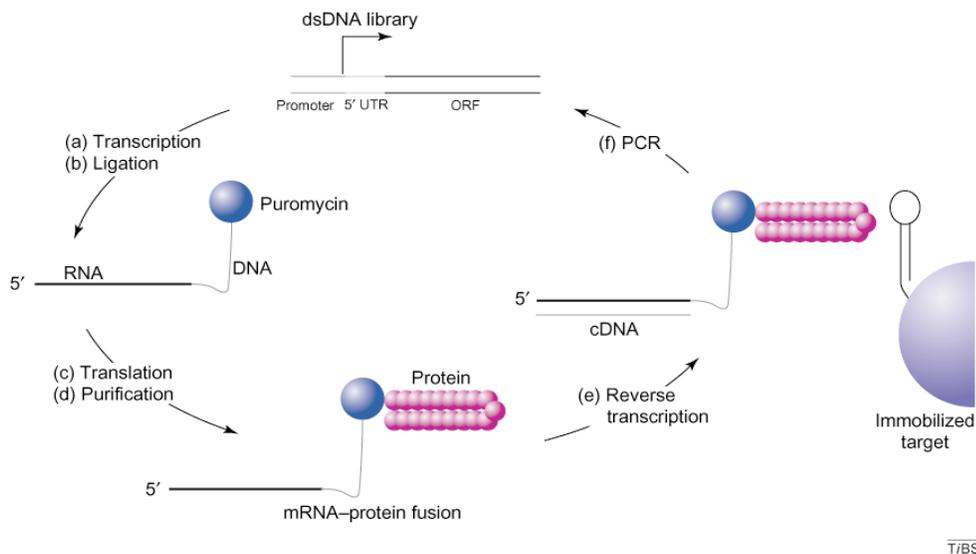
Figure 2.3. A typical mRNA Display selection cycle. **(a)** A library of double-stranded DNA sequences is transcribed to generate mRNA. **(b)** The mRNA is ligated to a puromycin oligonucleotide (blue) and used to program an *in vitro* translation reaction **(c)**. cDNA synthesis is performed **(e)** and the cDNA/mRNA-protein fusion is sieved using the target of interest. PCR is used to regenerate the full-length DNA construct **(f)**. For targets containing RNase or RNase H activity the cDNA can be crosslinked to the puromycin oligonucleotide to generate a cDNA-protein fusion (10).

both template and peptide acceptor if it contains a 3'-puromycin molecule. Puromycin serves as a chemically stable, small molecule mimic of aminoacyl tRNA (Figure 2.2). The selection cycle for a typical mRNA display experiment is shown in (Figure 2.3).

Detailed descriptions of experiments and protocols have been published elsewhere (8-12). Briefly, a synthetic oligonucleotide containing a 3' puromycin is ligated to the 3' end of an mRNA and the product is translated in rabbit reticulocyte lysate. The sequence present in the peptide is therefore encoded in the covalently attached mRNA, allowing the sequence information in the protein to be read and recovered after selection via reverse transcriptase (RT)-PCR. Thus, exceedingly small amounts of material can thus be amplified.

Since the original description of the mRNA display system in 1997 (1, 2), optimization has resulted in the ability to perform selection experiments on libraries containing more than 10^{13} molecules (8, 12, 13). Routinely 10–40% of the input mRNA template can be converted to fusion product, resulting in more than 5 x 10^{13} mRNA-protein fusions per milliliter of translation reaction. Overall, the mRNA display system allows libraries with sequence complexity approximately 10,000-fold that of phage display (4), 10^6-fold over yeast display or yeast two- and three- hybrid systems (6, 14-16), and approximately 10^9-fold over colony screening approaches (17).

In the majority of mRNA display experiments, polypeptides with relatively short chain lengths (10–110 amino acids) have been used. Larger proteins have been studied as well (e.g. λ protein phosphatase a 24 kDa enzyme (12) and β-lactamase, a 31 kDa enzyme; S. Li & R. W. Roberts unpublished observation), but these typically form fusion products with somewhat reduced efficiency. Even for such proteins, libraries can still be

readily constructed that are orders of magnitude larger than a typical phage display library. Another feature of the mRNA display constructs is that the mRNA appears to improve the solubility of the attached protein, enabling functional selection of sequences that can aggregate or are only partially soluble when expressed by themselves (see below).

Ligand Discovery with mRNA Display

Selections to discover new peptides and proteins with desired features have now been completed. Sequences have been isolated that bind RNA, small molecules, and proteins. These results illustrate three important principles: (i) larger library size does, in fact, result in higher affinity molecules; (ii) larger libraries result in a greater diversity of sequences with similar function; and (iii) the vast number of sequences recovered after selection can be analyzed using informational techniques, such as sequence covariation analysis (see below).

RNA-Binding Peptides

RNA binding proteins participate in regulation of transcription, splicing, and translation, and have been implicated in several diseases (18, 19). Selections for RNA-binding peptides also present a stringent functional test of mRNA display. Numerous mRNA display selections have isolated more than 100 chemically distinct RNA binding peptides (20-22). These selections demonstrate that even highly basic and unstructured molecules retain function and do not interact with the attached mRNA/cDNA hybrid. The majority of experiments have been conducted using the RNA binding domain from phage λ N protein as a model system due to its small size (22 amino acids), high affinity

(low nanomolar), and thorough characterization. Selections have resulted in numerous

peptides with nanomolar affinity for their cognate target (21). The highest complexity

RNA selection performed to date contained 10 random residues (X_{10}, where X is any of

the twenty amino acids) and more than 9 x 10^{12} different sequences in the initial library

(Figure 2.4A) (21). The selected peptides all bound the *boxB* RNA hairpin with very

high affinity (K_d = 0.5 to 5 nM) and most demonstrated equal or better specificity than the

wild-type sequence. However, the selected peptides showed striking chemical diversity
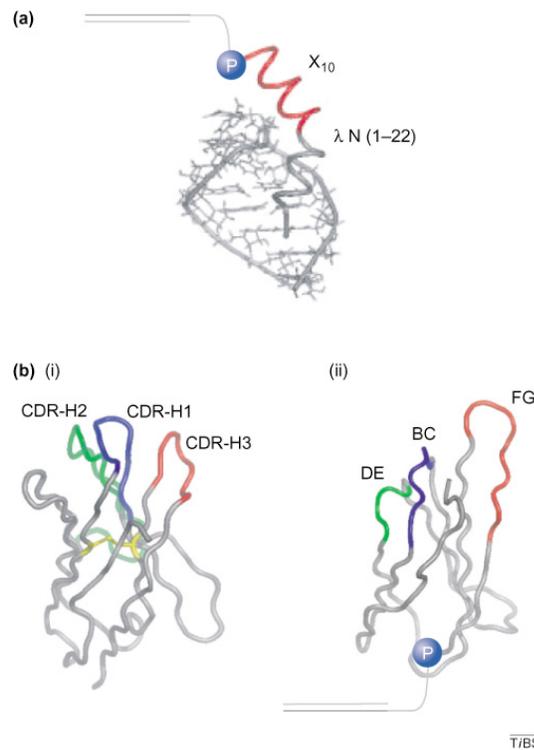
Figure 2.4. Examples of mRNA display libraries. **(a)** A library containing ten consecutive random residues (red) was constructed from the λ N RNA-binding protein and selected for binding RNA hairpins (21). **(b)** Structural comparison between V_{HH} (i) and the 10th fibronectin type III domain (^{10}Fn3) (ii). Three libraries derived from ^{10}Fn3 were constructed as antibody mimics of V_{HH} and selected for binding to TNF-α and leptin (35). Residues of ^{10}Fn3 that were randomized are shown in color. Abbreviations: CDR, complementarity-determining region; P, puromycin.

and bore little resemblance to wild-type. Only a single Arg at position 15 (glutamine in the wild-type) showed any significant conservation. Despite the lack of homology, sequence covariation analysis indicated that the molecules fold into helices, showing correlations between adjacent residues (i to i + 1) and residues located one turn away (i to i + 3 and i to i + 4) (20).

ATP Aptamers

Primordial proteins presumably evolved from random sequences and it is probable that one of these first proteins bound ATP. Keefe and Szostak (23) attempted to isolate a modern relative of these prebiotic proteins by selecting for ATP binding using a 109-mer protein containing an 80-amino acid random region. Libraries containing such large numbers of random positions present special problems since the probability of encountering a frame shift or stop codon can become substantial (12, 24). To solve these problems, the Szostak group utilized mRNA display to preselect library fragments for readability, selecting for the presence of N- and C-terminal epitope tags (13). The readable fragments were then digested and assembled into full-length libraries that contained greatly improved open reading frames (ORFs). After 8 rounds of selection for ATP binding, 4 distinct protein sequence families could be discerned (23). Further rounds of selection, combined with mutagenesis, resulted in a clone (18-19) that bound ATP with high affinity (K_d = 100 nM) and could discriminate ATP from other nucleotide triphosphates with up to 2000-fold specificity. These protein aptamers contain a conserved Cys-Xaa-Xaa-Cys (CXXC) motif and function in a metal-dependent manner. The fact that the aptamer uses metals might indicate that chelation provides a simple way to create stable, functional protein structures, which is consistent with the large energy

seen for protein-metal interactions (25). One feature of these aptamers is that only a fraction of each clone appears folded and functional; the proteins themselves tend to aggregate when expressed as free proteins. Thus, selection of these proteins was probably facilitated by the improved solubility imparted by the mRNA/cDNA tail, and argues that such sequences would not be found in a typical phage display selection. The fact that the functional clones are not well behaved likely reflects the relative paucity of proteins that are both folded *and* functional in the vastness of sequence space.

The structure of one of these aptamers has been recently determined by X-ray crystallography. By screening a few different constructs, Sollazzo and coworkers were able to obtain single crystals of one aptamer they termed artificial nucleotide-binding protein (ANBP) (26). The protein folds into a novel three-stranded β-sheet flanked by two α-helices and binds ATP in a manner similar to natural ATP binding proteins. It also exhibits biophysical characteristics reminiscent of natural proteins such as good NMR chemical shift dispersion and thermal unfolding.

Keefe and Szostak estimate that one in 10^{12} molecules in their initial library have the ability to bind ATP – approximately the same fraction seen for ATP-binding RNA aptamers (27). This result is somewhat surprising given the greater chemical diversity of proteins (twenty sidechains) relative to nucleic acids (four sidechains). While functionally impoverished, nucleic acid aptamers may benefit from the ease of forming higher-order structures through simple base-pairing interactions, in contrast to proteins, which require a hydrophobic core for folding. It remains an open question if catalytic proteins can also be found with similar frequencies to their nucleic acid counterparts.

Streptavidin Aptamers

Szostak and coworkers also created long open reading frames for a binary patterned library (13). This library contained a random region of 87-88 amino acids with an initial complexity of $\sim10^{13}$ sequences, and was assembled from two distinct 11 amino acid segments containing hydrophobic and polar amino acid patterning that result in either amphipathic α-helices or β-strands (28). mRNA display selections against streptavidin resulted in a number of sequences that bound streptavidin with nanomolar affinity ($K_d \sim5$ nM) (29), and bind 200- to 2,200-fold better than the *Strep*-tag II peptide obtained previously by phage display (30, 31).

Although the library had been patterned to form helices and sheets in reading frame one, all of the selected molecules were shifted into reading frame three, effectively eliminating the patterning. The shifted frame seems to have been greatly preferred due to the presence of His-Pro-Gln (HPQ) tripeptide sequences. The HPQ peptide represents the minimal core of the *Strep*-tag II peptide and has been shown to bind streptavidin (30, 31). Frame one of the patterned library contained very few HPQ sequences (1/45,000 clones), owing to the library design, whereas in the third frame, HPQ peptides were present in 1/64 sequences.

The majority of the sequences contained at least one HPQ motif, one similar tripeptide motif (e.g., HPQ, His-Pro-Ala (HPA), and Leu-Pro-Gln (LPQ)) and do not appear to contain any disulfide bonds. A 38 amino acid sequence, termed the "SBP-tag," has been used for one-step affinity purification on streptavidin agarose and western blot detection using streptavidin-horseradish peroxidase for visualization (32). Despite frameshifting, the patterned library still contained ~10,000-fold greater sequence

complexity than a standard phage display selection, likely leading to the high affinity of the resulting aptamers. Finally, this experiment demonstrates the difficulty in designing random libraries with imposed structural features *a priori*.

TNF-α Aptamers using the ^{10}Fn3 Domain

Monoclonal antibodies are useful both as a biochemical tool and as potential therapeutics. mRNA display has been used to isolate novel antibody mimetics based on a fibronectin domain. The tenth type III domain of human fibronectin (^{10}Fn3) displays an Arg-Gly-Asp (RGD) sequence involved in cell-surface recognition by integrins (33). The ^{10}Fn3 domain has a similar β-sheet architecture to antibody V_H domains, with three structurally analogous loops (Figure 2.4B). The antibody-like structure, exposure to the immune system, small size (94 amino acids), lack of disulfide bonds, high bacterial expression levels, and high stability (T_m = 90°C) all make the make the ^{10}Fn3 domain an excellent potential scaffold. However, previous attempts to isolate ^{10}Fn3 derivatives using phage display resulted in molecules with only modest affinity ($IC_{50, \text{ubiquitin}}$ = 5 μM) and relatively non-specific binding (34).

Xu et al. constructed three libraries based on ^{10}Fn3, randomizing either one loop (libraries 1 and 2) or all three loops simultaneously (library 3) (35). An mRNA display selection was then performed against tumor necrosis factor-α (TNF-α). After 9-10 rounds of selection, diverse, high affinity (K_d = 1 – 24 nM), and high specificity ligands were isolated, primarily originating from library 3. Further selection for a total of 14 rounds resulted in clones with sub-nanomolar affinity (K_d = 90-110 pM). Returning to round 8, mutagenic PCR was added to the selection cycle, duplicating the affinity

maturation process of antibodies. Further rounds of selection resulted in a clone with very high affinity (K_d = 20 pM). While less stable than wild-type [10]Fn3, the best clone (12.21) nonetheless was monomeric and showed good expression and protease resistance at 30°C. Immobilized versions of a round 9 clone (9.12) could be used to capture TNF-α from a solution of 10% fetal bovine serum, demonstrating the high specificity of these reagents, even when immobilized on a solid support.

Specificity and Interaction Analysis

Epitope Recovery

In addition to enriching sequences containing a known epitope (7), mRNA display can also be used to determine which sequences are critical for recognition. Baggio et al. used two random libraries to investigate the specificity of peptides binding the anti-*c-myc* antibody 9E10 (36) and bovine trypsin (37). Selection against the 9E10 antibody with a 27 random residue (X_{27}) library revealed a consensus sequence x[Q/E]xLISExx[L/M] (the *c-myc* tag is EQKLISEEDLN), demonstrating that the Leu-Ile-Ser-Glu (LISE) sequence was the core element recognized by the antibody. In the same work, a six random residue library (positions 3-8) was created using the *Ecballium elaterium* trypsin inhibitor two protein (EETI-II) as a scaffold (37). EETI-II, a 28-residue protein with three disulfide bonds, is a member of the knottin family and inhibits bovine trypsin via interaction at positions 3-8 (38). Selection against trypsin yielded a sequence consensus of Pro-Arg-Xaa-Leu-Xaa-Xaa (PRxLxx), with 20% of the selected clones matching the wild-type sequence of Pro-Arg-Ile-Leu-Met-Arg (PRILMR).

mRNA display has also been applied to define a recognition epitope for the oncogenic v-abl tyrosine kinase, which is a target of great biological and therapeutic interest (39). Initial experiments demonstrated that mRNA-peptide fusions containing a v-abl phosphorylation site (the tyrosine residue in EAI**Y**AAPFAKKK) could be phosphorylated by the v-abl kinase and immunoprecipitated with α4G10, an anti-phosphotyrosine monoclonal antibody. Libraries of the form GCGGX_5**Y**X_5GCG were subjected to phosphorylation with v-abl and precipitation with α4G10. The majority of clones contained an [Ile/Leu/Val]-Tyr-Xaa$_{1-5}$-[Pro/Phe] ([I/L/V]Yx_{1-5}[P/F]) consensus. Interestingly, despite the sequence variations, the kinase effectively phosphorylated all 12 of the consensus clones, indicating a broader specificity than previously thought.

Cellular Interaction Partners

mRNA display libraries constructed from cDNA offer the potential of isolating biologically relevant interaction partners. Hammond and coworkers used a random priming approach to create mRNA display libraries from several different human tissues (40). This approach yields libraries of various lengths and in three reading frames, and it also allows the experimenter construct libraries with tissue-specific primer tags. After selection, PCR using these primers can be used to deconvolute the library and obtain binders from specific tissues. Sieving cellular libraries against the anti-apoptotic protein Bcl-X$_L$ resulted in isolation of over 20 different proteins including the known interaction partners Bim, Bax and BCL2L12. The diversity in the cellular mRNA display libraries means that hundreds to millions of fragments of various lengths will be present from each gene. In that vein, the Bcl-X$_L$ selection demonstrated that alignment of multiple positive

clones is equivalent to typical deletion analysis, providing a clear indication of the sequence boundaries necessary for recognition.

Cellular libraries may also be used to characterize and discover cellular proteins or receptors that interact with a drug of interest. McPherson et al. used the immunosuppressive drug FK506 as a target for cellular libraries (41). This work resulted in isolation of the known target, FK506 binding protein (FKBP) and defined a region within FKBP that was necessary and sufficient for interaction with the drug (41).

Unique Applications of mRNA Display

Self-Assembling Protein Microarrays

The mRNA-protein fusions used in mRNA display can also be used for high-throughput screening applications. Protein chips offer the promise of quick analysis of the expressed protein content in a sample and performing *in vitro* interaction analysis. Weng et al. demonstrated that a standard DNA chip could be converted to a protein chip by hybridization of mRNA-protein fusions (Figure 2.5) (3). mRNA-protein fusions coding for the MYC, FLAG, or HA11 epitopes were synthesized and incubated with a DNA chip. The chip was imprinted with DNA complementary to a unique (MYC or FLAG or HA11) or common (5' or 3') nucleic acid portion of the fusions. Hybridization of the fusions to complementary DNA directs the self-directed assembly of the protein chip. The experiments demonstrate that at least for antibody-epitope interactions, these protein arrays preserve the functionality of the displayed proteins, present them in a uniform orientation, and have sub-attomole detection limits.
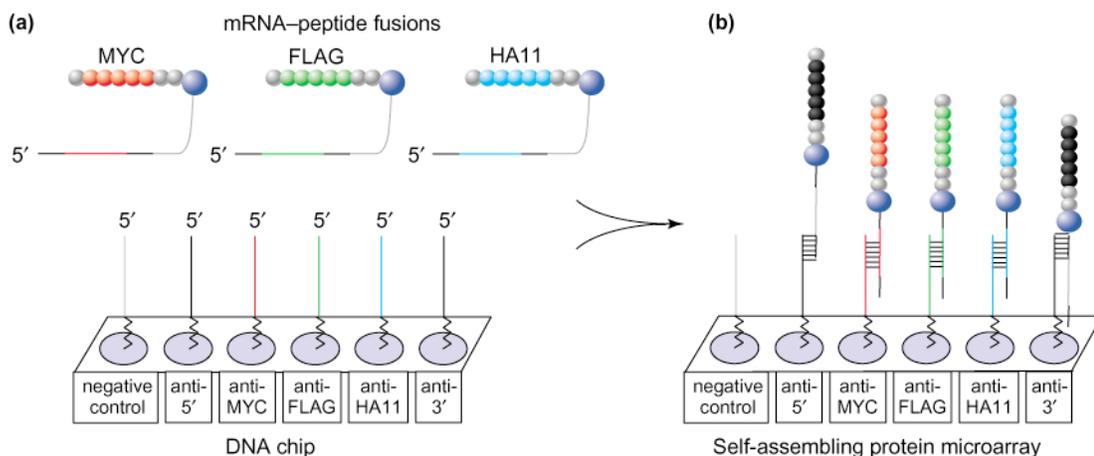
Figure 2.5. A self-assembling protein chip. **(a)** A mixture of mRNA-protein fusions containing either the MYC (red), FLAG (green), or HA11 epitope (light blue) (all three, black) epitope was incubated with a standard DNA chip. The nucleic acid component directs the fusions to regions on the chip containing complementary DNA. **(b)** A DNA complementary to the 5' or 3' sequence hybridizes to all three fusions, whereas the anti-MYC DNA will only isolate fusions containing the MYC epitope. This results in spatially addressable peptide micro arrays that can be recognized by monoclonal antibodies (35).

Non-natural Libraries

Phage display and the yeast two-hybrid system contain an obligate *in vivo* step and thus are generally limited to display only the 20 natural amino acids. Expansion of the amino acid alphabet would increase the chemical diversity that can be displayed and facilitate the discovery of greatly improved ligands. Unnatural amino acids can be introduced either during translation or post-translationally.

Recently, Li et al. demonstrated that the suppressor tRNA strategy for incorporating unnatural amino acids (42, 43) could be used to create mRNA display libraries bearing biocytin, a biotinylated lysine derivative (44) (Figure 2.6). After selection, the library was enriched in sequences containing the amber stop codon (TAG), which was suppressed by biocytin. The combination of these two powerful technologies

increases the chemical diversity that can be displayed and should facilitate discovery of ligands with improved affinity, specificity, stability or reactivity.

Using the nonsense suppression strategy, a maximum of three unnatural amino acids could be incorporated – one for each stop codon. However, recent work has shown that sense suppression is also possible, allowing the incorporation of up to 64 unnatural amino acids and essentially rewriting the genetic code. Frankel and Roberts used a biocytin-charged tRNA to select for a GUA codon that would allow unnatural amino acid incorporation with efficiencies comparable to that of nonsense suppression (45). Using the GUA codon, unnatural polymers (termed "encodamers") of N-methyl phenylalanine (N-MePhe) were synthesized and exhibited marked protease resistance (46).

Other work to incorporate unnatural amino acids has focused on post-translationally derivatizing libraries. A library bearing a pendant penicillin sidechain was used to select for peptides that increased the inhibitory activity of the attached penicillin to penicillin binding protein 2a (PBP2a) by more than 100-fold (47). This strategy should be applicable to improve a variety of small molecule compounds, and also increase the chemical diversity of mRNA display libraries.
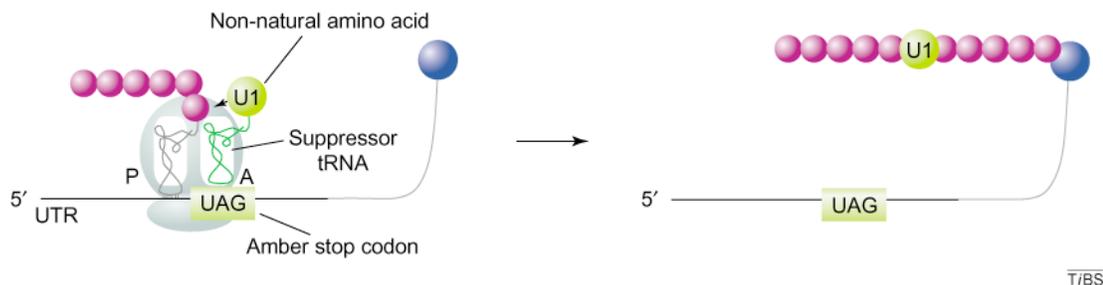
Figure 2.6. Inserting non-natural residues into mRNA display libraries. *In vitro* nonsense suppression using a chemically aminoacylated suppressor tRNA was used to insert biocytin into mRNA display libraries and select for the presence of the unnatural residue (44).

Conclusions

Techniques for performing mRNA display are now well established and allow facile synthesis and selection of mRNA-protein fusion libraries (8, 12, 13). Completed selections demonstrate that mRNA display is a powerful tool for both ligand discovery and interaction analysis. Notable features of the resulting ligands are high affinity (nanomolar to picomolar) and sriking sequence diversity present (21, 29, 35). The *in vitro* nature of the system provides a unique opportunity for *in vitro* affinity maturation and evolution (23, 35), inclusion of non-natural residues (44), chemical derivatization of libraries, and the opportunity for *in vitro* recombination experiments (48). Future applications point toward the isolation of new catalysts and the creation of libraries composed entirely of unnatural sidechains or non-peptidic backbones.

References

1.   Roberts, R.W. and Ja, W.W. *In vitro* selection of nucleic acids and proteins: What are we learning? (1999) *Curr Opin Struct Biol* **9**, 521-529.

2.   Nemoto, N., Miyamoto-Sato, E., Husimi, Y. and Yanagawa, H. *In vitro* virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. (1997) *FEBS Lett*. **414**, 405-408.

3.   Weng, S., Gu, K., Hammond, P.W., Lohse, P., Rise, C., Wagner, R.W., Wright, M.C. and Kuimelis, R.G. Generating addressable protein microarrays with PROfusion™ covalent mRNA-protein fusion technology. (2002) *Proteomics* **2**, 48-57.

4.   Smith, G. and Petrenko, V. Phage display. (1997) *Chem. Rev*. **97**, 391-410.

5.   Hanes, J. and Pluckthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. (1997) *Proc Natl Acad Sci U S A* **94**, 4937-4942.

6.   Fields, S. and Song, O.K. A novel genetic system to detect protein-protein interactions. (1989) *Nature* **340**, 245-246.

7.   Roberts, R.W. and Szostak, J.W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12297-12302.

8.   Barrick, J.E., Takahashi, T.T., Balakin, A. and Roberts, R.W. Selection of RNA-binding peptides using mRNA-peptide fusions. (2001) *Methods* **23**, 287-293.

9.   Kurz, M., G., K. and Lohse, P.A. An efficient synthetic strategy for the preparation of nucleic acid-encoded peptide and protein libraries for *in vitro* evolution protocols. (2000) *Molecules* **5**, 1259-1264.

10. Kurz, M., Gu, K., Al-Gawari, A. and Lohse, P.A. cDNA-protein fusions: covalent protein-gene conjugates for the *in vitro* selection of peptides and proteins. (2001) *Chembiochem* **2**, 666-672.

11. Kurz, M., Gu, K. and Lohse, P.A. Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA-protein fusions. (2000) *Nucleic Acids Res*. **28**, E83.

12. Liu, R., Barrick, J.E., Szostak, J.W. and Roberts, R.W. Optimized synthesis of RNA-protein fusions for *in vitro* protein selection. (2000) *Methods Enzymol*. **318**, 268-293.

13. Cho, G., Keefe, A.D., Liu, R., Wilson, D.S. and Szostak, J.W. Constructing high complexity synthetic libraries of long ORFs using *in vitro* selection. (2000) *J. Mol. Biol.* **297**, 309-319.

14. Boder, E.T. and Wittrup, K.D. Yeast surface display for screening combinatorial polypeptide libraries. (1997) *Nat. Biotechnol.* **15**, 553-557.

15. Gyuris, J., Golemis, E., Chertkov, H. and Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. (1993) *Cell* **75**, 791-803.

16. Sengupta, D.J., Zhang, B.L., Kraemer, B., Pochart, P., Fields, S. and Wickens, M. A 3-hybrid system to detect RNA-protein interactions *in vivo*. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8496-8501.

17. Arnold, F.H. and Volkov, A.A. Directed evolution of biocatalysts. (1999) *Curr. Opin. Chem. Biol.* **3**, 54-59.

18. Burd, C.G. and Dreyfuss, G. Conserved structures and diversity of functions of RNA-binding proteins. (1994) *Science* **265**, 615-621.

19. Draper, D.E. Themes in RNA-protein recognition. (1999) *J Mol Biol* **293**, 255-270.

20. Barrick, J.E. and Roberts, R.W. Sequence analysis of an artificial family of RNA-binding peptides. (2002) *Protein Science* **11**, 2688-2696.

21. Barrick, J.E., Takahashi, T.T., Ren, J., Xia, T. and Roberts, R.W. Large libraries reveal diverse solutions to an RNA recognition problem. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12374-12378.

22. Xia, T., Frankel, A., Takahashi, T.T., Ren, J. and Roberts, R.W. Context and conformation dictate function of a transcription antitermination switch. (2003) *Nat Struct Biol* **10**, 812-819.

23. Keefe, A.D. and Szostak, J.W. Functional proteins from a random-sequence library. (2001) *Nature* **410**, 715-718.

24. Labean, T.H. and Kauffman, S.A. Design of synthetic gene libraries encoding random sequence proteins with desired ensemble characteristics. (1993) *Protein Sci.* **2**, 1249-1254.

25. Kuntz, I.D., Chen, K., Sharp, K.A. and Kollman, P.A. The maximal affinity of ligands. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9997-10002.

26. Lo Surdo, P., Walsh, M.A. and Sollazzo, M. A novel ADP- and zinc-binding fold from function-directed *in vitro* evolution. (2004) *Nat Struct Mol Biol* **11**, 382-383.

27. Sassanfar, M. and Szostak, J.W. An RNA motif that binds ATP. (1993) *Nature* **364**, 550-553.

28. Kamtekar, S., Schiffer, J.M., Xiong, H., Babik, J.M. and Hecht, M.H. Protein design by binary patterning of polar and nonpolar amino acids. (1993) *Science* **262**, 1680-1685.

29. Wilson, D.S., Keefe, A.D. and Szostak, J.W. The use of mRNA display to select high-affinity protein-binding peptides. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3750-3755.

30. Schmidt, T., Koepke, J., Frank, R. and Skerra, A. Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. (1996) *J. Mol. Biol.* **255**, 753-766.

31. Voss, S. and Skerra, A. Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the *Strep*-tag II peptide and improved performance in recombinant protein purification. (1997) *Protein Eng.* **10**, 975-982.

32. Keefe, A.D., Wilson, D.S., Seelig, B. and Szostak, J.W. One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. (2001) *Protein Expr. Purif.* **23**, 440-446.

33. Schwarzbauer, J.E. and Sechler, J.L. Fibronectin fibrillogenesis: a paradigm for extracellular matrix assembly. (1999) *Curr. Opin. Cell. Biol.* **11**, 622-627.

34. Koide, A., Bailey, C.W., Huang, X. and Koide, S. The fibronectin type III domain as a scaffold for novel binding proteins. (1998) *J. Mol. Biol.* **284**, 1141-1151.

35. Xu, L., Aha, P., Gu, K., Kuimelis, R.G., Kurz, M., Lam, T., Lim, A.C., Liu, H., Lohse, P.A., Sun, L., Weng, S., Wagner, R.W. and Lipovsek, D. Directed

evolution of high-affinity antibody mimics using mRNA display. (2002) *Chem. Biol.* **9**, 933-942.

36. Evan, G.I., Lewis, G.K., Ramsay, G. and Bishop, J.M. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. (1985) *Mol. Cell. Biol.* **5**, 3610-3616.

37. Baggio, R., Burgstaller, P., Hale, S.P., Putney, A.R., Lane, M., Lipovsek, D., Wright, M.C., Roberts, R.W., Liu, R., Szostak, J.W. and Wagner, R.W. Identification of epitope-like consensus motifs using mRNA display. (2002) *J. Mol. Recognit.* **15**, 126-134.

38. Le-Nguyen, D., Heitz, A., Chiche, L., Castro, B., Boigegrain, R.A., Favel, A. and Coletti-Previero, M.A. Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. (1990) *Biochimie* **72**, 431-435.

39. Cujec, T.P., Medeiros, P.F., Hammond, P.W., Rise, C. and Kreider, B.L. Selection of v-Abl tyrosine kinase substrate sequences from randomized peptide and cellular proteomic libraries using mRNA display. (2002) *Chem. Biol.* **9**, 253-264.

40. Hammond, P.W., Alpin, J., Rise, C.E., Wright, M. and Kreider, B.L. *In vitro* selection and characterization of Bcl-X(L)-binding proteins from a mix of tissue-specific mRNA display libraries. (2001) *J. Biol. Chem.* **276**, 20898-20906.

41. Mcpherson, M., Yang, Y., Hammond, P.W. and Kreider, B.L. Drug receptor identification from multiple tissue using cellular-derived mRNA display libraries. (2002) *Chem. Biol.* **9**, 691-698.

42. Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C. and Schultz, P.G. A general method for site-specific incorporation of unnatural amino acids into proteins. (1989) *Science* **244**, 182-188.

43. Saks, M.E., R., S.J., Nowak, M.W., Kearney, P.C., Du, F., Abelson, J.N., Lester, H.A. and Dougherty, D.A. An engineered *Tetrahymena* tRNA[Gln] for *in vivo* incorporation of unnatural amino acids into proteins by nonsense suppression. (1996) *J. Biol. Chem.* **271**, 23169-23175.

44. Li, S., Millward, S. and Roberts, R.W. Applying an Unnatural Strategy to mRNA Display Libraries. (2002) *J. Am. Chem. Soc.* **124**, 9972-9973.

45. Frankel, A. and Roberts, R.W. *In vitro* selection for sense codon suppression. (2003) *Rna* **9**, 780-786.

46. Frankel, A., Millward, S.W. and Roberts, R.W. Encodamers: unnatural peptide oligomers encoded in RNA. (2003) *Chem Biol* **10**, 1043-1050.

47. Li, S. and Roberts, R.W. A novel strategy for *in vitro* selection of peptide-drug conjugates. (2003) *Chem Biol* **10**, 233-239.

48. Stemmer, W.P.C. Rapid evolution of a protein *in vitro* by DNA shuffling. (1994) *Nature* **370**, 389-390.

Chapter 3

Selection of RNA-Binding Peptides
Using mRNA-Peptide Fusions

Abstract

We are interested in the discovery of novel RNA binding peptides using *in vitro* selection. To do this, we use mRNA-protein fusions, peptides covalently attached to their encoding mRNA. Here, we report selection protocols developed using the arginine-rich domain of bacteriophage λ N protein and its binding target, the *boxBR* RNA. Systematic investigation of different selection paths has allowed us to design a reliable and efficient protocol to enrich RNA binding peptides from non-functional members in a complex mixture. This protocol should greatly facilitate the isolation of new molecules using the fusion system.

Introduction

There is great interest in creating peptides and proteins that bind nucleic acids with high affinity and specificity. *In vitro* and *in vivo* selection experiments are powerful techniques that provide functional solutions to nucleic acid recognition problems (1-5, and reviewed in 6). We have recently developed a novel strategy to perform *in vitro* peptide and protein selection using mRNA-protein fusions (mRNA display), proteins and peptides linked to their encoding mRNA (7, 8). Under optimal conditions, synthesis of up to 100 trillion (10^{14}) independent sequences is possible, the largest peptide or protein library available with any system (9).

In order to use mRNA display for the selection of RNA-binding peptides, we needed to design an efficient selection cycle. We have shown that the arginine-rich domain from the λ N protein retained the ability to bind its cognate *boxBR* target when synthesized as an mRNA-peptide fusion (9). However, after fusion synthesis, there is great flexibility in both the number and order of the steps that can be incorporated into the selection cycle.

The key to a successful selection experiment is the enrichment of functional sequences from non-functional sequences. Affinity selection, reverse transcription, and PCR are the only essential steps in a selection cycle (boldface, Figure 3.1). Other steps (e.g., affinity purification of the template, affinity purification of the peptide, a second affinity selection) may be added to improve the enrichment during the selection cycle. Experimental design must balance the advantages of additional steps (lower background, higher stringency selection) with the disadvantages (decreased product yield, increased

cycle time, PCR failure). A general goal is to achieve the maximum enrichment possible while still maintaining robust PCR after the selective step. An efficient selection cycle also maximizes the yield of product and reduces technical difficulties associated with multiple purification and enzymatic steps.
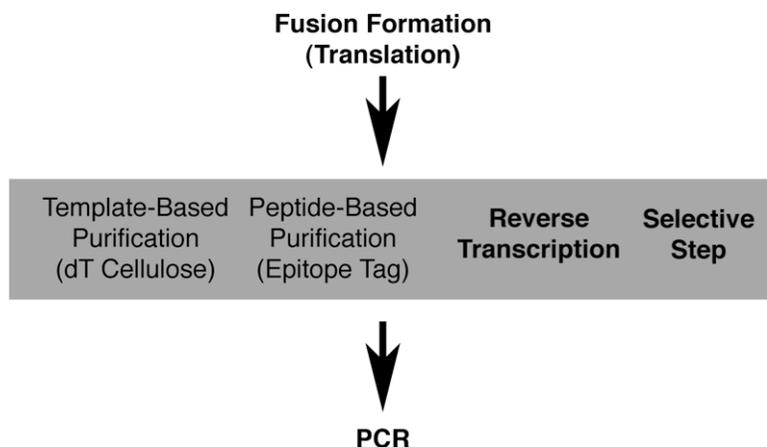
Figure 3.1. Path used for selection cycle using the fusion system. The order of steps follows from top-to-bottom, and left-to-right. Steps that cannot be omitted are shown in bold.

We have systematically investigated a number of selection paths with the goal of developing a robust *in vitro* selection cycle for RNA-binding peptides. Our optimized cycle represents a facile approach to isolate novel peptides with high affinity and specificity. Our current protocol follows all the steps in (Figure 3.1), from top to bottom and left to right. The methods we present should be generally applicable for the isolation of peptides and proteins that bind any immobilized target.

Results and Discussion

Fusion Synthesis

To begin a selection cycle, mRNA-peptide fusions must first be synthesized. The process has been described and optimized (7, 9). Briefly, mRNA containing a 3' puromycin is translated *in vitro* at ~400 nM template concentration. Monovalent and divalent cations (K^+, Mg^{+2}) are added after translation, facilitating fusion formation. The final product consists of an mRNA linked to the peptide it encodes through puromycin (Figure 3.2a). Reverse transcription allows conversion of the fusion product to the cDNA/mRNA hybrid (Figure 3.2b).
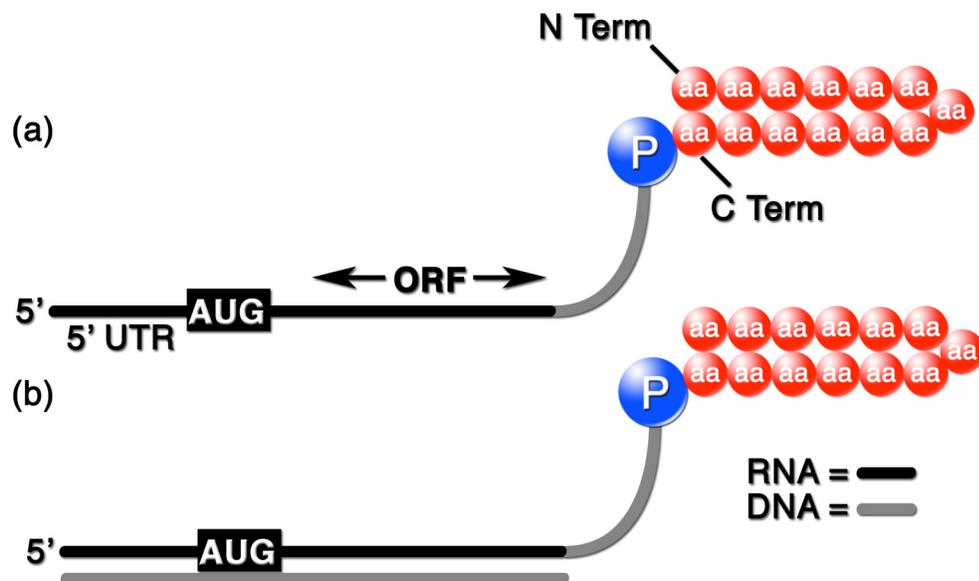
Figure 3.2. Two forms of mRNA-peptide fusions. **(a)** Schematic indicating the structure and connectivity of a mRNA-peptide fusion after synthesis on the ribosome. Linkage occurs between the C terminus of the peptide and the 3' end of the template through puromycin (P). **(b)** cDNA/mRNA-peptide fusion resulting from reverse transcription of the template.

Under these conditions, fusion synthesis is highly efficient and may be quantified in two ways: 1) the percent of the total synthesized peptide that is fused, or 2) the percent of input mRNA template that is converted to fusion product. After optimization, it is possible to convert up to 40% of the N template (Figure 3.3a) and up to 50% of the N peptide to fusion product (Figure 3.3b).
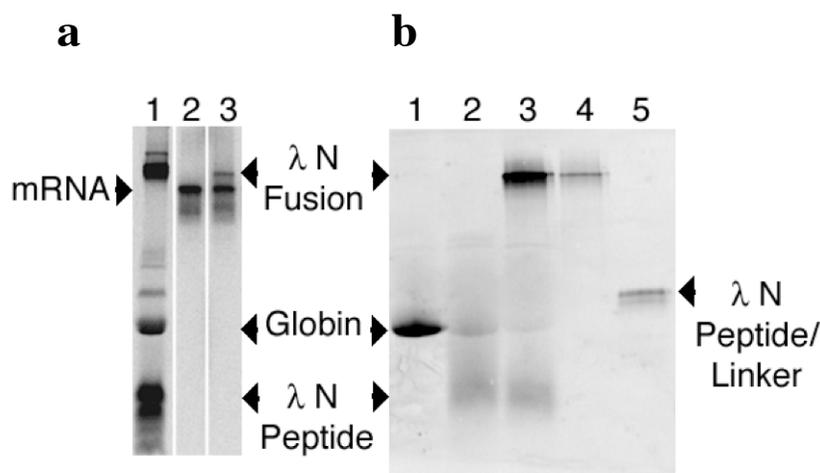
Figure 3.3. Efficiency of fusion synthesis. **(a)** The fraction of template converted to fusion. Translation of 400 nM ^{32}P-labeled template (lane 2) produces a fusion product with lower mobility (lane 3) as assayed by SDS-Tricine PAGE. Lane 1 shows ^{35}S-labeled fusion as a size standard. **(b)** The fraction of *in vitro* synthesized peptide converted to fusion. Translation of λ N ligated template (400 nM) in the presence of ^{35}S-methionine results in attachment of 50% of the peptide (lane 3). Subsequent dT-cellulose purification results in ^{35}S-labeled fusion (lane 4), which can be digest to peptide/DNA linker by RNase A (Lane 5). *In vitro* translated globin (MW ~16 kDa, Lane 1) and λ N peptide (Lane 2) are shown as a size standards.

Template-based Isolation

After translation and fusion formation, we isolate the input template from the translation reaction. Fusions are diluted into high salt buffer in the presence of dT-cellulose or dT-agarose, which hybridizes to the poly-dA repeat present in the end of the input template. The resulting product consists of a mixture of mRNA-peptide fusions,

unfused template, and any puromycin linker present in the reaction. Template-based isolation has many advantages as the first purification step after translation. First, it is highly efficient, allowing recovery of up to 90% of the input template (both fused and unfused alike). Second, it allows the removal of the bulk of the protein present in the translation reaction, including unfused proteins and nucleases or proteases present in the lysate (see below). Finally, it is very gentle, requiring no denaturants that could unfold the peptide or protein component.

Peptide/Protein-based Isolation

After dT purification, the sample contains a mixture of fused and unfused templates. It is possible to proceed directly to the selective step provided that it is stringent enough to remove nonfunctional sequences. Removal of unfused template is especially critical if the fusion efficiency is relatively low (7); a large excess of unfused template gives a very high background, making selective enrichment of functional sequences challenging. Maximal enrichments will likely be garnered if the unfused template is removed first.

In the initial demonstration of the system, fused molecules were purified from unfused molecules via disulfide bond chromatography (7). We compared the efficiency of the original protocol with immunoprecipitation using the FLAG epitope tag (DYKDDDDK). Constructs were generated that contained the FLAG epitope as well as a single cysteine near the C-terminus of the peptide sequence. The results (Table 3.1) show that the peptides bind and elute much more efficiently from the thiopropyl sepharose support as compared to the anti-FLAG support. Overall, thiopropyl sepharose

yields four times as much fusion material as does the FLAG affinity protocol, and was therefore incorporated into the selection cycle.

Table 3.1. Peptide-based purification of mRNA-peptide fusions

Separation Method	Binding (%)	Elution (%)	Total Yield (%)
Disulfide bond formation (thiopropyl-Sepharose)	80%	50%	40%
Immunoprecipitation (anti-FLAG agarose)	25%	40%	10%

Desalting

Often, the buffer conditions or concentration of fusions at the end of one step need to be changed in order to prepare for the next step. We have used two approaches to exchange the buffer: ethanol precipitation and ultrafiltration. Fusions and libraries containing the N peptide can be efficiently precipitated and resuspended (~90% overall yield) using linear acrylamide as carrier (see Materials and Methods). Ultrafiltration using filters of appropriate molecular weight cutoff also give similar results (see Materials and Methods).

When to RT?

In principle, reverse transcription (RT) may be performed at almost any point in the selection cycle. In practice, the RT step should be performed after purification of the fusion from the lysate but before the selective step due to the following observations.

The 3' Puromycin mRNA templates are quite stable in the reticulocyte lysate translation system. Northern analysis indicates little degradation over the course of an hour translation reaction or in the post-translational incubation step (A. Balakin and R.W. Roberts, unpublished observations). However, conversion of the mRNA to its DNA/RNA hybrid form (Figure 3.2b) in the lysate is likely to result in the destruction of the RNA portion of the template. During translation, addition of sub-stoichiometric amounts of oligonucleotide complementary to the RNA-DNA template junction causes significant degradation of the template (Figure 3.4). This observation is consistent with
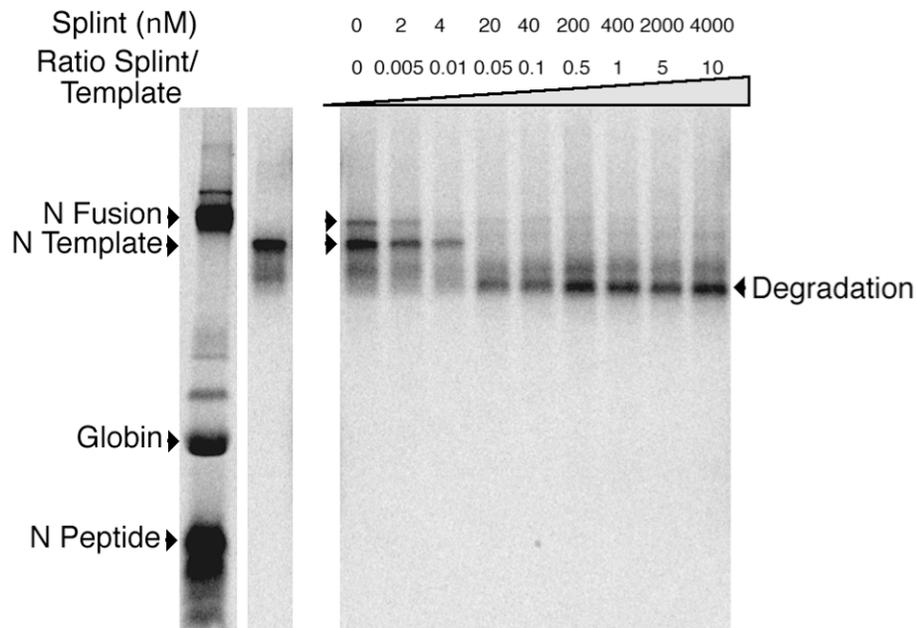
Figure 3.4. Splint-mediated degradation of template during translation. Varying amounts of splint oligonucleotide (see Materials and Methods) were added to translation reactions containing a ^{32}P-labeled template. The presence of sub-stoichiometric amounts of splint (0.05/1, splint to template) causes degradation of the template.

the presence of RNaseH activity in the lysate (10). Thus, the RNA portion of a cDNA/mRNA hybrid fusion will likely be degraded, destroying the physical linkage between the template and the peptide it encodes.

The efficiency of RT reactions depends on the amount of input template used. Prior to the selective step, there is generally enough template present such that the RT reaction proceeds with high efficiency (J. E. Barrick, T.T. Takahashi, R.W. Roberts unpublished observations). After the selective step, the amount of mRNA present is often 100-fold less than in the previous steps. Low template concentrations (below the Km of the enzyme) can result in inefficient or failure of reverse transcription, causing the selection cycle to fail. Finally, synthesis of the cDNA/RNA hybrid removes RNA tertiary structures from the library, greatly decreasing the likelihood of isolating RNA aptamers rather than functional peptides and proteins.

The Selective Step—Enrichment using an RNA Target

We have shown that N peptide synthesized in reticulocyte lysate is functional and binds to its immobilized RNA target (*boxBR*) (9). This peptide also showed a high degree of specificity in that it did not bind similar immobilized RNA structures, including the BIV-TAR site (11) and the HIV-RRE (12). The peptide also retained its RNA-binding activity when generated as an mRNA-peptide fusion, making it an excellent candidate for selection experiments (9).

Our preliminary results from selection experiments highlight the importance of controlling the stringency and specificity of binding. Selection experiments using an N

library indicate that the highest level of selective enrichment is attained when very large quantities of competitor are present in solution. Indeed, increasing the competitor (Yeast tRNA) concentration from 50 μg/mL to 5,000 μg/mL dramatically increases the efficiency of selection (13, Chapter 4). These results are consistent with those from the Pabo laboratory, where very large concentrations of competitor DNA were essential to isolate sequence specific zinc finger proteins (14). Varying other biophysical parameters such as increasing the temperature and/or salt concentration would also be expected to increase the stringency of selection.

Conclusions

Several systems have been developed or applied to isolate peptides and proteins that bind RNA including 1) bacterial suppression analysis (15), 2) the yeast 3-hybrid system (1), 3) a bacterial antitermination system (2), 4) a mammalian cell transcriptional activation system (4), and 5) phage display selections (16-18). *In vivo* systems have the advantage that they select for function in the context of cellular processes, whereas *in vitro* approaches provide access to larger libraries, reduced expression bias, and greater control over binding conditions.

In vitro selection using mRNA-peptide fusions presents a powerful addition to this list. Fusions containing the N peptide retain binding affinity and specificity to their cognate RNA (9, 13). It is therefore likely that other arginine-rich peptides (such as BIV-Tat and HIV-Rev) or other RNA binding domains (such as zinc fingers or the RNP motif) may also serve as facile starting points for fusion-based selections. Large sequence

complexities (up to 10^{14} sequences, 10,000-fold more than phage display) are accessible with the fusion system, and should allow the discovery of rare functional sequences that could not be isolated with other systems. Finally, once isolated, functional sequences may be rapidly optimized by addition of *in vitro* recombination (19, 20) and mutagenesis (21) to the PCR portion of the selection cycle

Materials and Methods

Construction of Fusion Template

A double stranded DNA template encoding the 22 amino acid RNA-binding domain of phage λ N (underlined) (22) fused to a FLAG epitope (DYKDDDDK) followed by the amino acid sequence, NSCA (peptide sequence: MDAQTRRRERRAEKQAQWKAAN-DYKDDDDKNSCA, N-FLAG-myc) was constructed from a synthetic deoxyoligonucleotide template (5'-GGGACAATTACTATTTACAATTACAATGGACGCCCAGACCCGCCGGCGCGAGCGCAGGGCCGAGAAGCAGGCCCAGTGGAAGGCCGCCAACGACTACAAGGACGACGATGACAAG-3') and two primers, Fmyc (5'-AGCGCAAGAGTTCTTGTCATCGTCGTCCTTGTAGTC-3'), and 42.108 (5'-TAATACGACTCACTATAGGGACAATTACTATTTACAATTACA-3'). A DNA pool was constructed in a similar fashion from a synthetic template (5'-GGGACAATTACTATTTACAATTACAATGGACGCCCAGACCNNCBNGCGCGAGCGCAGGGCCGAGAAGCAGGCCCAGTGGAAGGCCGCCAACGACTACAAGGACGACGATGACAAG-3'), containing the sequence 5'-NNCBNG-3' (N = A,T, G, or C; B = C,G, or T) at codons 6 and 7. mRNA was produced by T7 run-off transcription of

these templates (23) in the presence of RNAsecure (Ambion) followed by gel purification via denaturing urea PAGE and electroelution. The flexible DNA linker containing puromycin, F30P (5' $dA_{21}[C9]_3dACdCP$; C9 = triethylene glycol phosphoramidite (Spacer 9, Glen Research), P = CPG-puromycin, (Glen Research)), was synthesized using standard phosphoramidite chemistry, chemically phosphorylated using phosphorylation reagent II (Glen Research), and purified by OPC cartridge (Glen Research). Approximately 30% of the full-length mRNA transcript has the proper 3' nucleotide for ligation to the purified flexible DNA linker F30P using the DNA splint (5'-TTTTTTTTTTAGCGCAAGAGTT-3'). Ligation reactions consisted of mRNA, F30P, and splint in a 1:1.5:1 ratio, respectively, with 0.8 U of T4 DNA ligase (New England Biolabs) per pmol of template RNA. After ligation, the fusion template was gel purified, electroeluted, and desalted by ethanol precipitation.

Translation and Fusion Formation

Fusion template was translated in reticulocyte lysate (Novagen) using conditions optimized for N peptide translation (400 nM template, 1.0 mM MgOAc, 100 mM KOAc). Upon completion of translation, fusion formation was stimulated by addition of $MgCl_2$ and KCl to 50 mM and 0.50 M, respectively, and incubation at –20 ºC for more than 4 hours.

Template-Based (dT) Purification

Following post-translational incubation, the lysate was diluted and incubated with biotinylated dT_{25} bound to ImmunoPure® immobilized streptavidin (Pierce) at 4°C in

isolation buffer (IB) (100 mM Tris-HCl, pH 8.0, 1.0 M NaCl, 0.2% (v/v) Triton X-100 (Sigma)) for 1-2 hrs. In some experiments, fusion was isolated in isolation buffer using dT-cellulose (New England Biolabs). In both cases, bound fusions were washed with isolation buffer, eluted with ddH$_2$O, and concentrated by ethanol precipitation in the presence of 0.3 M NaOAc, pH 5.2, and linear acrylamide (20 μg/mL, Ambion).

Peptide-Based Purification

For disulfide bond chromatography, 20 μL of a 50/50 (v/v) slurry of prewashed thiopropyl (TP) sepharose (Pharmacia) and 500 μL of 0.2 M NaOAc buffer pH 4.0 were added to a mixture of fusions and incubated at 4 °C for 1-2 hours. Samples were washed three times with 0.1 M NaOAc buffer and eluted with 50 mM DTT in 1X TBE, pH ~8. Samples were desalted either by ethanol precipitation in the presence of linear acrylamide, or by ultrafiltration using preblocked YM-10 or YM-30 Microcon filters (Millipore). Filters were blocked by addition of 500 μL of blocking buffer (1% BSA (w/v) in 1X PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na$_2$HPO$_4$-7H$_2$O, 1.4 mM KH$_2$PO$_4$) and 0.2 μm-filtered), incubated for >2 hours, and centrifuged at 14,000xg. Filters were then washed once with 500 μL of H$_2$O, and centrifuged at 1000xg. For purification by FLAG immunoprecipitation, 20 μL of prewashed Anti-FLAG M2 Affinity Gel (Sigma) and 500 μL of 2X Tris-buffered saline (TBS, 50 mM Tris-HCl, pH 7.4, 150 mM NaCl) was added to dT purified material and the mixture incubated at 4 °C for one hour followed by three washes of 1X TBS. Fusions were eluted either by incubation with 0.2 μg/mL FLAG peptide at 4 °C for 1 hour or by incubation with 0.1 M glycine-HCl, pH 3.5 for 15 minutes and then quantified by scintillation counting.

Reverse Transcription

Reverse transcription with Superscript II RNAse H⁻ Reverse Transcriptase (BRL, Life Technologies) was according to the manufacturer's specifications. Addition of 2 to 10 equivalents of primer to template quantitatively generated cDNA-mRNA hybrid from the fusion template or fusion itself.

Splint Doping

λ N mRNA was generated using T7 runoff transcription in the presence of $(\alpha\text{-}^{32}P)$-UTP (NEN) and ligated to the F30P oligonucleotide as described above. 0.02, 0.04, 0.2, 0.4, 2, 4, 20, and 40 pmol of splint were added to the reaction tubes and dried. Translation mixes were directly added to these tubes (as described above) and the products run on a 5% (stack)/15% (separating) tricine gel (24) and quantified by Phosphorimager (Molecular Dynamics).

Selective Step

Biotinylated *boxBR* hairpin (5'-GGCCCUGAAAAAGGGCCAAA-Biotin-3') was immobilized on streptavidin agarose (Pierce) pre-equilibrated in N binding buffer (10 mM HEPES pH 7.5, 100 mM KCl, 1 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT; 0.01% (v/v) NP-40) with 50, 500, or 5000 μg/mL of yeast tRNA (Boehringer Manheim). cDNA/mRNA fusion was added to the binding reaction and incubated at 4 °C for 1 hour. The agarose beads were washed three times with 1X binding buffer. Thirty microliters of ddH_2O and 1 μL of RNaseA (Roche) were added and then incubated at 37 °C for 30

minutes to elute the bound fusions. One-tenth the volume of supernatant (~5 μL) was taken and 18 cycles of PCR performed to generate an enriched pool.

References

1.   Sengupta, D.J., Zhang, B., Kraemer, B., Pochart, P., Fields, S. and Wickens, M. A three-hybrid system to detect RNA-protein interactions *in vivo*. (1996) *Proc Natl Acad Sci U S A* **93**, 8496-8501.

2.   Harada, K., Martin, S.S. and Frankel, A.D. Selection of RNA-binding peptides *in vivo*. (1996) *Nature* **380**, 175-179.

3.   Greisman, H.A. and Pabo, C.O. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. (1997) *Science* **275**, 657-661.

4.   Tan, R. and Frankel, A.D. A novel glutamine-RNA interaction identified by screening libraries in mammalian cells. (1998) *Proc Natl Acad Sci U S A* **95**, 4247-4252.

5.   Danner, S. and Belasco, J.G. T7 phage display: A novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. (2001) *Proc Natl Acad Sci U S A* **98**, 23.

6.   Roberts, R.W. and Ja, W.W. *In vitro* selection of nucleic acids and proteins: What are we learning? (1999) *Curr Opin Struct Biol* **9**, 521-529.

7.   Roberts, R.W. and Szostak, J.W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. (1997) *Proc Natl Acad Sci U S A* **94**, 12297-12302.

8.   Roberts, R.W. Totally *in vitro* protein selection using mRNA-protein fusions and ribosome display. (1999) *Curr Opin Chem Biol* **3**, 268-273.

9.   Liu, R., Barrick, J.E., Szostak, J.W. and Roberts, R.W. Optimized synthesis of RNA-protein fusions for *in vitro* protein selection. (2000) *Methods Enzymol* **318**, 268-293.

10.   Cazenave, C., Frank, P. and Busen, W. Characterization of ribonuclease H activities present in two cell-free protein synthesizing systems, the wheat germ extract and the rabbit reticulocyte lysate. (1993) *Biochimie* **75**, 113-122.

11.   Puglisi, J.D., Chen, L., Blanchard, S. and Frankel, A.D. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. (1995) *Science* **270**, 1200-1203.

12.   Battiste, J.L., Mao, H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D. and Williamson, J.R. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. (1996) *Science* **273**, 1547-1551.

13.   Barrick, J.E., Takahashi, T.T., Ren, J., Xia, T. and Roberts, R.W. Large libraries reveal diverse solutions to an RNA recognition problem. (2001) *Proc Natl Acad Sci U S A* **98**, 12374-12378.

14.   Rebar, E.J., Greisman, H.A. and Pabo, C.O. Phage display methods for selecting zinc finger proteins with novel DNA-binding specificities. (1996) *Methods Enzymol* **267**, 129-149.

15.   Franklin, N.C. Clustered arginine residues of bacteriophage lambda N protein are essential to antitermination of transcription, but their locale cannot compensate for *boxB* loop defects. (1993) *J Mol Biol* **231**, 343-360.

16.   Laird-Offringa, I.A. and Belasco, J.G. Analysis of RNA-binding proteins by *in vitro* genetic selection: identification of an amino acid residue important for

locking U1A onto its RNA target. (1995) *Proc Natl Acad Sci U S A* **92**, 11859-11863.

17.   Friesen, W.J. and Darby, M.K. Phage display of RNA binding zinc fingers from transcription factor IIIA. (1997) *J Biol Chem* **272**, 10994-10997.

18.   Blancafort, P., Steinberg, S.V., Paquin, B., Klinck, R., Scott, J.K. and Cedergren, R. The recognition of a noncanonical RNA base pair by a zinc finger protein. (1999) *Chem Biol* **6**, 585-597.

19.   Stemmer, W.P. Rapid evolution of a protein *in vitro* by DNA shuffling. (1994) *Nature* **370**, 389-391.

20.   Stemmer, W.P. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. (1994) *Proc Natl Acad Sci U S A* **91**, 10747-10751.

21.   Cadwell, R.C. and Joyce, G.F. Mutagenic PCR. (1994) *PCR Methods Appl* **3**, S136-140.

22.   Cilley, C.D. and Williamson, J.R. Analysis of bacteriophage N protein and peptide binding to *boxB* RNA using polyacrylamide gel coelectrophoresis (PACE). (1997) *RNA* **3**, 57-67.

23.   Milligan, J.F. and Uhlenbeck, O.C. Synthesis of small RNAs using T7 RNA polymerase. (1989) *Methods Enzymol* **180**, 51-62.

24.   Schagger, H. and Von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. (1987) *Anal Biochem* **166**, 368-379.

Chapter 4

Large Libraries Reveal Diverse Solutions
to an RNA Recognition Problem

Abstract

RNA loops that adopt the characteristic GNRA "tetraloop" fold are common in natural RNAs. We selected peptides that bind an example of this RNA loop motif using mRNA display. Starting with the RNA recognition domain from the λ N protein, we constructed libraries containing 150 and 1,600 different peptide sequences as mRNA-peptide fusions and isolated those capable of high-affinity RNA binding. These selections have resulted in almost 20 different peptides that bind the same RNA loop with nanomolar affinity. Analysis of one peptide complex by fluorescence spectroscopy and NMR spectroscopy suggest a different binding mode from the wild-type peptide. Our work demonstrates that multiple, chemically and conformationally distinct solutions exist for a particular RNA recognition problem.

Introduction

The ability to construct high-affinity, high-specificity peptide ligands provides a means to target RNA molecules of interest. Genetic approaches have been developed to isolate novel arginine-rich RNA-binding peptides *in vivo* (1-3). These systems allow selection in the context of living systems, but limit library sizes to a maximum of ~10^5 to 10^6 sequences, allowing only four residues to be searched exhaustively ($20^4 = 1.6$ x10^5). *In vitro* selection experiments to isolate RNA-binding peptides have not been demonstrated. A totally *in vitro* approach affords precise control over the selection conditions and the ability to explore much larger libraries, enabling the isolation of rare sequences.

We sought to use *in vitro* selection to isolate peptides that bind RNA tetraloops. The tetraloop fold is a common element in many functional RNAs, enhancing duplex stability and participating in tertiary folding interactions (4-6). We developed mRNA display (also known as the mRNA-peptide fusion system) to perform *in vitro* selection of peptides and proteins (7). In mRNA display, cycles are carried out entirely *in vitro* and libraries larger than 10^{13} independent sequences can be constructed (Figure 4.1, (7-10)).

The arginine-rich peptide corresponding to the RNA-binding domain of the λ N protein served as a starting point for our experiments and libraries. This short peptide (22 amino acids) recognizes the *boxB* RNA hairpin with high affinity and specificity as a bent α-helix (11-14). The hairpin contains a five-base RNA loop that adopts a tetraloop fold with one base extruded (15, 16). Here, we have used mRNA display to isolate peptides that bind the *boxBR* RNA motif.
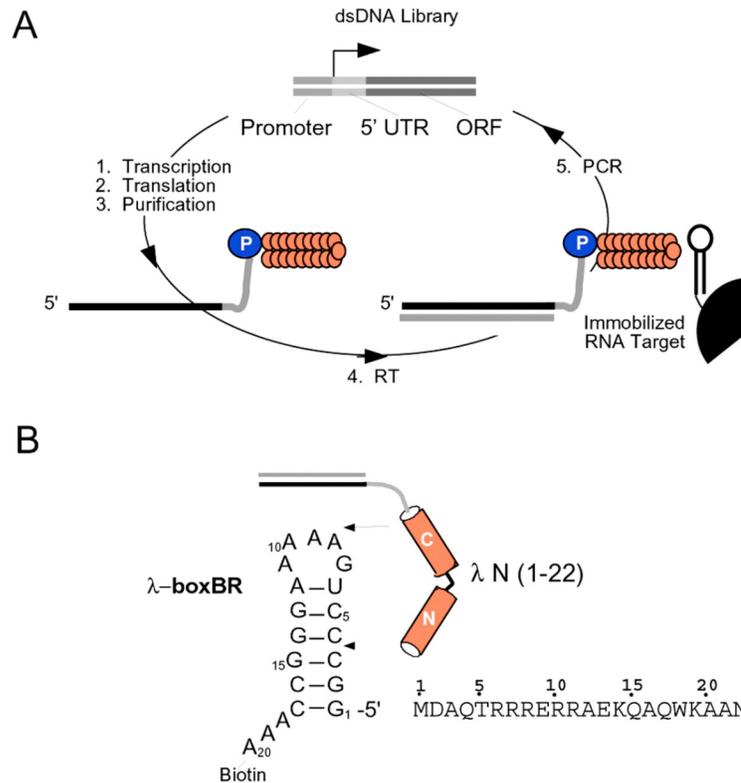
Figure 4.1. **(a)** Selection cycle. In the selection cycle, a double-stranded DNA library containing randomized codons is transcribed, generating a pool of mRNA templates. These templates are then ligated to a flexible DNA oligonucleotide containing puromycin at its 3' end. Translation of these ligated templates *in vitro* produces peptides covalently attached via their C-terminus to the 3' end of their own message by way of a stable amide linkage a mRNA-peptide fusion. These fusions are then converted into cDNA/mRNA hybrid fusions by using reverse transcriptase and subjected to selection on an affinity matrix. **(b)** Constructs used. λ N (1-22, amino acid sequence shown) binds to *boxB* as a bent α-helix.

Results and Discussion

Peptide Fusion Binding

It was unclear *a priori* whether an RNA-binding peptide would be functional as either an mRNA-peptide or cDNA/mRNA-peptide fusion. We synthesized mRNA-peptide fusions containing the λ N RNA binding domain (17) (Figure 4.1B) and tested the ability of these molecules to bind an immobilized RNA target. (Figure 4.2) shows

that both the mRNA-peptide fusion and the cDNA/mRNA-peptide fusion specifically

bind the *boxBR* RNA target. These assays demonstrate that a significant fraction of the

desired complex can be isolated (30-60%) whereas little (1-5%) is retained to noncognate

biotinylated RNA targets (Rev Response Element, RRE; U1 RNA hairpin loop II, U1).

In Vitro Selection of λ N Peptides

We designed two selections of increasing complexity to test our ability to isolate

novel tetraloop-binding peptides. In selection 1, we randomized positions 6 and 7 of the
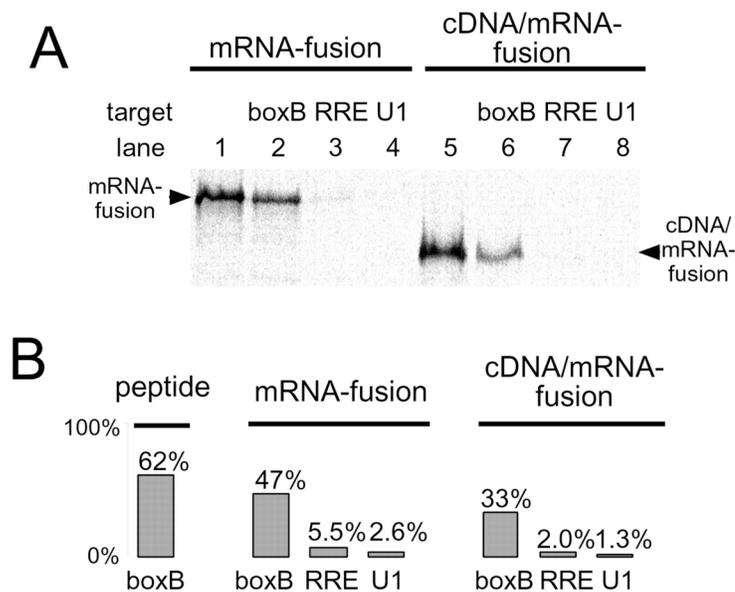
Figure 4.2. Binding and specificity of N peptide N mRNA-peptide fusion, and N cDNA/mRNA-peptide fusion constructs. **(a)** Gel analysis of binding. Binding of ^{35}S-Met-labeled N mRNA-peptide fusions and cDNA/mRNA-peptide fusions to immobilized (i) *boxBR* (lanes 2 and 6), (ii) HIV-RRE (lanes 3 and 7) and (iii) U1 hairpin (lanes 4 and 8). Lanes 1 and 5 show the total amount of fusion before binding. Complexes were eluted and resolved on SDS tricine PAGE (27). **(b)** Scintillation analysis of binding and specificity. ^{35}S-Met-labeled peptide, mRNA, and mRNA/cDNA fusions bound to immobilized *boxBR*, HIV-RRE, and U1 hairpins. The cDNA/mRNA and mRNA fusions bind the cognate *boxBR* RNA efficiently, whereas only 1-5% bind the RRE and U1 targets.

wild-type construct (R6X, R7X, henceforth R67X). Previous mutagenesis and structural work indicated a single sequence, containing arginine at both positions (R6 and R7, henceforth R6R7), should be optimal for binding (1, 13). The cassette contained 150 different amino acid combinations, only one of which was R6R7, and was designed such that the R6R7 sequence could be recognized and cut by the restriction enzyme NgoM IV (Figure 4.3). We performed selection 1 (one round), three times in parallel, using increasing concentrations of competitor yeast tRNA to provide high stringency selection. Selection under stringent conditions is expected to increase the enrichment of functional sequences (18-20).

Restriction analysis by NgoM IV indicated that a majority (>85%) of the selected clones contained R6R7 when 5 mg/mL tRNA competitor was used (Figure 4.3). Sequencing of clones selected under these conditions showed that 8/11 (72%) contained R6R7 (Figure 4.4) and 50% of other sequences contained R7, expected to be the more important of the two positions (1). Thus, increased competitor allowed good enrichment (50- to 150-fold per round) without forcing us to make the target the limiting reagent (21).

In selection 2 (Figure 4.5A, library 2) we randomized positions 7, 14, and 15, for a total of 1,600 possible combinations. In the wild-type N/*boxBR* complexes, residues 14 and 15 lie at the RNA-peptide interface (15, 16) and are important for binding (1, 14). After four rounds of selection, pool binding was similar to the wild-type sequence (Figure 4.5B). Analysis of 39 round 4 clones yielded 19 different sequences (Figure 4.5C). Each
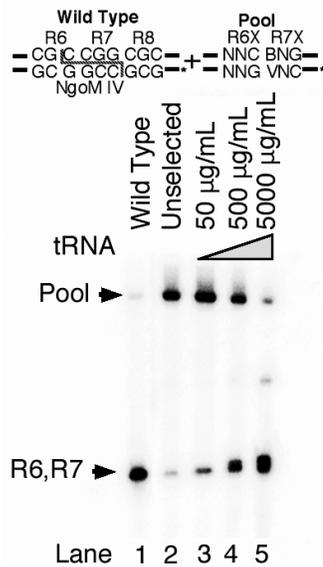
Figure 4.3. Restriction analysis of the R67X Pool. The R67X Pool (Lane 2) was subjected to one round of *in vitro* selection in the presence of increasing concentrations of competitor yeast tRNA (Lanes 3-5). DNA was PCR amplified with a ^{32}P-labeled primer and cut with NgoM IV. Only the wild-type sequence (R6R7) is cut by NgoM IV (Lane 1). Under the highest concentration of competitor tRNA (5 mg/mL) almost all (>85%) of the recovered sequences are cut by NgoM IV, and therefore correspond to the wild-type sequence.

Figure 4.4. Sequencing of the R67X pool after one round of selection with 5 mg/mL tRNA competitor. Sequences corresponding to the wild-type sequence R6R7 are shown in blue. For clarity, only amino acids differing from the wild-type sequence are shown.

A   Library 2   MDAQTR4RERRAEXXAQWKAANDYKDDDDKNSCA

B

C

```
                                              Relative
                                            Immobilized
                                            boxBR binding
1         10          20          30
------R------KQ------------------          100
------R------KQT-----------------           83
------R------AR-R----------------           83
------R------ERV-----------------           83
------R------KI------------------           83
------R------LR------------------           83
------R------LQ------------------           66
------R------MR------------------           61
------R------NR------------------           51
------R------KT------------------           46
------R------KR------------------           40
------R------KA------------------           39
------R------RS------------------           34
------R------RT------------------           33
------R------MF------------------           31
------R------KM------------------           30
------R------SR------------------           28
------R------RR------------------           24
------R------KH------------------           20
```
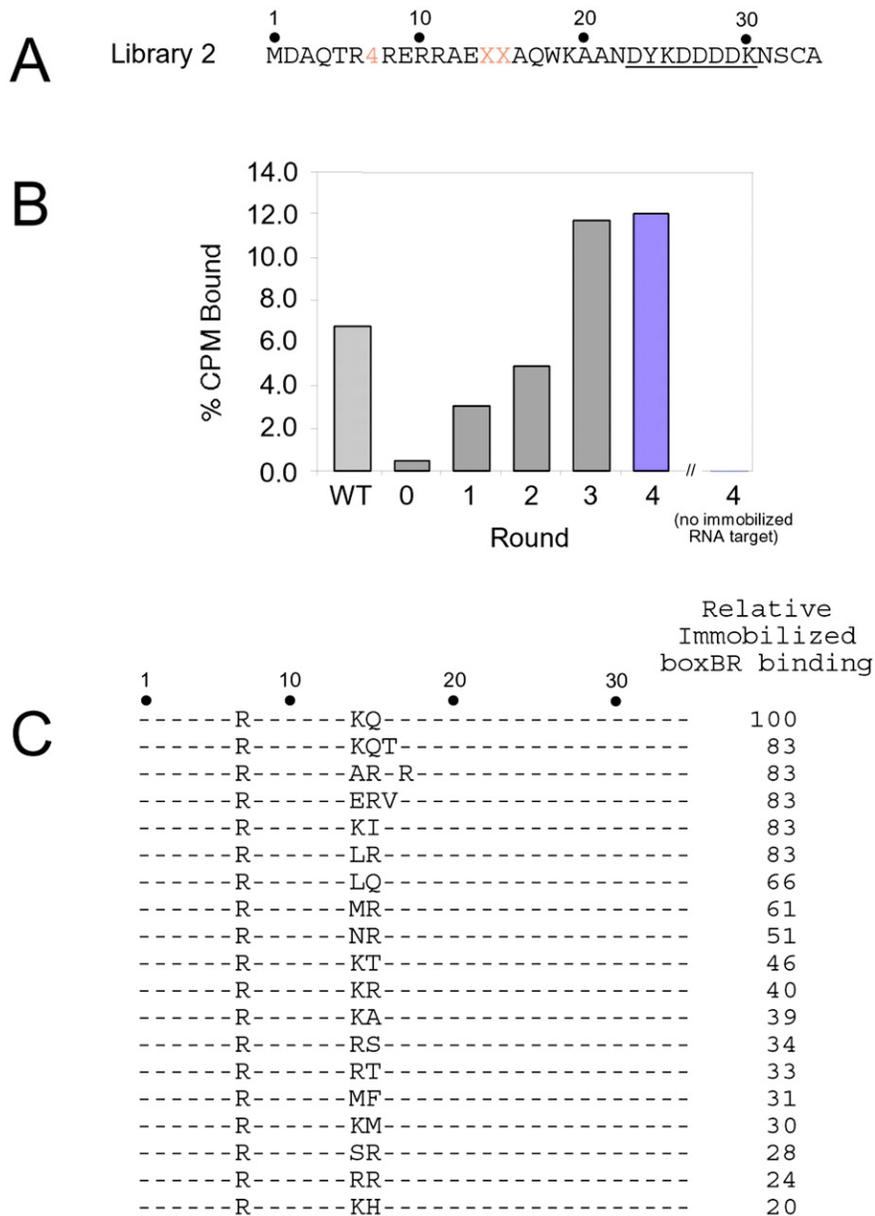
Figure 4.5. Selection of *boxBR* binding peptides from library 2. **(a)** Sequence of library 2. The target is the *boxBR* RNA (Figure 4.1B). Peptide position 7 contains a cassette encoding 4 amino acids, whereas positions 14 and 15 encode all 20 possible amino acids and one stop codon (NNG/C codons). **(b)** Binding to immobilized *boxBR* RNA for rounds 0-4 measured by using ^{35}S-Met-labeled fusions. Controls using the wild-type complex (WT) and agarose with no immobilized RNA (no *boxBR*) are shown. Material from the fourth round (colored bar) was cloned and sequenced. **(c)** Peptide sequences and peptide-binding affinities of selected clones. Binding affinities are given relative to the wild-type sequence in the same construct.

peptide was constructed by *in vitro* translation, and assayed for binding to immobilized *boxBR* RNA (Figure 4.5C) by means of a binding depletion assay (Figure 4.S3). All of the corresponding peptides bind *boxBR* and contain arginine at position 7. Aside from the wild-type (K14Q15), only three of the highest-affinity clones retain one of the two original residues. Indeed, a different consensus sequence containing R15 emerges as a dominant motif; roughly half of the selected sequences contain R15. These sequences include E14R15, which replaces lysine-14 with glutamic acid, the consensus sequence in a selection randomizing residues 13-22 of λ N (22).

Characterization of the E14R15-*boxB* Complex

Peptide affinities were determined by fluorescence titration against 2-aminopurine (2AP) labeled *boxBR*. 2-aminopurine is a fluorescent nucleoside analog of adenine that has been previously substituted in RNA (23). Peptides corresponding to wild-type λ N(1-22) and E14R15(1-22) were constructed synthetically and titrated against *boxBR* labeled at position 7 (2AP-7). Both peptides bind with high affinity to the target site (1.0 nM vs. 8.5 nM for wild-type and E14R15, respectively, Table 4.1). The difference is small considering that a single change of K14 to A14 results in loss of more than 2 kcal/mol (>30-fold) (13). The charge reversal of glutamic acid for lysine at position 14 is puzzling and may result from favorable interaction of glutamic acid with the helix dipole of residues 12-22 (24, 25).

Addition of wild-type N(1-22) produces strong quenching of 2AP fluorescence, whereas addition of E14R15 enhances fluorescence (Figure 4.6). In previous work, 2AP

quenching has been attributed to increased stacking, while fluorescence enhancement is attributed to increased exposure to aqueous solvent (23). Quenching of 2AP in the N (1-22)/*boxBR* complex is expected, as W18 stacks on position 7 (15, 16) and tryptophan fluorescence is quenched in the complex (13), and see below).

Table 4.1. Binding constants and specificity of selected and control peptides. K_d values are in nM.

Peptide	*boxBR* (2AP-7)
λ N(1-22)	1.0 ± 0.2
E14R15(1-22)	8.5 ± 2.0
Controls	
λ N(1-11)	$1,290 \pm 20$
E14R15(1-15)	303 ± 8

Binding constants were determined by fluorescence titration at 20 °C, 50 mM KOAc, 20 mM TrisOAc, pH 7.5. All peptides contain a free amino and carboxy terminus. λ N(1-11) and E14R15(1-15) contain a C-terminal GY sequence to facilitate quantitation. 2AP-7 denotes a 2'-methoxy 2-aminopurine residue inserted at the second loop position. Error estimates indicate the precision of individual fits. Data for ERV for selection 2 indicates the binding is similar to E14R15(1-22) (unpublished observation).
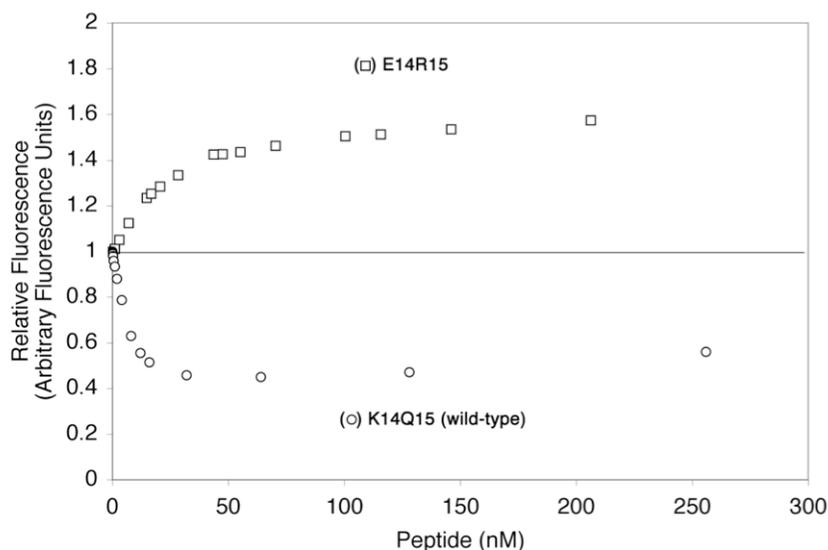
Figure 4.6. Binding isotherms for K14Q15 (wild-type) N peptide (circles) and E14R15 peptide (squares). The E14R15 peptide was added to 2AP-7 labeled *boxB* RNA resulting in an increase in fluorescence. Addition of wild-type peptide to 2AP-7 labeled RNA results in an initial decrease, followed by a gradual increase in fluorescence because of tryptophan fluorescence.

We also investigated tryptophan fluorescence in the E14R15 complex. Figure 4.7 shows that addition of *boxB* RNA to wild-type peptide results in a dramatic decrease in W18 fluorescence, as reported previously (13). However, addition of *boxB* RNA to the E14R15 peptide results in an increase of W18 fluorescence by ~10% relative to the unbound peptide. Taken together, these data imply that W18 is less stacked on adenine 7 in the E14R15 complex.

One- and two-dimensional NMR of the N/*boxBR* complexes provide information about the structure of both the bound peptide and RNA. Binding of the wild-type peptide results in a ~ 1 p.p.m. upfield shift of the Trp indole proton relative to the unbound peptide because of ring current effects in the stacked state (15, 16). However, in the E14R15 complex, W18 is unshifted, indicating that it is not stacked in the complex (Figure 4.8); the chemical shift of the E14R15 W18 indole proton is similar to that of the
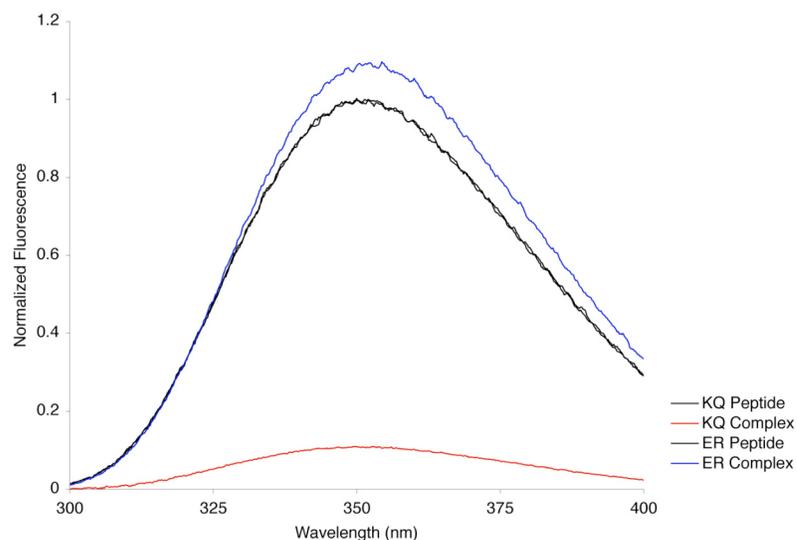
Figure 4.7. Tryptophan fluorescence of N peptide and E14R15 peptide free and in complex with *boxB* RNA. Addition of *boxB* RNA to wild-type N peptide results in a dramatic quenching of Trp fluorescence (red line). Addition of *boxB* RNA to E14R15 peptide results in a slight increase in Trp fluorescence (blue line).

free unbound peptide. Both the wild-type and E14R15 complexes show the same number of imino protons, including the G6-A10 sheared pair imino at ~10.8 p.p.m. (15, 16) implying that the RNA structure is similar in both complexes. The ^{15}N-HSQC of the free E14R15(A16V) peptide is similar to the free wild-type peptide (Figure 4.9). Both spectra show poor chemical dispersion in the proton dimension, characteristic of unfolded peptides. Addition of unlabeled *boxBR* RNA to E14R15 results in peak shifts and increased chemical shift dispersion in the proton dimension, suggesting a folded structure (26). The Ala3 amino proton is shifted downfield to ~9.3 p.p.m. in the E14R15 complex, suggesting that the N-terminal domain of the peptide binds in a manner similar to the wild-type.
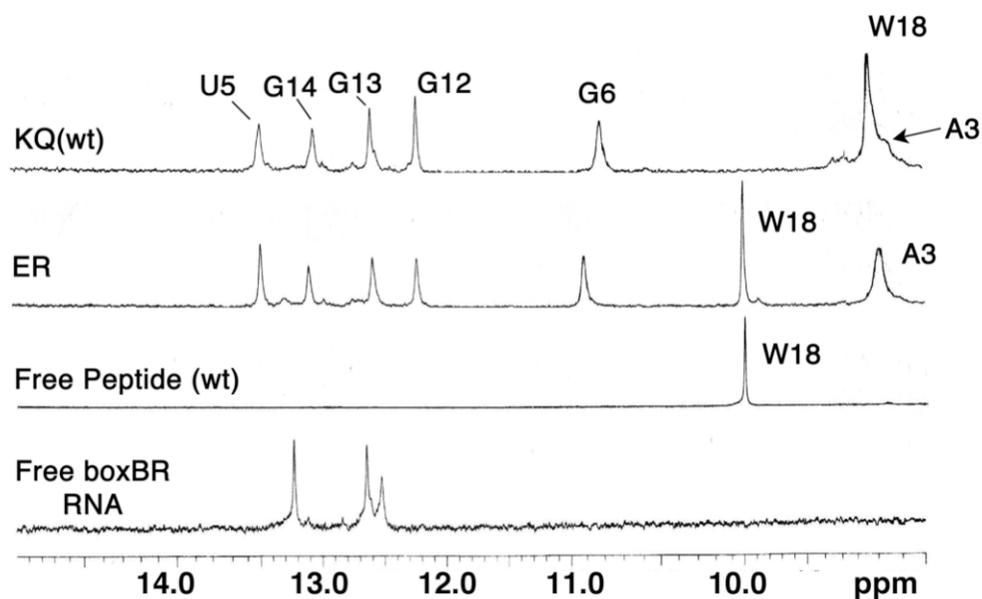
Figure 4.8. ¹H-NMR spectra of wild-type and ER peptides in complex with *boxB* RNA. Both complexes show similar patterns of imino spectra. Free wild-type peptide and *boxB* are shown as controls.
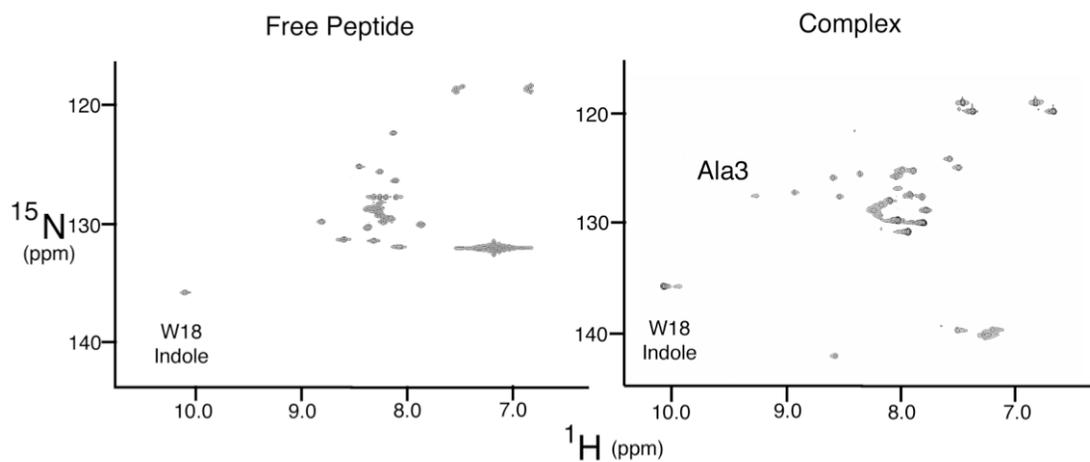
Figure 4.9. ¹⁵N-HSQC of the E14R15(V16) peptide alone (left) and in complex with *boxB* RNA (right). Upon addition of the mRNA increased chemical shift dispersion in the ¹H dimension is observed, consistent with the peptide folding. Highlighted are the W18 indole and Ala3 amino peaks.

Neither peptide shows any appreciable structure in the absence of RNA as judged by circular dichroism (CD) (Figure 4.10). The difference spectra of the complexes indicate that both peptides fold into α-helices when *boxBR* RNA is added (Figure 4.10). Although globally similar, the two complexes display differences in regions indicative of peptide folding (200-225 nm) and RNA folding (260-300 nm). E14R15 is somewhat less helical than the wild-type peptide.
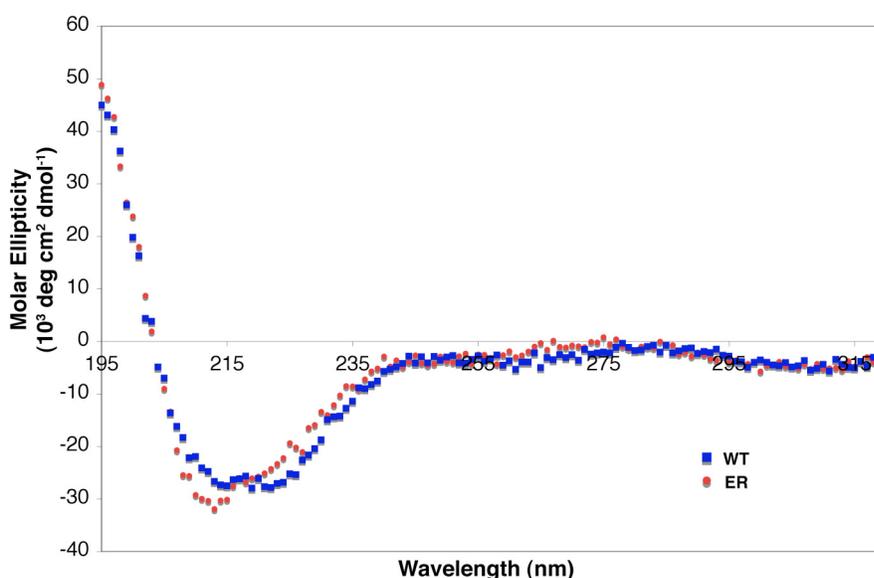
Figure 4.10. CD spectra of wild-type N and E14R15 peptides in complex with *boxB* RNA.

The biophysical data described above suggest that the E14R15 complex with *boxB* is similar in many respects to the wild-type complex. First, the E14R15 peptide binds with near wild-type affinity and folds into an α-helix upon RNA binding. In the bound structure, the RNA and the N-terminal region of E14R15 are structurally similar to the wild-type. However, the tryptophan in the E14R15 complex does not stack on the RNA, and therefore the C-terminus of E14R15 is most likely in a different conformation

than in the wild-type structure. This change in structure could disrupt native contacts between residues 16-22 of the E14R15 peptide and the RNA; the loss of the W18-RNA stacking interaction results in a loss of more than 2 kcal/mol of binding free energy (13).

We synthesized two model peptides, N (1-11) and E14R15(1-15), to address the role of the residues between positions 13 and 22. The E14R15(1-15) peptide reveals that addition of R15 cannot be the sole feature that confers low nanomolar binding to the selected peptides. Adding the sequence containing E14R15 to the N (1-11) peptide confers only a 3- to 5-fold enhancement in K_d, a G° of 0.7 to 1.0 kcal/mol (Table 4.1 E14R15(1-15)). This result indicates that the amino acids between positions 16 and 22 contribute ~3 kcal/mol of binding free energy to the interaction.

Conclusions

Our results demonstrate that high-affinity RNA ligands may be isolated using mRNA display. Starting with the RNA binding domain of λ N, we isolated nearly 20 peptides that retain the ability to bind the *boxB* RNA. A non-wild-type consensus, containing arginine at position 15, emerges in the selected peptides leading to the question of why nature did not select for R15 in the λ N peptide. Perhaps the E14R15 peptide suggests a potential answer: while the E14R15 peptide possesses similar affinity to the wild-type, W18 is not stacked in the complex suggesting a different structure in the C-terminus of the peptide. It is interesting that a change of only two amino acids results in such a dramatic change in structure, and this structural change has little affect on binding affinity. Future experiments will address several of these issues. These data are consistent with a model where RNA-peptide interactions are highly context dependent.

In that vein, the combinatorial approach we have used provides a powerful design strategy for interactions where rules may be idiosyncratic at best.

Materials and Methods

Construction of Fusion Template

Templates and constructs used to assay fusion binding (Figure 4.2) correspond to the N-myc construct described previously (17).

Binding Analysis

Chemically synthesized 3' biotinylated RNA oligonucleotides of the *boxBR* hairpin (5'-GGCCCUGAAAAGGGCCAAA-biotin 3', Figure 4.1B), HIV Rev RRE binding site (5'-GGUCUGGGCGCAGCGCAAGCUGACGGUACAGGCCAAA-biotin 3'), and U1 stem-loop (5'-AAUCCAUUGCACUCCGGAUUAAA-biotin 3') were immobilized on streptavidin agarose beads. Samples were added to 400 μL of N binding buffer (10 mM HEPES pH 7.5, 0.5 mM EDTA, 100 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 0.01% (v/v) Nonidet P-40, 10% glycerol (v/v)) containing 50 μg/mL yeast tRNA (Roche) and 200 pmol of immobilized RNA. This mixture was incubated at 4°C for one hour and then washed in a filter centrifuge tube with binding buffer without tRNA. Samples were eluted with either 2X tricine sample buffer (0.1 M Tris-HCl, pH 6.8, 24% glycerol (v/v), 8% SDS (w/v), 0.2 M DTT, 0.02% Coomassie blue G-250 (w/v)) or 100 mM $MgCl_2$.

Library Construction

Library 1 was constructed beginning with the N-FLAG-myc synthetic single-stranded DNA (5'-GGGACAATTACTATTTACAATTACAATGGACGCCCAGACCNNCBNG CGCGAGCGCAGGGCCGAGAAGCAGGCCCAGTGGAAGGCCGCCAACGACTAC

AAGGACGACGATGACAAG-3', (17)) that was amplified using the Fmyc primer (5'-AGCGCAAGAGTTCTTGTCATCGTCGTCCTTGTAGTC-3'), and 42.108 primer(5'-TAATACGACTCACTATAGGGACAATTACTATTTACAATTACA-3'). Library 1 contains the sequence 5' NNCBNG 3' (N = A, T, G, or C; B = C, G, or T) at codons 6 and 7 that encodes the peptide MDAQTX$_6$X$_7$RERRAEKQAQWKA ANDYKDDDDKNSCA. The random position X$_6$ encodes the amino acids ACDFGHILNPRSTVY while the random position X$_7$ encodes AEGLPQRSVW. The single nucleotide sequence that encodes R6R7 encodes an overlapping NgoM IV restriction site. Library 2 was constructed as library 1 using a ssDNA template (5'-GGGA CAATTACTATTTACAATTACAATGGACGCCCAGACCCGCCNGCGCGAGCGCA GGGCCGAGNNSNNSGCCCAGTGGAAGGCCGCCAACGACTACAAGGACGACG ATGACAAG-3'). Library 2 contains 5' CNG 3' at codon 7 and the sequence NNS (S = G, C) at codons 14 and 15. This combination encoded L, P, Q, or R at position 7 and all 20 aa at positions 14 and 15. Arginine at position 7 encodes an NgoM IV restriction site. Round 0 pools for library 1 and 2 were sequenced (see below) and are shown in **Figure 4.S1** and Figure, respectively.

Selection Experiments

The library 1 and library 2 selection rounds were carried out essentially as described (17), except that the samples were desalted by ultrafiltration (YM-30, Millipore) and not gel purified after transcription.

Restriction Digest and Quantitation

λ N wild-type template (5'-GGGACAATTACTATTTACAATTACAATGGACGCC CAGACCCGCCGGCGCGAGCGCAGGGCCGAGAAGCAGGCCCAGTGGAAGGCC GCCAACGACTACAAGGACGACGATGACAAG-3') and aliquots of the unselected and selected samples were amplified with 7 cycles of PCR using the 42.108 primer (5' TAATACGACTCACTATAGGGACAATTACTATTTACAATTACA 3') and ^{32}P-labeled 24.108 (5' CTTGTCATCGTCGTCCTTGTAGTC 3'). The products were doubly purified using the Qiaquick PCR purification kit (Qiagen), digested with NgoMIV (NEB) for one hour, and then run on a 10% denaturing PAGE followed by quantitation by PhosphorImager (Molecular Dynamics).

Quantitation by Affinity Precipitation

Peptides for each clone were produced by ^{35}S-Met-labeled *in vitro* translation of gel purified mRNA under conditions identical to fusion synthesis (17). Ten microliters of crude lysate were mixed with 1 mL 200 nM immobilized *boxBR* in N binding buffer, washed five times with N-binding buffer and eluted with RNase A (Roche Molecular Biochemicals). The eluted peptides were run adjacent to a sample of the translation reaction on a tricine-SDS polyacrylamide gel (27), and the percentage of peptide bound was determined by PhosphorImager quantitation (Molecular Dynamics) of the respective gel bands. A representative gel is shown in Figure.

Fluorescence-Binding Measurements

Peptides were constructed by means of automated synthesis and fluorenylmethloxy-carbonyl (Fmoc) or t-butyloxycarbonyl (tBoc) monomers. Crude peptides were purified as single peaks by means of reversed-phase HPLC and the identity was confirmed by MALDI-TOF mass spectrometry. RNA hairpins containing 2-aminopurine at the second loop position (denoted 2AP7, Table 4.1; substitution for A7) were constructed by automated RNA synthesis by using 2'- O-methyl 2-aminopurine phosphoramidite (Glen Research, Sterling, VA).

Fluorescence measurements were performed essentially as in Lacourciere et al. (23), with excitation and emission wavelengths of 310 and 370 nm, respectively. Concentrated peptide was added stepwise to a stirred solution of 20 nM to 800 nM 2AP-7 RNA hairpin and the temperature was maintained at 20°C.

CD Spectroscopy

Spectra were taken on an Aviv 62 DS CD spectrometer at 25°C. The samples contained 10 μM RNA and 10 μM peptide in 10 mM potassium phosphate buffer (pH 7.9). The spectra of the bound peptides were determined by subtracting the free RNA and excess free peptide spectra from the spectra of the complex.

NMR Sample Preparation

BoxBR RNA 5'-GCCCUGAAAAAGGGC-3' (15-mer) was synthesized by *in vitro* transcription by using T7 RNA polymerase. The RNA was purified by 20% urea-PAGE, desalted on a NAP column (Amersham Pharmacia), freeze-dried, and resuspended in

NMR buffer (10 mM phosphate, pH 6, 50 mM NaCl) in H_2O/D_2O (90:10 (v/v)).
Complexes between the wild-type N (1-22) or E14R15 (1-22) and *boxBR* RNA were
generated by addition of concentrated (10 mM) peptide to *boxBR* RNA (280 μM) with
the stoichiometry monitored by inspecting the imino-proton spectra. The final sample
concentrations were 250 μm for the free RNA and 280 μM for both RNA and peptide in
the complexes.

NMR Spectroscopy

NMR spectra were collected at 25°C on a Varian INOVA 600-MHz spectrometer. A
modified double gradient echo Watergate solvent-suppression pulse sequence was used to
suppress the solvent peak (28). Assignments were based on reported work (13, 15, 16).

Sequencing

N pool template, unselected, and material selected in 100X tRNA binding buffer were
amplified with 7 cycles of PCR using the 42.108 and 24.108 primers and purified on a
4% agarose gel and extracted using the Qiaquick gel extraction kit (Qiagen). NgoMIV
digestion was performed on a portion of the selected material and the uncut material was
extracted and purified similarly. TA cloning was performed on the purified PCR
products using the TOPO cloning kit for sequencing (Invitrogen). Single colonies were
grown overnight in 5 mL of LB broth containing 100 μg/mL Ampicillin, purified using
the Qiaprep Spin Miniprep kit (Qiagen) and sequenced.

Binding to BoxB RNA

N mRNA was translated in rabbit reticulocyte lysate in the presence of of ^{35}S-methionine

(NEN) for one hour. Five μL of translation extract was taken and incubated for one hour

at 4 °C with 200 pmol of BoxB RNA, 20 μL of 50% steptavidin-agarose in 500 mL of N

Binding Buffer. The beads were washed three times with 1 mL Binding buffer and the

remaining beads counted by liquid scintillation counting.

Phosphorylation of Oligonucleotides

Reactions typically consisted of ~60 pmol of (γ^{32}-P)-ATP (NEN), 60 pmol of primer, and

1 μL (10 U) of polynucleotide kinase (NEB) in 25 μL of 1X kinase buffer (70 mM Tris-

HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT). Samples were incubated at 37 °C for 30

minutes followed by 10 minutes at 95 °C to kill the enzyme.

Supplemental Information

```
Clone              Peptide Sequence

          1         10        20        30
          •         •         •         •
wt        MDAQTRRRERRAEKQAQWKAANDYKDDDDK
Pool      -----XX----------------------

8            -GA----------------------
10        -----GR----------------------
11        -----RV----------------------
12        -----PA----------------------
13        -----D.------------S----------
15        -----PG----------------------
B         -----IA-SAGPSSRPSGRPPTTTRTTMT
D         -----AG----------------------
G3        -----GL-------GPVEGRQRLQGRR.Q
B1        V----VG----------------------
B3        -----FG--------------------N-
B6        -----RL----------------------
```

Figure 4.S1. Round 0 sequences from Selection 1 (R67X). Pool DNA was PCR amplified, T/A cloned, and sequenced (See Materials and Methods). The FLAG epitope tag is underlined. Stop codons are denoted by a red period.

```
Clone              Peptide Sequence

          1         10        20        30
          •         •         •         •
wt        MDAQTRRRERRAEKQAQWKAANDYKDDDDKNSCA
Pool      ------4------XX-------------------

1         ------R------LY-------------------
2         ------Q------QM-------------------
4         ------R------YS-------------------
5         -----P------F.-------------------
7         -----RS----AS-------------------
13        ------P------LT-------------------
14        ------P------AC-------------------
16        ------R------QS-------------------
17        -----P------.P-------------------
```

Figure 4.S2. Round 0 sequences from Selection 2. Pool DNA was PCR amplified, T/A cloned, and sequenced (See Materials and Methods). The FLAG epitope tag is underlined. Stop codons are denoted by a red period. Position 7 encodes one of four amino acids (L, P, Q or R).
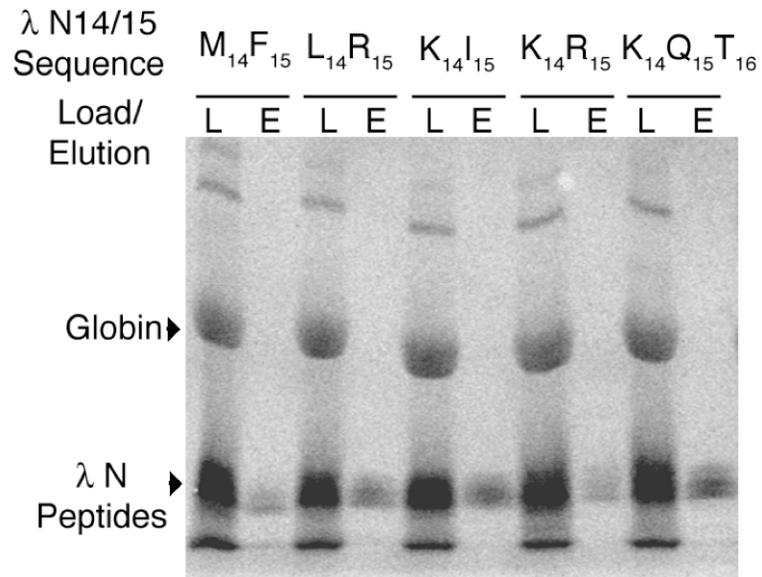
Figure 4.S3. Binding depletion of N peptides. N peptide variants were *in vitro* translated and bound to immobilized *boxB*. The beads were washed and the bound material eluted and run on a tricine gel. Crude translation extract (Load) is run on an adjacent lane to estimate the amount of peptide added to the reaction.

References

1. Franklin, N.C. Clustered arginine residues of bacteriophage lambda N protein are essential to antitermination of transcription, but their locale cannot compensate for *boxB* loop defects. (1993) *J Mol Biol* **231**, 343-360.

2. Harada, K., Martin, S.S. and Frankel, A.D. Selection of RNA-binding peptides *in vivo*. (1996) *Nature* **380**, 175-179.

3. Jain, C. and Belasco, J.G. A structural model for the HIV-1 Rev-RRE complex deduced from altered-specificity rev variants isolated by a rapid genetic strategy. (1996) *Cell* **87**, 115-125.

4. Heus, H.A. and Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. (1991) *Science* **253**, 191-194.

5. Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R. and Doudna, J.A. RNA tertiary structure mediation by adenosine platforms. (1996) *Science* **273**, 1696-1699.

6. Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. Crystal structure of a group I ribozyme domain: principles of RNA packing. (1996) *Science* **273**, 1678-1685.

7. Roberts, R.W. and Szostak, J.W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. (1997) *Proc Natl Acad Sci U S A* **94**, 12297-12302.

8. Roberts, R.W. Totally *in vitro* protein selection using mRNA-protein fusions and ribosome display. (1999) *Curr Opin Chem Biol* **3**, 268-273.

9.   Liu, R., Barrick, J.E., Szostak, J.W. and Roberts, R.W. Optimized synthesis of RNA-protein fusions for *in vitro* protein selection. (2000) *Methods Enzymol* **318**, 268-293.

10.   Cho, G., Keefe, A.D., Liu, R., Wilson, D.S. and Szostak, J.W. Constructing high complexity synthetic libraries of long ORFs using *in vitro* selection. (2000) *J Mol Biol* **297**, 309-319.

11.   Lazinski, D., Grzadzielska, E. and Das, A. Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. (1989) *Cell* **59**, 207-218.

12.   Tan, R. and Frankel, A.D. Structural variety of arginine-rich RNA-binding peptides. (1995) *Proc Natl Acad Sci U S A* **92**, 5282-5286.

13.   Su, L., Radek, J.T., Hallenga, K., Hermanto, P., Chan, G., Labeots, L.A. and Weiss, M.A. RNA recognition by a bent alpha-helix regulates transcriptional antitermination in phage lambda. (1997) *Biochemistry* **36**, 12722-12732.

14.   Cilley, C.D. and Williamson, J.R. Analysis of bacteriophage N protein and peptide binding to *boxB* RNA using polyacrylamide gel coelectrophoresis (PACE). (1997) *Rna* **3**, 57-67.

15.   Legault, P., Li, J., Mogridge, J., Kay, L.E. and Greenblatt, J. NMR structure of the bacteriophage lambda N peptide/*boxB* RNA complex: recognition of a GNRA fold by an arginine-rich motif. (1998) *Cell* **93**, 289-299.

16.   Scharpf, M., Sticht, H., Schweimer, K., Boehm, M., Hoffmann, S. and Rosch, P. Antitermination in bacteriophage lambda. The structure of the N36 peptide-*boxB* RNA complex. (2000) *Eur J Biochem* **267**, 2397-2408.

17. Barrick, J.E., Takahashi, T.T., Balakin, A. and Roberts, R.W. Selection of RNA-binding peptides using mRNA-peptide fusions. (2001) *Methods* **23**, 287-293.

18. Rebar, E.J., Greisman, H.A. and Pabo, C.O. Phage display methods for selecting zinc finger proteins with novel DNA-binding specificities. (1996) *Methods Enzymol* **267**, 129-149.

19. Boder, E.T. and Wittrup, K.D. Optimal screening of surface-displayed polypeptide libraries. (1998) *Biotechnol Prog* **14**, 55-62.

20. Greisman, H.A. and Pabo, C.O. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. (1997) *Science* **275**, 657-661.

21. Roberts, R.W. and Crothers, D.M. Specificity and stringency in DNA triplex formation. (1991) *Proc Natl Acad Sci U S A* **88**, 9397-9401.

22. Barrick, J.E., Takahashi, T.T., Ren, J., Xia, T. and Roberts, R.W. Large libraries reveal diverse solutions to an RNA recognition problem. (2001) *Proc Natl Acad Sci U S A* **98**, 12374-12378.

23. Lacourciere, K.A., Stivers, J.T. and Marino, J.P. Mechanism of neomycin and Rev peptide binding to the Rev responsive element of HIV-1 as determined by fluorescence and NMR spectroscopy. (2000) *Biochemistry* **39**, 5630-5641.

24. Richardson, J.S. and Richardson, D.C. Amino acid preferences for specific locations at the ends of alpha helices. (1988) *Science* **240**, 1648-1652.

25. Dasgupta, S. and Bell, J.A. Design of helix ends. Amino acid preferences, hydrogen bonding and electrostatic interactions. (1993) *Int J Pept Protein Res* **41**, 499-511.

26.   Dahiyat, B.I. and Mayo, S.L. Probing the role of packing specificity in protein design. (1997) *Proc Natl Acad Sci U S A* **94**, 10172-10177.

27.   Schagger, H. and Von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. (1987) *Anal Biochem* **166**, 368-379.

28.   Liu, M., Mao, X., Ye, C., Huang, H., Nicholson, J.K. and Lindon, J. Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy. (1998) *J. Magn. Res.* **132**, 125-129.

Chapter 5

Context and Conformation Dictate Function of a Transcription Antitermination Switch

This work has been adapted from the following publication:

Xia, T., Frankel, A., Takahashi, T.T., Ren, J. and Roberts, R.W. Context and conformation dictate function of a transcription antitermination switch. (2003) *Nat Struct Biol* **10**, 812-819.

Abstract

In bacteriophage λ, transcription elongation is regulated by the N protein, which binds a nascent mRNA hairpin (termed *boxB*) and enables RNA polymerase to read through distal terminators. We have examined the structure, energetics and *in vivo* function of a number of N–*boxB* complexes derived from *in vitro* protein selection. Trp18 fully stacks on the RNA loop in the wild-type structure, and can become partially or completely unstacked when the sequence context is changed three or four residues away, resulting in a recognition interface in which the best binding residues depend on the sequence context. Notably, *in vivo* antitermination activity correlates with the presence of a stacked aromatic residue at position 18, but not with N–*boxB* binding affinity. Our work demonstrates that RNA polymerase responds to subtle conformational changes in *cis*-acting regulatory complexes and that approximation of components is not sufficient to generate a fully functional transcription switch.

Introduction

The regulation of transcription elongation is important in both prokaryotic and eukaryotic gene regulation (1-5). However, much less is known about the regulation of elongation than of initiation, even in classical systems such as bacteriophage λ, in which termination (6) and elongation regulation were first described (7, 8). The N protein is a key regulator of transcription elongation in phage λ (9-11), interacting with (i) the nascent mRNA (ii) the transcription factor NusA and (iii) RNA polymerase, despite being only 107 amino acids long and totally unfolded in solution (Figure 5.1a,b) (12). N increases the processivity of RNA polymerase, allowing it to read through both intrinsic and Rho-dependent termination signals (9-11), and initiates this antitermination activity by specifically binding the nascent mRNA at the *boxB* stem loop within the sequence known as *nut* (N-utilization) (13, 14) (Figure 5.1b,c). Despite recent crystallographic work describing prokaryotic transcription machinery (15), various features of this antitermination switch remain mysterious. For example, controversy exists over whether approximation of components in the antitermination machinery is sufficient for function (16) or whether the detailed conformations of *cis*-acting elements (17) have any role.

The RNA-binding domain of the λ N(1–22) peptide folds into a bent α-helix upon binding the *boxB* RNA and enforces the formation of a GNRA tetraloop fold with one base extruded (Figure 5.1c) (17-21). In the complex, Trp18 stacks on top of the tetraloop fold, effectively extending the RNA stack by one residue (Figure 5.1c). Previous work *in vitro* and *in vivo* indicates that Trp18 is important for both RNA binding and processive antitermination (19, 22). We have previously used the RNA-binding domain of the λ

N(1–22)–*boxB* complex as a model system to isolate high-affinity RNA-binding peptides by mRNA display (23, 24). We were surprised when selections randomizing residues 13–22 of N showed no conservation of tryptophan at position 18 in the selected peptides (23). Instead, a loose consensus containing Glu14 and Arg15 emerged, which also differed from the energetically important wild-type residues Lys14 and Gln15 (19).
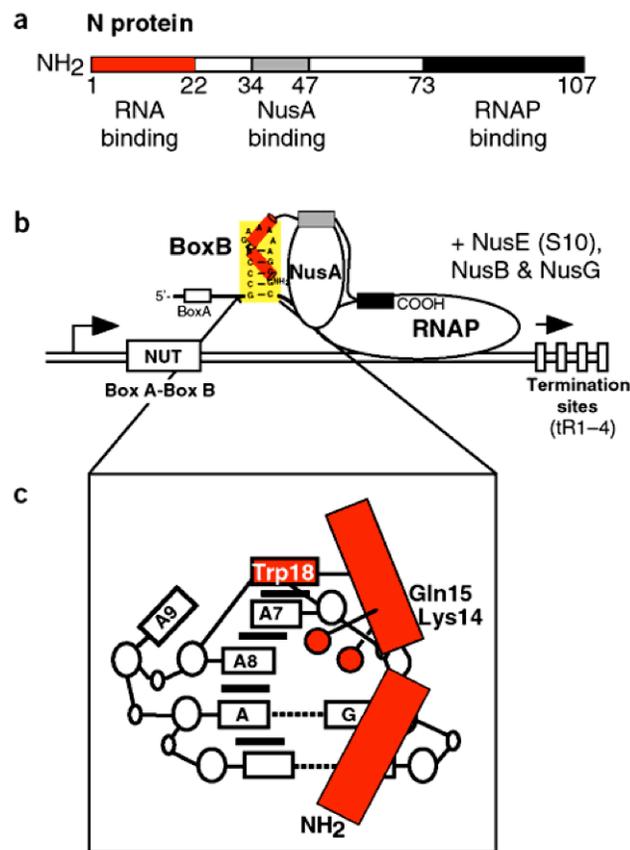
Figure 5.1. Structure and function of the N protein. **(a)** Domain structure of N. Regions for *boxB* RNA binding (residues 1–22), NusA binding (residues 34–47) and RNA polymerase binding (residues 73–107) are indicated (39). **(b)** Formation of the processive antitermination complex. The minimal antitermination complex (50) forms when N binds the *boxB* in the nascent mRNA and is stabilized by NusA binding. Association of the other Nus proteins results in a processive antitermination complex. **(c)** Schematic of N peptide–*boxB* RNA complex indicating the peptide and loop stacking geometry.

Notably, selections in which residues 7, 14 and 15 were also randomized resulted in a sequence containing the Glu14Arg15 pair in the presence of Trp18. Here we address the structural, energetic and functional consequences of these mutations.

Results

Context-Dependent Energetics

Our previous selections hinted that certain residues in the RNA-binding domain of N might be energetically coupled (23). Subsequent sequence-covariation analysis of more than 60 different RNA-binding peptides isolated through *in vitro* selection experiments supported this idea, indicating a statistically significant linkage between (i) positions 14 and 15 and (ii) positions 15 and 18, and general linkage between (iii) adjacent residues and (iv) those positioned one helical turn away (25).

We were curious to examine the structural and functional consequences of these changes and assess why these selected sequences are not found in the wild-type phage. We constructed 20 peptides with systematic variations at positions 14, 15 and 18 and measured the binding of each one to *boxB* RNA to investigate any energetic coupling experimentally. In wild-type λ N(1–22; K14Q15), the residues at position 18 with highest binding affinities were tryptophan and tyrosine, whereas arginine, alanine and glutamate variants showed similar stability (Figure 5.2). In other sequence contexts, both the best residues at position 18 and the rank order shifted substantially. For example, peptides containing Trp18, Ala18 and Arg18 all showed similar affinity when residues 14 and 15 were glutamate and arginine (that is, in the E14R15 variant), respectively. However, the peptide with Arg18 was the best binder in a designed sequence containing
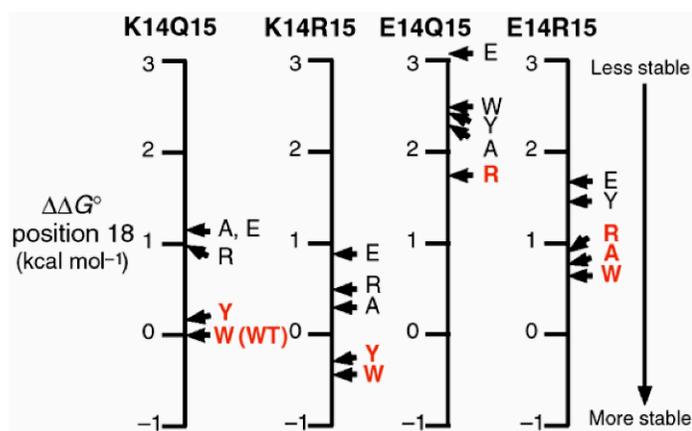
Figure 5.2. Relative stability of N peptide–*boxB* RNA complexes that differ at positions 14, 15 and 18. The dissociation constant (K_d) for the wild-type (WT) complex was determined to be 1.2 nM for 2AP-7–labeled *boxB* RNA. Uncertainties in K_d values were <10%. For complexes with small fluorescence change at 2AP-7 (such as E14Q15), 2AP-8 and 2AP-9 *boxB* RNA were used. Complexes are ranked by their free energy change relative to the wild-type N–*boxB* complex (G°; kcal mol^{-1}), with the most stable shown in red. Complexes are indicated by the peptide residue identity at positions 14 and 15 (top) and residue 18 (adjacent to arrows).

E14Q15. Thus, the energetic landscape at position 18 depends on the sequence context one helical turn away at positions 14 and 15.

Probing Sequence Preferences with mRNA Display

We next investigated the context-dependent energetics of RNA binding using *in vitro* selection experiments to rank the amino acids at position 18 in relation to different sequence contexts at positions 14 and 15. We inserted a saturation cassette (NNG/C codon) into the position corresponding to residue 18 of the protein. We then carried out mRNA display–based peptide selections (23) for five libraries with variants differing at positions 14 and 15: (i) K14Q15 (ii) L14Q15 (iii) K14H15 (iv) K14R15 and (v) E14R15. After two rounds of selection against *boxB* RNA (23, 24), 20 sequences were cloned from each library and assessed (Table 5.1). We observed good correlation between the number of times a peptide sequence occurred and its binding affinity for the *boxB* RNA,

Table 5.1. Amino acid composition at position 18 after two rounds of *in vitro* **selection for binding** *boxB* **RNA**

Library	Trp	Tyr	Phe	His	Total aromatic	Total other residues (%)
K14Q15-X18	19.2	19.2	7.7	3.8	50	50
L14Q15-X18	30.0	35.0	0	5.0	70	30
K14H15-X18	0	17.9	17.9	3.6	39.4	61.6
K14R15-X18	4.2	4.2	4.2	12.4	25	75
E14R15-X18	0	0	5.3	15.8	21.1	78.9

For each library, fractions (%) of sequences with tryptophan, tyrosine, phenylalanine and histidine at position 18 are listed separately, and their sum is listed as Total aromatic; the rest of the residues are listed as Total others. The numbers of sequences for each library are as follows: 26 for the K14Q15-X18 library, 20 for L14Q15-X18, 28 for K14H15-X18, 24 for K14R15-X18 and 19 for E14R15-X18.

confirming our approach. The selected sequences showed strong context effects for *boxB* RNA binding and peptide positions 14, 15 and 18 that were unique to each library. For the K14Q15-X18 and L14Q15-X18 libraries, we observed a strong preference for tryptophan and tyrosine at position 18. Fluorescence measurements of these variants showed quenching of 2-aminopurine (2AP)-7-substituted *boxB* similar to that of the wild-type peptide, consistent with -stacking of the aromatic side chain on the RNA loop (19, 22) (Figure 5.3b). Our selection results correlated well with previous genetic work demonstrating that the L14Q15 variant is highly functional and that tryptophan and tyrosine are the only functional substitutions at position 18 of N protein (22). In contrast, our selections showed that aromatic and planar residues rarely occur at position 18 in the K14R15 and E14R15 variants. The selections therefore provided an experimental demonstration that the fitness landscape for *boxB* binding is rugged for the interfacial residues, and were suggestive of structural differences that would explain the observed sequence preferences.
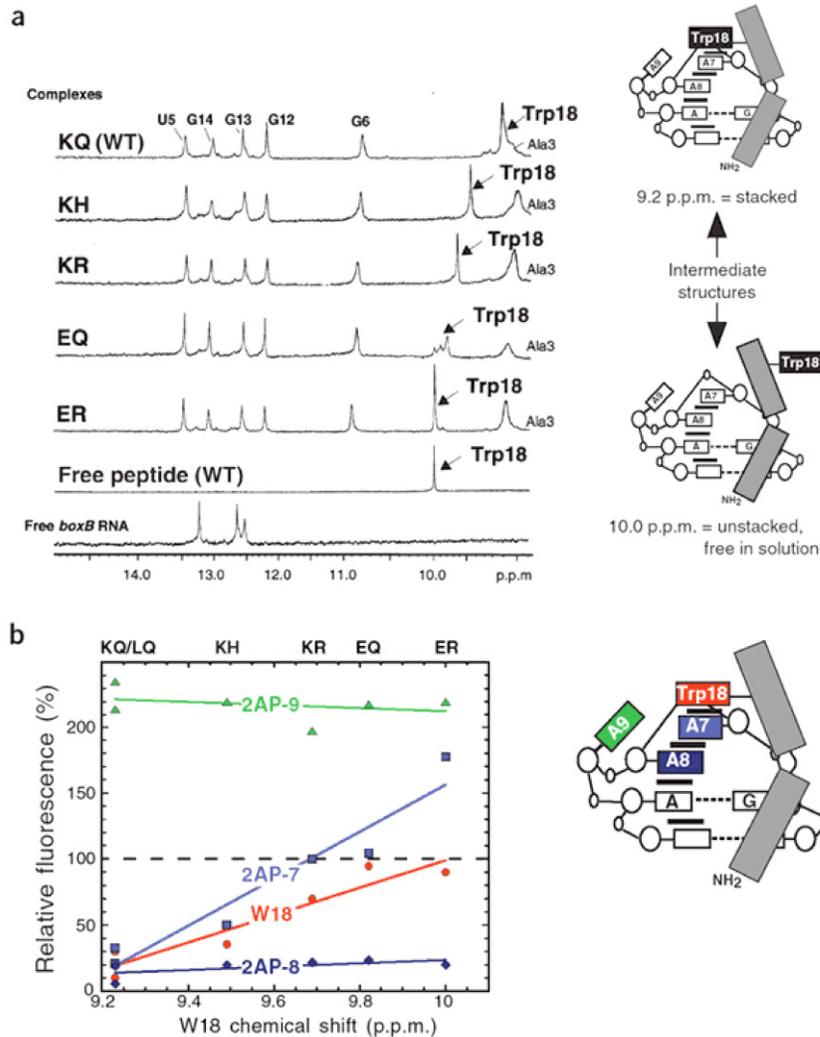
Figure 5.3. Structure and folding characteristics of N–*boxB* complexes as detected by NMR and fluorescence spectroscopy. **(a)** One-dimensional NMR spectra of peptide–RNA complexes containing Trp18 and residues K14Q14 (KQ; wild-type, WT), K14H15 (KH), K14R15 (KR), E14Q15 (EQ) or E14R15 (ER) at positions 14 and 15. The positions of the imino protons and the tryptophan indole proton are indicated. One-dimensional spectra are also shown for the free peptide (wild-type) and free *boxB* RNA. **(b)** Correlation of N–*boxB* fluorescence with Trp18 chemical shift. The relative fluorescence of Trp18 (unlabeled complex) or 2-aminopurine at RNA position 7, 8 or 9 (2AP-7, 2AP-8 and 2AP-9) is plotted versus the 1H chemical shift of Trp18 (p.p.m.). The identities of peptide residues 14 and 15 are indicated at the top of the figure. Linear fits to the data and the stacking schematic of the wild-type complex are also shown.

An Adaptive RNA-Peptide Interface

NMR provided a means to examine the structural origins of the context dependences we had observed. The indole NH proton of Trp18 and the RNA imino protons provide well-resolved spectroscopic handles in our RNA–peptide complexes. We reasoned that structural variations in the complexes should be revealed by intensity or chemical-shift changes of some of these resonances. All of the free peptides examined were unfolded in solution (23) and showed induced α-helix formation upon binding the *boxB* RNA (Figure 5.S1). The free *boxB* RNA showed three imino resonances; the rest were lost to exchange with solvent (Figure 5.3a). In the wild-type N(1–22)–*boxB* complex (K14Q15, Figure 5.3a), the RNA showed two more resonances from imino protons (five total), four corresponding to stem base pairs (U5, G14, G13 and G12) and one upfield at 11 p.p.m. for the G6–A10 sheared pair. The most notable change in the peptide spectra occurred at Trp18, which shifted upfield by 0.8 p.p.m. as a result of the ring-current effect of the adjacent, π-stacked RNA base (A7), consistent with previous structural work (17-19).

All complexes gave the same number and pattern of imino proton peaks at 10–14 p.p.m. (including the G6–A10 sheared pair imino), indicating that the mutant peptides induce similar RNA stem structure in all cases. Consistent with this idea, the N-terminal peptide (N(1–11)) present in all our mutants also induced the same RNA spectra with *boxB* (data not shown), indicating that stem recognition occurs predominantly with the N-terminal 11 residues. However, the Trp18 indole NH resonance in five of the RNA–peptide complexes showed considerable variability (Figure 5.3a). Whereas both the wild-type complex and L14Q15 complex (data not shown) showed almost complete

stacking on the loop as indicated by the large upfield Trp18 chemical shift, in the E14R15

complex, Trp18 appeared almost completely unstacked as indicated by the same chemical

shift as in the free peptide. The other complexes (K14H15, K14R15, and E14Q15)

showed intermediate levels of stacking, as gauged by their Trp18 chemical shifts between

the wild-type and E14R15.

The NMR observations correlated well with fluorescence experiments used to

determine the K_d values for RNA–peptide complex formation. In these experiments, we

measured the fluorescence ratio of the complex to free peptide or free 2AP-labeled RNA.

Quenching indicates increased stacking of the fluorophore, whereas enhancement

indicates increased exposure to aqueous solvent (26). The stacked structure seen in the

wild-type complex showed quenching of both position 7 (2AP-7) of the RNA and Trp18

of the peptide, whereas fluorescence enhancement for 2AP-7 was seen for the unstacked

E14R15 variant. The relative fluorescence of 2AP-7 and Trp18 showed a linear

correlation, as did 2AP-7 fluorescence and the Trp18 chemical shift we observed by

NMR (Figure 5.3b). RNA positions adjacent to this interface (2AP-8 and 2AP-9) showed

a consistent amount of fluorescence quenching (22–28%) or enhancement (200–235%)

irrespective of the Trp18 chemical shift for different complexes (Figure 5.3b).

Taken together, the NMR, CD and fluorescence data indicate that the complexes

have similarity in the RNA stem, the sheared GA pair, the flipped base (position A9) in

the loop and the induced helical structure of the peptides. However, the critical Trp18–

A7 interaction showed structural and energetic variability depending on the sequence

context at positions 14 and 15, one helical turn away. Structurally, variable amounts of

stacking between Trp18 and A7 were observed, with the wild-type and L14Q15

complexes being fully stacked and the E14R15 complex completely unstacked. It is likely that different complexes have different population distributions of stacked and unstacked conformations in dynamic exchange (27). In terms of the energetics and selection, the wild-type and L14Q15 sequence contexts showed that Trp18 and Tyr18 are the clear winners, whereas the best residues in other sequence backgrounds varied somewhat, and in the E14R15 context, Arg18, Ala18 and Trp18 appeared energetically similar.

Stacking, Not Energetics, Dictates *In Vivo* Function

Previous work from a number of groups has supported the idea that binding alone between N and *boxB* is sufficient to construct a functional antitermination switch. *In vitro* and *in vivo* studies have indicated that arginines 6, 7, 8, 10 and 11, as well as Trp18 or Tyr18, are critical for *in vivo* function as a result of their essential role in binding *boxB* RNA, whereas other positions such as Thr5, Glu9 and Glu13 can be readily substituted *in vitro* and *in vivo (19, 22)*. Previous *in vitro* selection work from our laboratory supported the idea that any substitutions for Arg6 and Arg7 disrupted stringent *boxB* RNA binding (23). Additionally, NMR and alanine scanning mutagenesis has also indicated that positions 3, 14 and 19 are important in recognition by phage N protein (17-19, 28). Finally, one group of researchers had reported an *in vivo* screen based solely on approximation between a nascent transcript and N mutants containing arginine-rich RNA-binding domains at their amino terminus (29). However, this system alone generated only 2% as much antitermination as the wild-type N–*boxB* system, suggesting that the distinct conformation of the components has an important role in generating a fully functional antitermination switch. Furthermore, some RNA mutations that support

N binding do not support the formation of higher-order complexes with the transcription elongation-termination factor NusA (13, 17). The identities of RNA bases in the pentaloop of *boxB* affect N binding (19) and antitermination (30) in somewhat different ways. G6 and A10 are absolutely required to maintain the GA sheared base pair, and purine bases are highly preferred for positions 7–9, with most pyrimidine substitutions resulting in <10% of wild-type antitermination activities except the A9U mutation (the flipped-out base), which has 35% of wild-type activity (30). Work in our laboratory had previously shown that wild-type N peptide discriminated between *boxB* RNA and a GAAA tetraloop with approximately ten-fold specificity (23).

These previous studies therefore motivated us to examine the *in vivo* activity of the N mutants we had studied. Our selected proteins are one-, two- and three-residue N mutants that preserve the known essential sequence elements that had been studied previously. Our measurements here of these mutants showed a wide range of stabilities (3.5 kcal mol^{-1}) and marked structural differences at the Trp18–A7 peptide-RNA interface. We were interested to see if differences in the conformation or stability of these complexes would affect *in vivo* antitermination activity. We inserted these sequences into a β-galactosidase reporter construct and measured transcription antitermination efficiency (22) in solution (Figure 5.4a) and on agar plates (Figure 5.4b,c). The activities of the complexes showed a full dynamic range, with some having activity equal to or better than that of the wild-type complex, whereas others show no activity, comparable to strains lacking N altogether (Figure 5.4a). In the wild-type sequence context (K14Q15), tryptophan and other aromatic residues were necessary at position 18 for efficient antitermination, consistent with previous *in vivo* work (22).
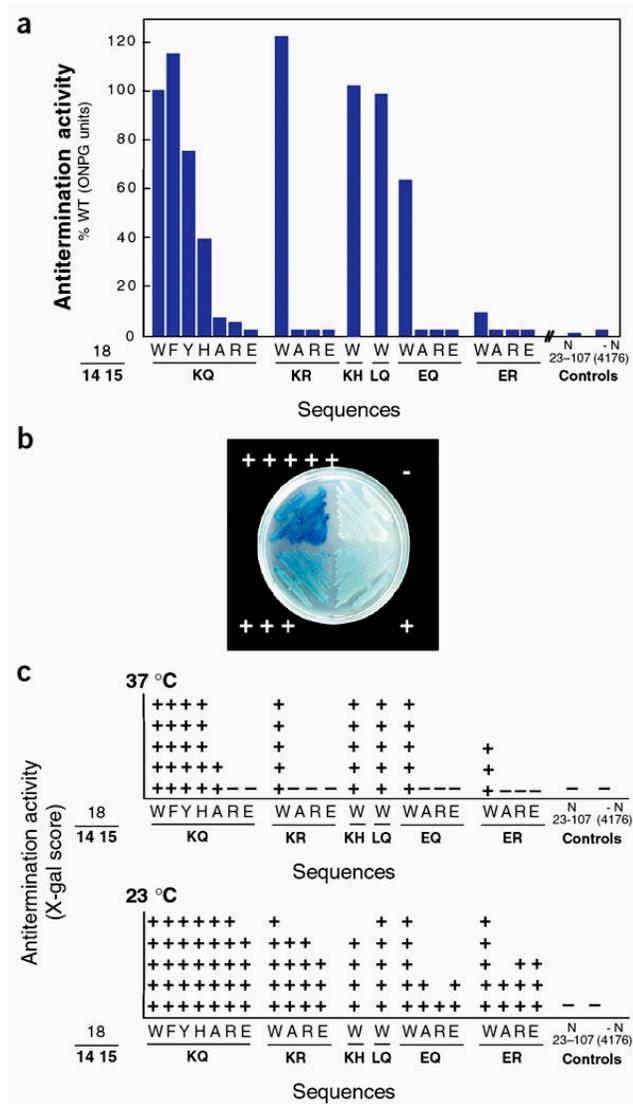
Figure 5.4. Antitermination activities of N mutants, as determined using *E. coli* strains carrying a two-plasmid reporter system (22). Functional antitermination through four terminators (tR1–4) results in expression of β-galactosidase. **(a)** Antitermination activity in solution, determined by quantifying β-galactosidase using a colorimetric assay based on *o*-nitrophenyl-β-*D*-galactopyranoside. β-galactosidase activity is expressed as a percentage of that of the wild-type (WT) N reporter construct (51). The residue identities at positions 14, 15 and 18 are indicated. Control constructs in which N lacks an RNA-binding domain (N(23–107)) or the N expressor plasmid is absent (N 4176) (22) are indicated. **(b)** Plate-scoring assay of mutants quantified on the basis of X-gal signal. **(c)** Plate-scoring assay of temperature sensitivity in the antitermination activity of N mutants grown at 37 °C or 23 °C.

However, assessment of the unstacked E14R15 mutant indicated that Trp18 alone is not sufficient for efficient antitermination, resulting in only about 9% the activity of the wild-type. This loss of activity did not seem to be caused by lack of expression, as both the E14R15 mutant and wild-type N proteins were expressed *in vivo* as judged by northwestern analysis (Figure 5.S2). Thus, variations at positions 14 and 15 that tend to unstack Trp18 resulted in little or no antitermination activity. Finally, constructs lacking the RNA-binding domain of N(23–107) showed no activity in this assay, indicating that the signal does not result from *boxB*-independent antitermination by N (31).

Antitermination activity tested on agar plates expanded the lower dynamic range of the assay and revealed that a number of our variants have temperature-sensitive phenotypes (Figure 5.4c). All variants that were positive at 37 °C also showed full activity at 23 °C. However, many variants that were inactive at 37 °C showed some level of antitermination activity at 23 °C. In particular, the fact that the E14R15 mutant has activity at lower temperature suggests that it binds *boxB* specifically *in vivo* and that its lack of functionality at higher temperature is not due to nonspecific RNA binding. Previous work has shown that some Nus proteins are dispensable for growth at lower temperatures (30 °C) (32), consistent with our observations that many more of the complexes are functional at the permissive temperature.

Our data demonstrated that the structure of the N–*boxB* complex interface, rather than its stability, is the principal determinant of functional antitermination (Figure 5.5). Our structural data suggested that the wild-type and mutant peptides have structural variation at the Trp18–A7 interface, while having essentially the same stem-helix interaction and overall loop folding (A9 extruded). The relative stability of the

complexes showed little correlation with antitermination activity (Figure 5.5a). Six of the complexes we examined showed stability within 1 kcal/mol of the wild-type complex but <10% antitermination activity, whereas one (E14Q15-W18), which binds more than 2.5 kcal mol^{-1} more weakly than the wild-type complex, had 63% of wild-type activity *in vivo* (Figure 5.5a). By contrast, complexes containing a stacked Trp18–A7 interface were fully active for antitermination (Figure 5.5b). Complexes in which Trp18 is unstacked (E14R15, Figure 5.5b). seemed to hinder or block processive antitermination. These data led us to conclude that there are different levels of control on the full dynamic range of transcription antitermination by N. The structure of the Trp18–A7 interface is critical in forming a fully functional antitermination switch, in addition to the effect of simple approximation of components (29).
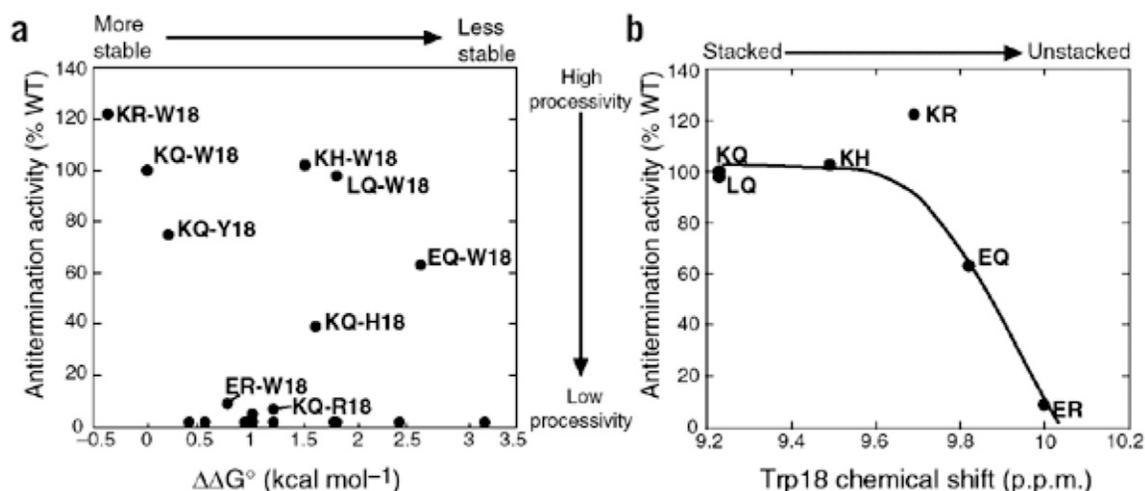
Figure 5.5. Energetic and structural correlation of antitermination activity. **(a)** No correlation is observed between the stability of the N–*boxB* complexes and their antitermination activity in solution. **(b)** Antitermination activity does correlate with Trp18-A7 stacking as detected by the chemical shift of Trp18 or the fluorescence of 2AP-7 (not shown).

Role of NusA

Our conclusion raised the questions of how the N-*boxB* conformation is sensed by the elongating polymerase, and how a single stacking interaction can dictate the processivity of a 400-kDa molecular machine. NusA is the most likely suspect as a conformational sensor. NusA has an important role in regulating the processivity of RNA polymerase (33-36). Previous work has indicated that NusA can discriminate between N–*boxB* complexes differing in a single nucleotide (13, 17) and can bind to N with high affinity at a site distinct from the RNA-binding domain (12). We therefore examined the binding and fluorescence effects of NusA on various 2AP-labeled N–*boxB* complexes (Figure 5.6).

NusA showed no detectable interaction with complexes constructed from an N peptide (either wild-type or E14R15 N(1–22)) and a 2AP-labeled *boxB* (data not shown), indicating that the N-*boxB* interface alone may not be sufficient to recruit NusA binding with high affinity. This observation is consistent with previous data showing that N protein is needed to bind NusA in order for NusA to bind RNA (37). NusA bound to complexes containing full-length N protein (wild-type or E14R15 N(1–107)) (Figure 5.6a and Figure 5.S3). In the wild-type N(1–107)–*boxB* complex, NusA quenched the fluorescence of the extruded base (2AP-9), but modestly enhanced the emission from 2AP at the crucial Trp18-A7 interface. In the E14R15 N(1–107)–*boxB* complex, NusA quenched both the extruded base (2AP-9) and the base located at the top of the loop (2AP-7). Thus, NusA seems to drive the binary N protein–*boxB* complexes toward similar, but distinct, final conformational states. Furthermore, NusA binds with similar, high-affinity K_d values to both wild-type and E14R15 complexes (14 ± 6 versus 9 ± 2
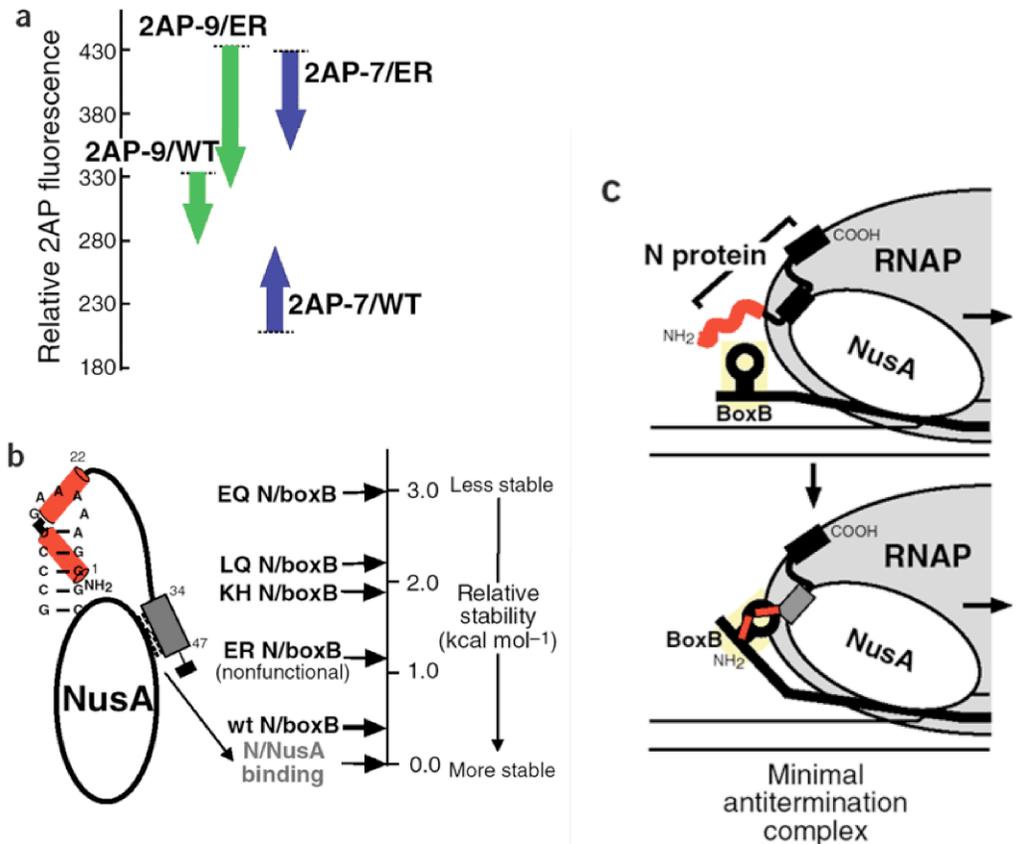
Figure 5.6. Characteristics of N-*boxB*-NusA interactions. **(a)** Fluorescence changes induced by NusA binding to N–*boxB* complexes. Binary complexes were constructed using wild-type N(1–107) or E14R15 N(1–107) and 2AP-7 (blue) or 2AP-9 (green) labeled *boxB* RNA. The arrows indicate the 2AP fluorescence change (starting to final) induced when NusA is added. **(b)** Relative energetics of NusA, N protein and *boxB* interactions. The N-NusA interaction is stronger than N binding to the *boxB* RNA. The wild-type (WT), K14H15, L14Q15 and E14Q15 complexes show >50% wild-type antitermination function in vivo, whereas ER shows <10% activity. **(c)** Model for formation of the minimal antitermination complex in vivo. The tight binding of N to NusA implies that N is prebound to polymerase during elongation and scans the message as it emerges from the transcription elongation complex. As the *boxB* hairpin emerges from the transcription bubble, the N-terminal domain of N engages the message to generate the minimal antitermination complex.

nM, respectively; Figure 5.S4), suggesting that NusA affinity is dictated by protein-protein contacts between N protein and NusA, rather than the N-*boxB* interface, in line with previous observations (12).

Notably, the K_d for NusA and N protein determined by van Gilst and von Hippel under more stringent temperature and salt conditions (12) actually indicates stronger binding between NusA and N protein (70 nM) than that of N protein to *boxB* RNA (K_d 127 nM under the same conditions; Figure 5.S5) (Figure 5.6b). An energetic ranking of a number of functional and nonfunctional N-*boxB* interactions indicates that the N-NusA interaction is as much as 3 kcal mol^{-1} more favorable than the RNA-protein interactions needed to construct a functional antitermination switch *in vivo* (Figure 5.6b).

Discussion

A Novel Model for Antitermination

In vivo, the N–*boxB* RNA complex is believed to serve the bipartite function of tethering NusA to the *nut* site RNA and sequestering N to interact directly with RNA polymerase (38, 39). The relative energetics of N-NusA and N-*boxB* interactions lead us to propose a new model for assembly of the minimal antitermination complex (Figure 5.6c). The very high affinity between N and NusA implies that N can be bound to the elongating polymerase, via interactions with NusA, where NusA associates with elongating polymerase with similar affinity (30–100 nM) (33, 34). This complex can be assembled stably before the exit of the *nut* site from polymerase. This view differs from the model generally described for antitermination, whereby N is recruited to the

transcription elongation complex (TEC) via interactions with *boxB* (Figure 5.1b) (13, 38, 40, 41). Previous structural modeling of TECs supports the idea that polymerase constantly scans the nascent transcript for termination signals via the flap-tip helix and the action of bound NusA (42). Our work implies that N protein modifies NusA, altering the function of RNA polymerase through NusA and RNAP interactions when it properly engages an appropriate RNA structure. This view of N action is consistent with experiments showing that N supports efficient transcription antitermination even in the absence of *boxB (16, 36) and with modeling efforts focused on understanding transcription termination (42). We conclude that conformational sensing of the Trp18-A7 interface occurs in the context of a TEC, containing N, NusA and RNA polymerase itself, after N has engaged the nascent boxB* hairpin. Essentially, the unfolded N acts as a template onto which the components of the antitermination complex assemble. This action of N is apparently dependent on the exact conformation of the interface of Trp18-A7, where the presence or absence of the stacking directs NusA and polymerase into either different orientations and spatial arrangements or different physical contacts. Supporting this view is previous work indicating that the quaternary NusA–N–*nut* polymerase complex can be formed and is stable with respect to competition from exogenous components (36). Although these components represent a minimal antitermination complex, assembly of the complete complex (containing N, NusA, NusB, NusE, NusG, *boxA*, *boxB* and RNAP) is essential for processive antitermination under stringent conditions *in vivo (14).*

Implications

Our results obtained under stringent conditions suggest that proper assembly of the RNA-protein interface is indispensable for full antitermination activity *in vivo*. One consequence of our observations is the conclusion that genetic screens using antitermination to isolate RNA-binding peptides (29) will likely be influenced by constraints of functional antitermination.

The N-*boxB* interface provides a clear example of adaptive recognition in which the sequence context dictates the structural meaning of interfacial residues. *In vivo*, the various conformations we observed had profound effects on polymerase function and demonstrated that approximation of components is not sufficient to generate a functional transcriptional switch. Rather, the transcription elongation complex has the means to sense the conformation of *cis*-acting regulatory elements, much like a stockbroker reading a ticker tape, profoundly altering its processivity accordingly. Similarly, it seems likely that a more complicated eukaryotic elongation complex could distinguish between various *cis*-acting regulatory signals such as those provided by HIV Tat (43, 44) and Nrd1p (45-47). In addition to revealing the structural origin of this control process, our work indicates that ligands targeting elements of elongation complexes (48) offer potential tools to better understand and control the mechanisms of gene expression in both prokaryotic and eukaryotic systems.

Materials and Methods

In vitro Peptide Selection

Generation and purification of mRNA-peptide fusions and selection cycles were done as previously described (23). Approximately 5 pmol of purified ^{35}S-labeled fusions were incubated with 200 pmol of biotinylated *boxB* RNA (5'-GGCCCUGAAAAGGG CCAAA-biotin-3') in the selection step in the presence of yeast tRNA competitor (Roche).

Synthesis of RNAs and Peptides

The 2-aminopurine (2AP)-labeled *boxB* RNAs, 5'-GGCCCUGAAAAAGGGCC-3' (with underlined nucleotides individually labeled 2AP-7, 2AP-8, 2AP-9), were generated by solid-phase synthesis using standard phosphoramidite chemistry, and then deprotected and purified by 20% urea-PAGE. RNA without the 2AP label for NMR experiments (5'-GCCCUGAAAAAGGGC-3') was synthesized using T7 RNA polymerase and a synthetic DNA template (49). Peptides MDAQTRRRERRAEXXAQXKAAN(gy) (where X represents variable residues and gy represents extra residues for quantification purposes for peptides lacking Trp18) were made by automated Fmoc synthesis (using an ABI 432 peptide synthesizer), deprotected, and purified by reversed-phase HPLC.

Binding Studies

K_d values for N-peptide binding to *boxB* RNA were determined by monitoring the fluorescence change of RNAs labeled with 2AP at the seventh, eighth or ninth position

upon addition of concentrated peptide and were fitted as previously described (23). NusA binding was performed by titrating NusA solution into preformed N–*boxB* complex. Free energies were calculated on the basis of complex stability at 20 °C in standard binding buffer (50 mM potassium acetate, 20 mM Tris, pH 7.5). Individual K_d values were reproducible with an error of 0.2 kcal mol^{-1}.

NMR Spectroscopy

Spectra were collected on a Varian Inova 600 at 25 °C in a buffer of 50 mM NaCl, 10 mM sodium phosphate, 0.5 mM EDTA, pH 6. Complex formation was monitored by inspecting imino protons of the RNA. The concentrations of the complexes were 150–300 μM.

Circular Dichroism

CD spectra were collected on an Aviv 62 DS CD spectrofluorimeter in 10 mM potassium phosphate buffer at 20 °C. Spectra for the peptide portion in the complex were calculated by subtracting spectra for free RNA and excess peptide from that of the complex. Helical content was calculated as described previously (48).

Antitermination Assay

Strains expressing wild-type and mutant N protein were constructed using the two plasmid reporter system previously described (22). All sequences were verified by sequencing. N-mutant strains were plated on tryptone agar supplemented with 0.05 mM IPTG, 0.08 mg ml^{-1} X-gal and the appropriate antibiotics. Plates were incubated at either

23 °C (permissive temperature) or 37 °C (nonpermissive temperature) and scored for blue color. Temperature sensitivity was defined here as the loss of relative β-galactosidase units in comparing the blue color of colonies grown at permissive and nonpermissive temperatures. All assays were performed in triplicate. Intensities of blue color are shown for sample N–mutant strains K14Q15E18 (-), K14Q15A18 (+), Nun-N fusion (+++) and K14Q15W18 (+++++).

Protein Expression and Purification

Wild-type and E14R15 N protein were expressed from plasmid pET-N1 in BL21 (DE3) gold cells as described (12), except that a SP-cation exchange column (Amersham) was used for the final purification step and the protein was dialyzed into fluorescence binding buffer (50 mM KCl, 20 mM Tris-HCl, pH 7.5, 1 mM β-mercaptoethanol). The NusA gene was amplified from plasmid pMS7, subcloned into vector pET19b (Novagen) and verified by sequencing. BL21(DE3) cells harboring the resulting plasmid were induced at an OD_{600} of 0.4 with 1 mM IPTG, grown for 3 h, harvested by centrifugation and resuspended 20 ml buffer A (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM β-mercaptoethanol, 1 mM imidazole) per liter of culture. The cells were lysed by three passes through a French press and clarified by two centrifugations at 18,000x*g* for 30 min at 4 °C. The supernatant was loaded on a Superflow Ni-NTA column (Qiagen), washed with 5 column volumes of buffer A, with 5 column volumes of buffer A containing 500 mM NaCl, and eluted with 400 mM imidazole in buffer A. The eluted protein was dialyzed against fluorescence binding buffer.

N Protein Expression Level

Expression of wild-type N protein, E14R15 and N(23–107) variants was accomplished as above except that the cells were induced with 0.5 mM IPTG at an OD_{600} of 0.3 and grown for 3 h, and 0.1 OD units of culture was concentrated and resuspended in 100 μl of 6x SDS loading buffer (0.28 M Tris-Cl/SDS, pH 6.8, 30% (v/v) glycerol, 1% (w/v) SDS, 0.5 M DTT). Five microliters of this solution was separated by 15% SDS-PAGE and then transferred to nitrocellulose (Bio-Rad) in transfer buffer (10 mM NaHCO3-3 mM Na2CO3, pH 9.9, 20% (v/v) methanol, 0.05% (w/v) SDS) at 4 °C. The nitrocellulose was blocked at room temperature in wash buffer (10 mM Tris-HCl, pH 7.5,100 mM NaCl, 1 mM EDTA, pH 8.0, 0.01% (v/v) NP-40) containing 5% (w/v) nonfat dry milk and 50 μg per milliliter of yeast tRNA, and then 100 nM biotinylated *boxB* RNA (5'-GGCCCUGAAAAGGGCCAAA-biotin-3') was added. The nitrocellulose was washed three times with wash buffer, probed with 0.5 g ml^{-1} streptavidin–horseradish peroxidase (Amersham) in wash buffer containing 0.5% (w/v) BSA, washed four times and visualized using Supersignal West Pico substrate (Pierce).

Acknowledgments

Competing Interests Statement:

The authors declare that they have no competing financial interests.

Supplemental Information
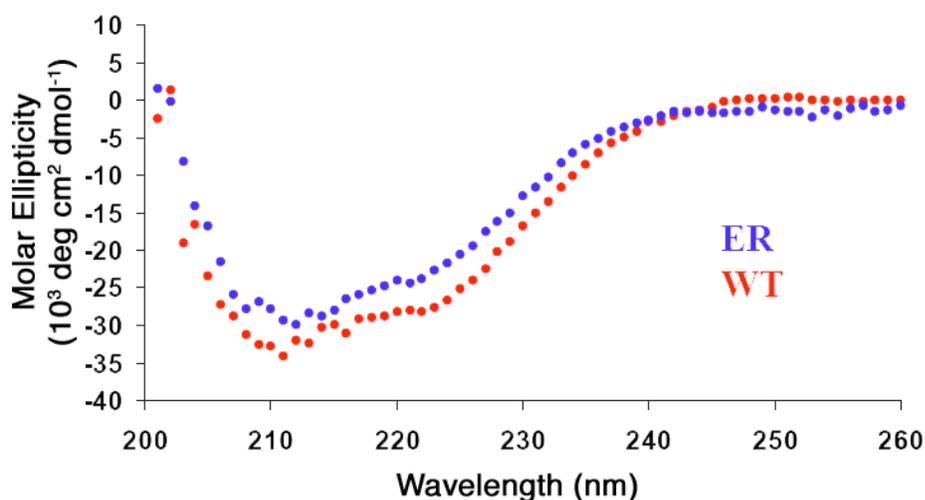
Figure 5.S1. CD spectra of the wild-type and ER peptides in complex with *boxB* RNA. Spectra were collected in 10 mM phosphate buffer at 20 $^{\circ}$C. The contributions of the peptide portion in the complex were calculated by subtracting the RNA signal from the complex signal as described previously (48). The average number of amino acid residues in helical form is 16.4 and 13.8 for wild-tye and ER, respectively.
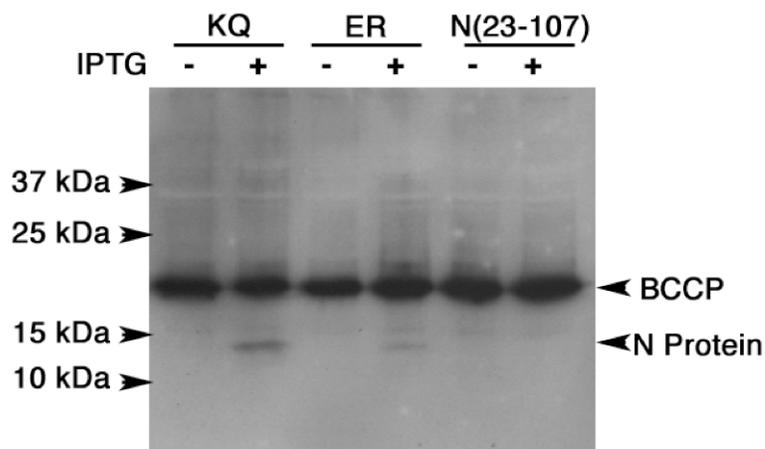
Figure 5.S2. Expression levels of full-length wild-type N protein and ER protein. Cell lysates containing wild-type N protein (KQ) or ER protein (ER) (+ IPTG lanes) were run on an SDS-PAGE gel and transferred to nitrocellulose. Wild-type and ER proteins were probed with biotinylated *boxB* RNA and visualized with streptavidin-HRP. The N(23-107) protein lacking the N-terminal RNA binding domain and lanes where cells were uninduced (- IPTG lanes) show no RNA binding. An endogenous biotinylated *E. Coli* protein (possibly BCCP) also reacts with streptavidin-HRP.
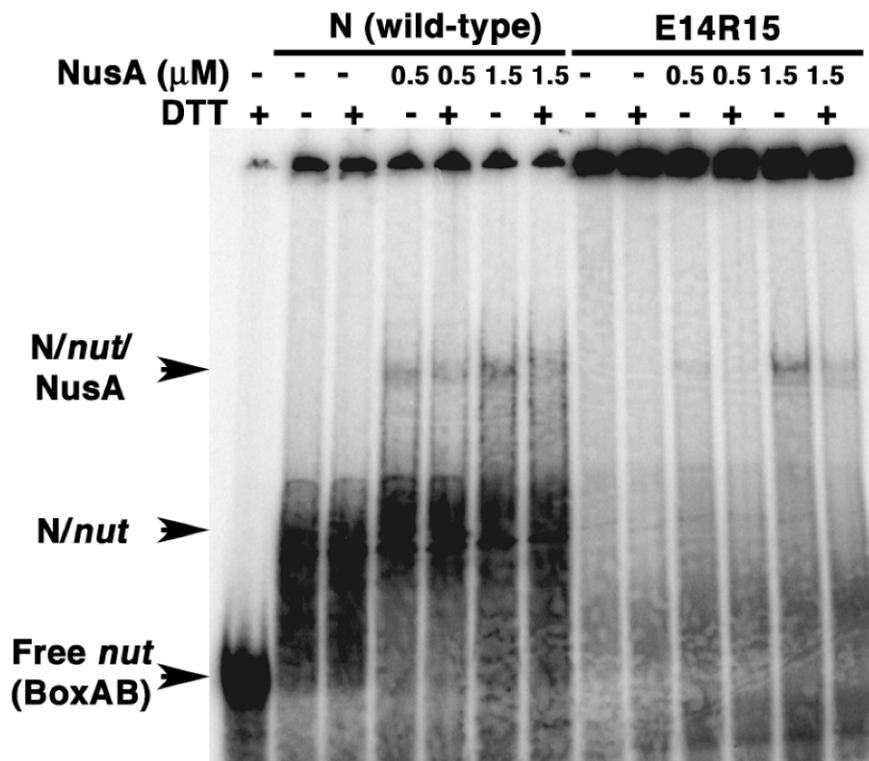
Figure 5.S3. Interaction of NusA with full-length wild-type N protein or ER protein. ^{32}P-labeled nut RNA (*boxA/boxB*) was incubated with either wild-type N protein (N (wt)) or E14R15 protein (ER) with increasing concentrations of NusA. Samples were run in the presence (+) or absence (-) of DTT. N and ER bind the *nut* RNA giving a band of slower mobility which can be supershifted with the addition of NusA. NusA supershifts both wild-type/*nut* and ER/*nut* complexes.
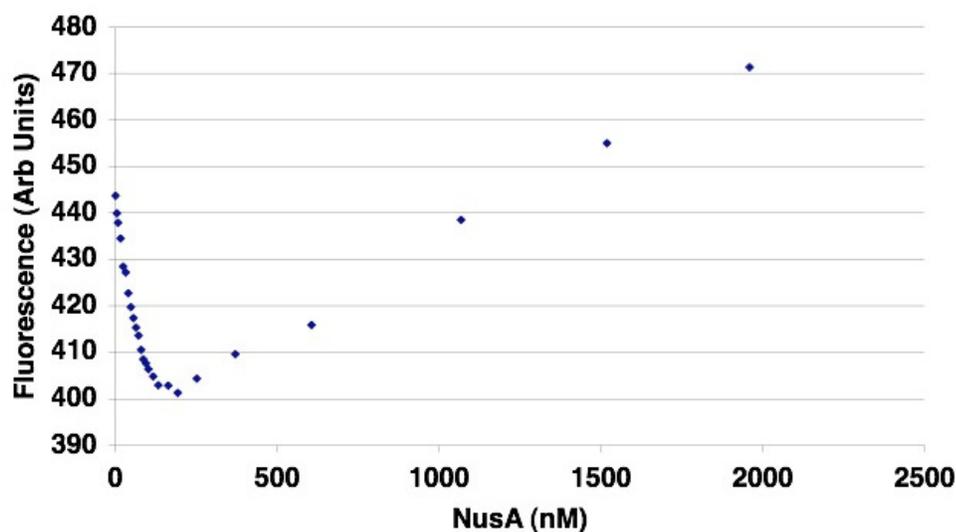
Figure 5.S4. A representative fluorescence titration isotherm of NusA binding to wild-type N in complex with *boxB* labeled at position 9. N protein and *boxB* complexes were preformed at 100 nM (1:1) in 50 mM KOAc, 20 mM Tris-OAc, pH 7.5 buffer at 20 $^{\circ}$C. Measured dissociation constants are: 14 ± 6 nM (2AP-7), and 3 ± 1 nM (2AP-9) for NusA binding to the wildtype N-*boxB* complex; 9 ± 2 nM (2AP-7), and 35 ± 5 nM (2AP-9) for NusA binding to the ER-*boxB* complex.
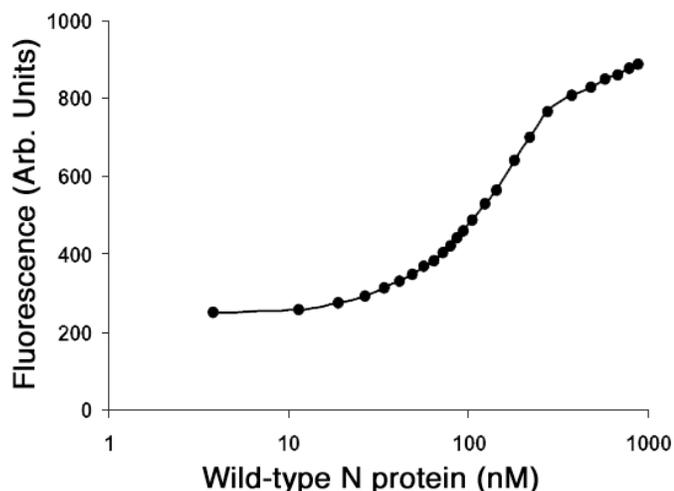
Figure 5.S5. Fluorescence titration isotherms of wild-type N protein to *boxB* RNA labeled at position 9 with 2-aminopurine (2AP-9). *BoxB* concentration is 100 nM. The titration is performed in 150 mM KOAc, 40 mM Tris-OAc, 1 mM DTT, 0.1 mM EDTA, pH 7.5 buffer at 30 oC. Measured dissociation constant is 127 nM.

References

1.   Von Hippel, P.H. An integrated model of the transcription complex in elongation, termination, and editing. (1998) *Science* **281**, 660-665.

2.   Yarnell, W.S. and Roberts, J.W. Mechanism of intrinsic transcription termination and antitermination. (1999) *Science* **284**, 611-615.

3.   Uptain, S.M., Kane, C.M. and Chamberlin, M.J. Basic mechanisms of transcript elongation and its regulation. (1997) *Annu Rev Biochem* **66**, 117-172.

4.   Zorio, D.A. and Bentley, D.L. Transcription elongation: the 'Foggy' is lifting... (2001) *Curr Biol* **11**, R144-146.

5.   Maniatis, T. and Reed, R. An extensive network of coupling among gene expression machines. (2002) *Nature* **416**, 499-506.

6.   Roberts, J.W. Termination factor for RNA synthesis. (1969) *Nature* **224**, 1168-1174.

7.   Greenblatt, J. Positive control of endolysin synthesis *in vitro* by the gene N protein of phage lambda. (1972) *Proc Natl Acad Sci U S A* **69**, 3606-3610.

8.   Salstrom, J.S. and Szybalski, W. Coliphage lambdanutL-: a unique class of mutants defective in the site of gene N product utilization for antitermination of leftward transcription. (1978) *J Mol Biol* **124**, 195-221.

9.   Greenblatt, J., Nodwell, J.R. and Mason, S.W. Transcriptional antitermination. (1993) *Nature* **364**, 401-406.

10.   Friedman, D.I. and Court, D.L. Transcription antitermination: the lambda paradigm updated. (1995) *Mol Microbiol* **18**, 191-200.

11. Weisberg, R.A. and Gottesman, M.E. Processive antitermination. (1999) *J Bacteriol* **181**, 359-367.

12. Van Gilst, M.R. and Von Hippel, P.H. Assembly of the N-dependent antitermination complex of phage lambda: NusA and RNA bind independently to different unfolded domains of the N protein. (1997) *J Mol Biol* **274**, 160-173.

13. Mogridge, J., Mah, T.F. and Greenblatt, J. A protein-RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the lambda N protein. (1995) *Genes Dev* **9**, 2831-2845.

14. Zhou, Y., Mah, T.F., Yu, Y.T., Mogridge, J., Olson, E.R., Greenblatt, J. and Friedman, D.I. Interactions of an Arg-rich region of transcription elongation protein NusA with NUT RNA: implications for the order of assembly of the lambda N antitermination complex in vivo. (2001) *J Mol Biol* **310**, 33-49.

15. Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O. and Darst, S.A. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. (2002) *Science* **296**, 1285-1290.

16. Gusarov, I. and Nudler, E. Control of intrinsic transcription termination by N and NusA: the basic mechanisms. (2001) *Cell* **107**, 437-449.

17. Legault, P., Li, J., Mogridge, J., Kay, L.E. and Greenblatt, J. NMR structure of the bacteriophage lambda N peptide/*boxB* RNA complex: recognition of a GNRA fold by an arginine-rich motif. (1998) *Cell* **93**, 289-299.

18. Scharpf, M., Sticht, H., Schweimer, K., Boehm, M., Hoffmann, S. and Rosch, P. Antitermination in bacteriophage lambda. The structure of the N36 peptide-*boxB* RNA complex. (2000) *Eur J Biochem* **267**, 2397-2408.

19. Su, L., Radek, J.T., Hallenga, K., Hermanto, P., Chan, G., Labeots, L.A. and Weiss, M.A. RNA recognition by a bent alpha-helix regulates transcriptional antitermination in phage lambda. (1997) *Biochemistry* **36**, 12722-12732.

20. Jucker, F.M., Heus, H.A., Yip, P.F., Moors, E.H. and Pardi, A. A network of heterogeneous hydrogen bonds in GNRA tetraloops. (1996) *J Mol Biol* **264**, 968-980.

21. Heus, H.A. and Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. (1991) *Science* **253**, 191-194.

22. Franklin, N.C. Clustered arginine residues of bacteriophage lambda N protein are essential to antitermination of transcription, but their locale cannot compensate for boxB loop defects. (1993) *J Mol Biol* **231**, 343-360.

23. Barrick, J.E., Takahashi, T.T., Ren, J., Xia, T. and Roberts, R.W. Large libraries reveal diverse solutions to an RNA recognition problem. (2001) *Proc Natl Acad Sci U S A* **98**, 12374-12378.

24. Barrick, J.E., Takahashi, T.T., Balakin, A. and Roberts, R.W. Selection of RNA-binding peptides using mRNA-peptide fusions. (2001) *Methods* **23**, 287-293.

25. Barrick, J.E. and Roberts, R.W. Sequence analysis of an artificial family of RNA-binding peptides. (2002) *Protein Sci* **11**, 2688-2696.

26.   Lacourciere, K.A., Stivers, J.T. and Marino, J.P. Mechanism of neomycin and Rev peptide binding to the Rev responsive element of HIV-1 as determined by fluorescence and NMR spectroscopy. (2000) *Biochemistry* **39**, 5630-5641.

27.   Xia, T., Becker, H.C., Wan, C., Frankel, A., Roberts, R.W. and Zewail, A.H. The RNA-protein complex: direct probing of the interfacial recognition dynamics and its correlation with biological functions. (2003) *Proc Natl Acad Sci U S A* **100**, 8119-8123.

28.   Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A. and Patel, D.J. Solution structure of P22 transcriptional antitermination N peptide-*boxB* RNA complex. (1998) *Nat Struct Biol* **5**, 203-212.

29.   Harada, K., Martin, S.S. and Frankel, A.D. Selection of RNA-binding peptides *in vivo*. (1996) *Nature* **380**, 175-179.

30.   Doelling, J.H. and Franklin, N.C. Effects of all single base substitutions in the loop of *boxB* on antitermination of transcription by bacteriophage lambda's N protein. (1989) *Nucleic Acids Res* **17**, 5565-5577.

31.   Das, A., Pal, M., Mena, J.G., Whalen, W., Wolska, K., Crossley, R., Rees, W., Von Hippel, P.H., Costantino, N., Court, D., Mazzulla, M., Altieri, A.S., Byrd, R.A., Chattopadhyay, S., Devito, J. and Ghosh, B. Components of multiprotein-RNA complex that controls transcription elongation in *Escherichia coli* phage lambda. (1996) *Methods Enzymol* **274**, 374-402.

32.   Friedman, D.I. (1998) The Bacteriophages (Plenum, New York.

33.   Whalen, W., Ghosh, B. and Das, A. NusA protein is necessary and sufficient *in vitro* for phage lambda N gene product to suppress a rho-independent terminator placed downstream of nutL. (1988) *Proc Natl Acad Sci U S A* **85**, 2494-2498.

34.   Gill, S.C., Weitzel, S.E. and Von Hippel, P.H. *Escherichia coli* sigma 70 and NusA proteins. I. Binding interactions with core RNA polymerase in solution and within the transcription complex. (1991) *J Mol Biol* **220**, 307-324.

35.   Mason, S.W., Li, J. and Greenblatt, J. Host factor requirements for processive antitermination of transcription and suppression of pausing by the N protein of bacteriophage lambda. (1992) *J Biol Chem* **267**, 19418-19426.

36.   Rees, W.A., Weitzel, S.E., Yager, T.D., Das, A. and Von Hippel, P.H. Bacteriophage lambda N protein alone can induce transcription antitermination in vitro. (1996) *Proc Natl Acad Sci U S A* **93**, 342-346.

37.   Mah, T.F., Kuznedelov, K., Mushegian, A., Severinov, K. and Greenblatt, J. The alpha subunit of *E. coli* RNA polymerase activates RNA binding by NusA. (2000) *Genes Dev* **14**, 2664-2675.

38.   Chattopadhyay, S., Garcia-Mena, J., Devito, J., Wolska, K. and Das, A. Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage lambda. (1995) *Proc Natl Acad Sci U S A* **92**, 4061-4065.

39.   Mogridge, J., Legault, P., Li, J., Van Oene, M.D., Kay, L.E. and Greenblatt, J. Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. (1998) *Mol Cell* **1**, 265-275.

40. Barik, S., Ghosh, B., Whalen, W., Lazinski, D. and Das, A. An antitermination protein engages the elongating transcription apparatus at a promoter-proximal recognition site. (1987) *Cell* **50**, 885-899.

41. Mason, S.W. and Greenblatt, J. Assembly of transcription elongation complexes containing the N protein of phage lambda and the *Escherichia coli* elongation factors NusA, NusB, NusG, and S10. (1991) *Genes Dev* **5**, 1504-1512.

42. Toulokhonov, I., Artsimovitch, I. and Landick, R. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. (2001) *Science* **292**, 730-733.

43. Marciniak, R.A., Calnan, B.J., Frankel, A.D. and Sharp, P.A. HIV-1 Tat protein trans-activates transcription *in vitro*. (1990) *Cell* **63**, 791-802.

44. Richter, S., Ping, Y.H. and Rana, T.M. TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication. (2002) *Proc Natl Acad Sci U S A* **99**, 7928-7933.

45. Steinmetz, E.J. and Brow, D.A. Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. (1996) *Mol Cell Biol* **16**, 6993-7003.

46. Steinmetz, E.J. and Brow, D.A. Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. (1998) *Proc Natl Acad Sci U S A* **95**, 6993-7003.

47. Steinmetz, E.J., Conrad, N.K., Brow, D.A. and Corden, J.L.   RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. (2001) *Nature* **413**, 327-331.

48. Austin, R.J., Xia, T., Ren, J., Takahashi, T.T. and Roberts, R.W.   Designed arginine-rich RNA-binding peptides with picomolar affinity. (2002) *J Am Chem Soc* **124**, 10966-10967.

49. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. (1987) *Nucleic Acids Res* **15**, 8783-8798.

50. Rees, W.A., Weitzel, S.E., Das, A. and Von Hippel, P.H.   Regulation of the elongation-termination decision at intrinsic terminators by antitermination protein N of phage lambda. (1997) *J Mol Biol* **273**, 797-813.

51. Peled-Zehavi, H., Smith, C.A., Harada, K. and Frankel, A.D.   Screening RNA-binding libraries by transcriptional antitermination in bacteria. (2000) *Methods Enzymol* **318**, 297-308.

Chapter 6

In Vitro Selection of Peptides Targeting Human Telomerase RNA

Abstract

Telomerase is an attractive anti-cancer target since it is expressed in most tumor cells but not in most normal cells. It is a ribonucleoprotein, composed of telomerase RNA and telomerase reverse transcriptase. We have targeted two functionally vital domains of telomerase RNA, P6.1 and CR7, using mRNA display. From peptide libraries of 10 or 100 trillion members, we have isolated L- and D-peptides that bind with low nanomolar affinity and are highly specific for their cognate RNA target. These peptides are highly basic and are predicted to form α-helicies, however non-electrostatic interactions contribute to overall peptide binding. Our data argue that very few random peptides bind RNA with high affinity and specificity. We expect that the peptides we have selected should bind in the context of telomerase RNA and will inhibit telomerase activity.

Introduction

Telomerase is a reverse transcriptase that is responsible for maintaining the telomeres at the ends of eukaryotic chromosomes (1). It is minimally composed of telomerase RNA (TR) and a protein component, telomerase reverse transcriptase (TERT), but other accessory proteins are known to associate with the telomerase complex (2). The enzyme utilizes an RNA template provided by TR to add d(TTAGGG) repeats to the ends of telomeres, preventing the progressive telomere shortening associated with DNA replication. Telomere shortening leads to cellular senescence and growth arrest, but cell crisis and death result if cells continue to divide after senescence should have been reached.

Roughly 85-90% of all human cancer cells constituatively express telomerase (3). Expression of telomerase is one of three events required for the oncogenic transformation of normal human cells (4). Most normal cells show little expression of telomerase, with the exception of stem and germ-line cells. Telomere lengths of normal cells are generally longer than those of tumor cells (5). Telomerase, therefore, represents an attractive target for anti-tumor specific therapy.

Several strategies currently exist for inhibiting telomerase. One class of compounds inhibits telomerase by stabilizing G-quartet structures, preventing telomerase from acting on the G-rich telomeres (6, 7). Other groups have used reverse transcriptase inhibitors to target hTERT (8). Lastly, several compounds that basepair with human TR (hTR) have been used to target the template region (9-11) and prevent assembly of the human TERT (hTERT)/hTR complex (12).

No group has targeted specific structural elements of hTR, possibly because of the difficulty in designing ligands to structural, rather than sequence elements. Targeting structural elements has the potential to be more specific than targeting sequence-specific elements. It is also more difficult for drug resistance to develop against structural-based compounds since there are few mutations that would prevent ligand binding but retain interactions with natural binding partners.

We have previously used mRNA display to isolate novel, high affinity, RNA binding peptides (13-15). mRNA display enables the selection of peptides by covalently coupling the peptides to their encoding mRNA (Figure 6.1), allowing the peptide sequence to be recovered by amplification of the RNA (16). Libraries containing up to 100 trillion independent sequences (10^{14}) can be constructed by mRNA display, the
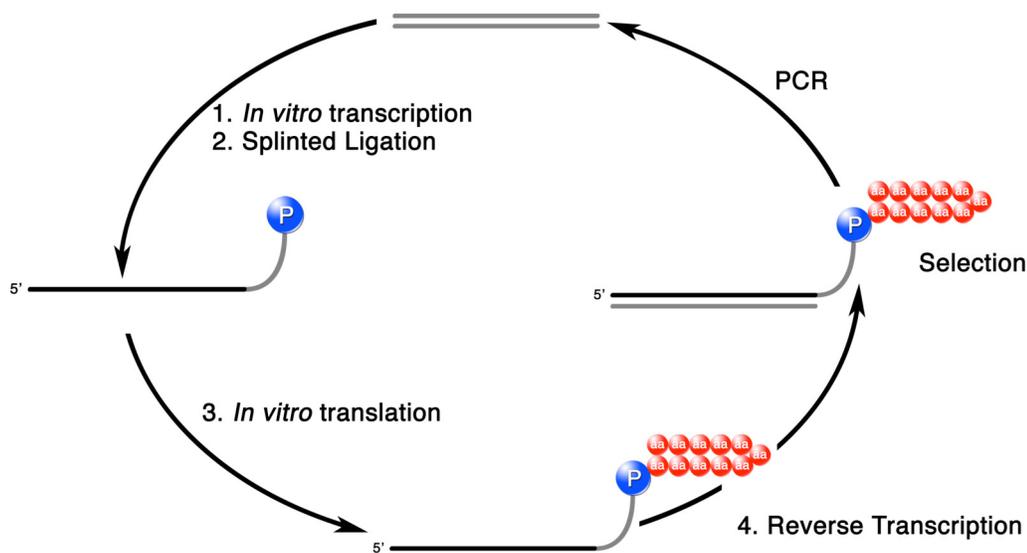
Figure 6.1. An mRNA display selection cycle. A double stranded DNA library is transcribed using T7 RNA polymerase [1] , and the resulting mRNA ligated to a synthetic oligonucleotide containing puromycin [2]. *In vitro* translation of the ligated product results in attachment of a peptide to its encoding mRNA [3]. Reverse transcription generates a cDNA/mRNA hybrid [4], which is used in affinity selection. PCR generates an enriched pool which is used in further cycles of selection.

largest of any peptide library. We sought to extend our original methodology and develop methods for the selection of RNA binding peptides targeting RNAs for which no natural ligand is known. We chose to target telomerase RNA since hTR-binding peptides would be useful tools to study telomerase as well as potential drug candidates.

Our overall design strategy to isolate novel RNA binding ligands is outlined in (Figure 6.2). Based on the RNA secondary structure of telomerase, we identified regions important for activity by their high sequence conservation or by loss of function when mutagenized, and constructed RNA targets based on these domains. The RNA targets were structurally
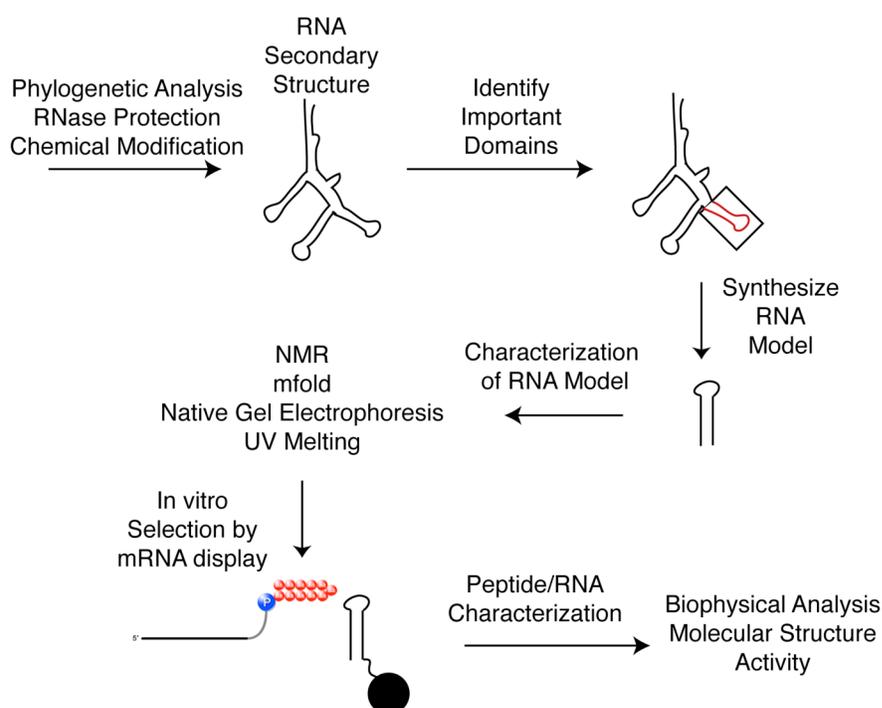
Figure 6.2. Strategy for selection of novel RNA binding peptides. Using a phylogenetic or experimentally derived RNA secondary structure, domains important for RNA activity are identified (shown in red). Small RNA models based on important domains are synthesized and characterized using NMR and native gel electrophoresis to insure that the RNA is folded and monomeric. A biotinylated target is synthesized and used as a target in an mRNA display selection. Peptide sequences from the selection are synthesized, tested for activity, as well as characterized biophysically and structurally.

characterized using NMR and native gel electrophoresis to confirm the RNA is folded and monomeric. We then synthesized biotinylated RNAs (both natural and unnatural, see below) and used them as targets in separate mRNA display selections. Peptides resulting from the selection were characterized using a variety of biophysical and structural techniques and will be tested for activity both *in vivo* and *in vitro*. This selection strategy should be applicable to isolate peptides that bind any RNA target.

Based on our previous experience, we expect that at least some of the peptides we select will have an affect on telomerase activity (15). RNA binding peptides often contain arginine rich domains (17) and such highly basic sequences can cross cellular membranes (18, 19). However, peptides are usually degraded by biological systems, reducing their overall effectiveness. We aim to select peptides that have better stability in biological systems using a reflection selection strategy (20, 21) (Figure 6.3). Selection
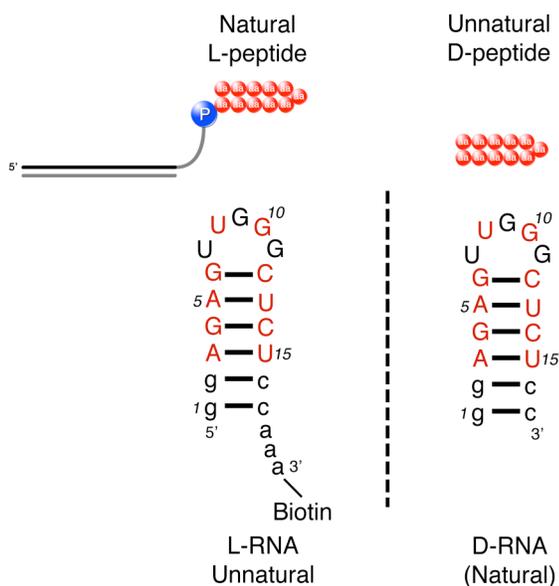
Figure 6.3. Reflection selection strategy. An unnatural (opposite chirality) L-RNA is synthesized and used as a target in an mRNA display selection using natural L-amino acids. Winning sequences are synthesized as D-peptides, which bind to natural D-RNA and are resistant to protease degradation.

is performed against an unnatural RNA target composed of L-nucleotides instead of D-nucleotides resulting in L-peptides that bind L-RNA. D-peptides of the same sequences should bind the natural D-RNA, resulting in peptides that are resistant to degradation by proteases (22).

Results and Discussion

RNA Target Determination

The ideal RNA target would have the following features: (i) Be important for telomerase function so that peptides that bind it would inhibit telomerase (ii) Be small enough to be synthesized by chemical methods so that unnatural bases could be introduced (e.g., 2-aminopurine, a fluorescent base used for biophysical studies) and since smaller RNAs are more amenable to structure determination by NMR (iii) Be stable and well folded (iv) Be monomeric.

Although we could theoretically target the full-length telomerase RNA, we were concerned that hTR would not be an ideal target for a number of reasons. First, full-length hTR may not fold correctly in the absence of hTERT, resulting in peptides that would bind non-physiologically relevant structures (23). A large RNA also has more potential for interaction with the nucleic acid component of mRNA-protein fusions resulting in selection for RNA-RNA or RNA-DNA interactions, instead of peptide-RNA interactions. Lastly, once peptides were isolated after selection, determining where individual peptides bind would be complicated since there are presumably many sites on hTR that could serve as targets for our peptides.

The phylogenetically determined secondary structure of vertebrate telomerase RNA (Figure 6.) contains eight conserved regions (CR1-8) (24, 25). CR1 corresponds to the template region and has been targeted by a number of antisense compounds (26). The pseudoknot CR2/CR3 domain and the CR4/CR5 domain are both important for
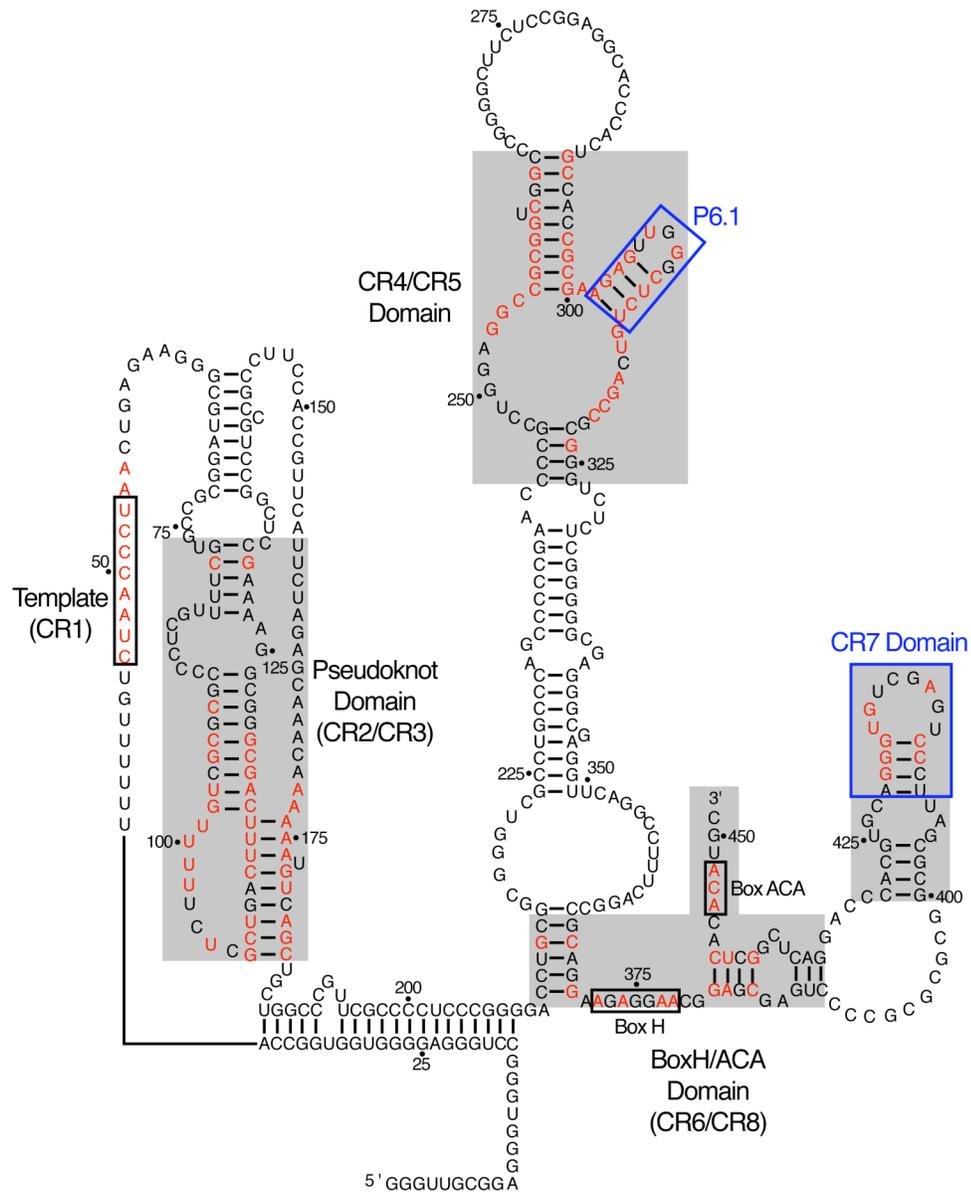
Figure 6.4. Proposed secondary structure of human telomerase RNA (hTR). Adapted from (24). Four highly conserved domains of hTR are shaded in grey. Regions targeted for selection are the P6.1 hairpin of the CR4/CR5 domain and the CR7 domain (blue boxes). Nucleotides conserved among vertebrate telomerase RNAs (>90% conservation) are red.

catalytic activity. The CR6/CR8 (boxH/ACA) and CR7 domains have been implicated in RNA processing and localization (27).

We focused our initial efforts on the CR4/CR5 domain and the CR7 domain, which are important for telomerase function. Construction of model RNA targets based on these domains was straightforward since both domains contain RNA hairpins, the targets of other selections using mRNA display (13, 14). The CR1 domain was also a potential target, however we feared it would be difficult for small peptides to recognize this single-stranded domain.

The CR4/CR5 domain encompasses almost 60 nt, with many highly conserved nucleotides located in the CR5 region. Part of the CR5 region folds into a small (13 nt), universally conserved hairpin, designated P6.1 (25). P6.1 is critical for telomerase activity; disrupting the stem or mutating the highly conserved U307 or G309 abolishes function in mouse telomerase. However, it is likely that these nucleotides would also be important in human telomerase because of their strict conservation. The P6.1 loop adopts a backbone conformation similar to UNCG tetraloops (28), with one base (G309) extruded (Figure 6.). In the P6.1 NMR structure (29), U306 and G310 form a GU wobble pair and the bases in the loop are highly exposed to solvent, suggesting that they interact with hTERT or other regions of hTR (29, 30).
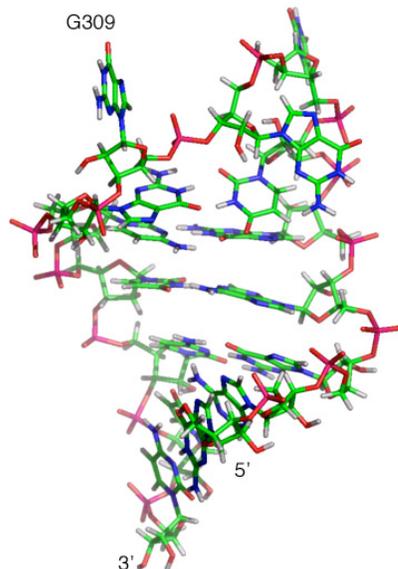
Figure 6.5. NMR structure of the P6.1 hairpin. The P6.1 pentaloop adopts a UNCG-like tetraloop fold, extruding G309 from the structure (29). The 5' and 3' ends of the hairpin are indicated.

The CR7 domain is a 30 nt domain that forms a hairpin containing an internal bulge. Many highly conserved bases are located in the CR7 terminal hairpin loop (U407-A422) and mutations or deletions in this region prevent hTR accumulation *in vivo* (31). The CR7 loop contains a self-complementary sequence (5'-AGCU-3') that could potentially form kissing complexes (32). The ability of hTR to dimerize may play a role *in vivo* as functional telomerase complexes are dimeric (33).

We designed and synthesized two RNA hairpins based on the CR4/CR5 P6.1 hairpin (A302-C314) and the terminal CR7 hairpin (U407-A422)(Figure 6.). One and two additional GC pairs were added to the CR7 and P6.1 stems, respectively, to stabilize the hairpins and allow the RNA to be synthesized by *in vitro* transcription (34).

U G G¹⁰
U   G
G — C
₅A — U
G — C
A — U₁₅
g — c
¹g — c
a   3'
a
⁵'a
Biotin

A G C U
G   G
U   U
₅C — G
C — G¹⁵
C — G
U — A
¹g — c
a   3'
a
⁵'a
Biotin

P6.1   CR7
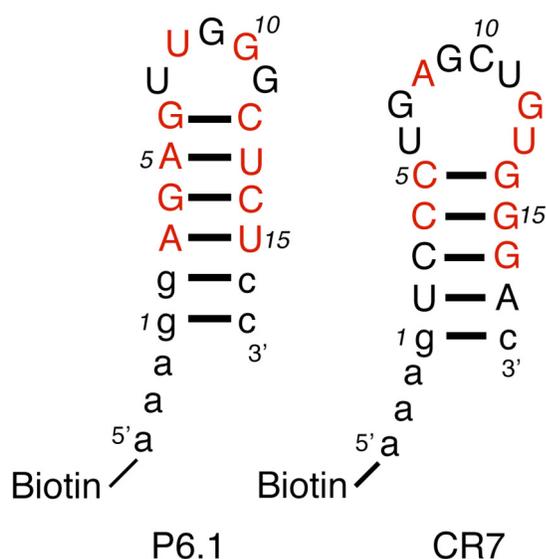
Figure 6.6. Sequence and secondary structure of telomerase RNA models. Highly conserved positions are shaded red. RNA models used for native gel electrophoresis and NMR lack the 5' biotin and rArArA spacer. Nucleotides added for stability and as spacers are depicted in lower case.

Characterization of Designed Sequences

Native gel electrophoresis of the RNA hairpins showed significant dimerization of the RNAs (Figure 6.). However, annealing by rapid heating and cooling of the sample showed that more than 90% of the RNA could be shifted to the monomeric form (Figure 6.). P6.1 remained monomeric for more than 4 hours while CR7 quickly dimerized, although the dimer form generally composed less than 10% of the sample (data not shown).

Imino protons in 1D ^{1}H-NMR spectra provide useful information about RNA secondary structure. Generally, a base-paired imino proton is protected from exchange with solvent resulting in a peak in the NMR spectrum. NMR of the annealed RNA hairpins showed reasonably sharp, well defined peaks in the imino region implying that
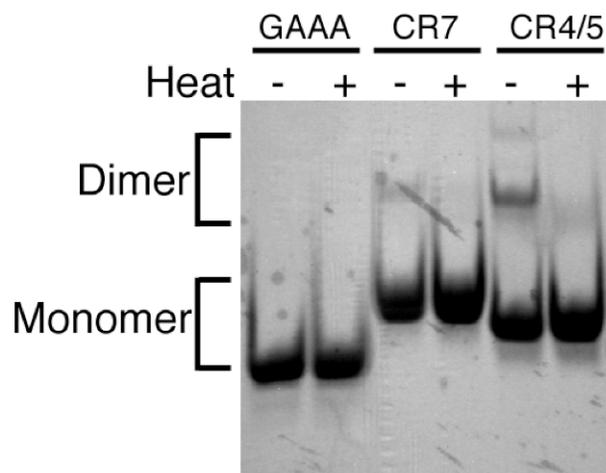
Figure 6.7. Native gel analysis of telomerase RNA targets. RNAs were annealed (+ heat lanes) by heating to 90 °C and rapidly cooling on ice. A control hairpin derived from *boxB* RNA containing the sequence 5'-GAAA-3' in the loop is shown as a monomeric control. Monomeric and dimeric species are indicated.
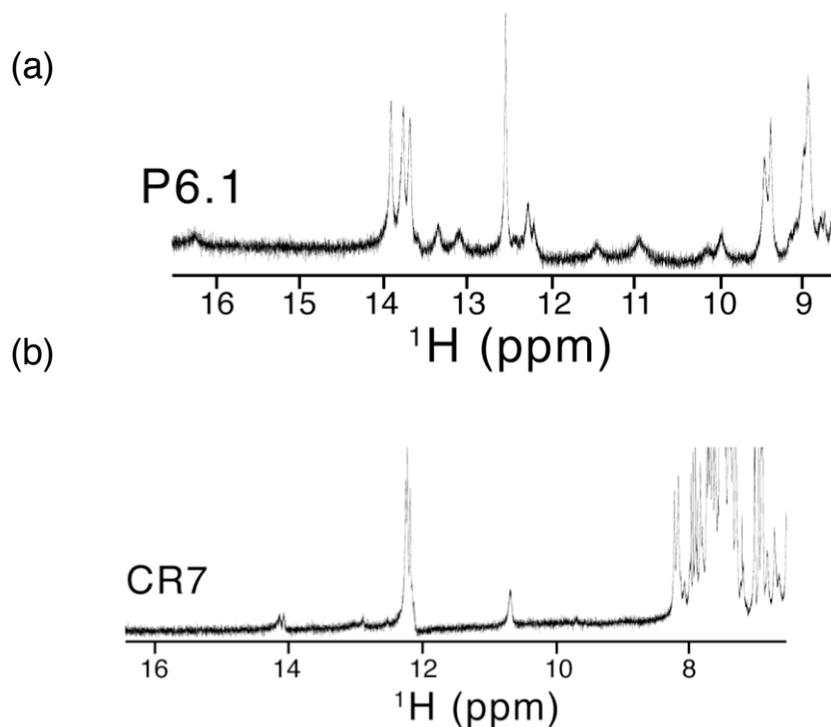
Figure 6.8. NMR spectra of RNA targets. **(a)** D-P6.1 hairpin was heated and cooled rapidly before the NMR experiment. The imino region of the spectrum is shown. Peaks near 14 p.p.m. are most likely from G nucleotides while the peak at 12.5 p.p.m. results from a U nucleotide. Small peaks near 11 and 13 p.p.m. are possibly from RNA dimerization. **(b)** NMR of D-CR7. The sample was prepared as in **(a)**.

the hairpins are folded (Figure 6.8). However, several small peaks in the imino region were observed, suggesting the presence of a small amount of dimeric RNA. The high concentrations of hairpin required for NMR probably enhanced dimerization even though the hairpins were annealed before the spectra were taken.

These data suggested that the P6.1 and CR7 models were both folded and mostly monomeric assuring us that these RNA would be good targets for selection. Moreover, the low concentrations of target required for selection (~0.2 μM) would further reduce the dimer population in our samples. Thus, we synthesized two biotinylated D-RNA targets corresponding to D-P6.1 and D-CR7. We also synthesized an L-P6.1 target for the reflection selection strategy.

Library Design

There are no known natural ligands for the P6.1 hairpin on which to base the design of our library (13). We therefore constructed two random libraries that differed on the degree of randomization. The first library contained 27 random amino acids (X27) and has been previously used to discover peptides that bind Methuselah, a *Drosophila melanogaster* G-protein coupled receptor (B. Ja. and R.W. Roberts, in preparation). We reasoned that the totally random library would sample sequence space evenly, preventing any bias in the solutions obtained. The library contained a 3' SpeI cloning site, coding for the amino acid sequence TS and a GGLRASAI C-terminal constant region.

The second library was based on the observation that many RNA binding peptides contain arginine-rich domains (arginine-rich motif, ARM) (35). Alignment of a number of ARM sequences (Figure 6.3) showed many sequences possessed an RRXRR sequence.

We designed a random library containing a central RRXRR motif where each Arg was doped at ~50% (the probability of Arg at all four positions is 12.5%). The library also incorporated a 3' NgoMIV cloning site, coding for the amino acid sequence AG, a C-terminal tyrosine to allow for easy quantitation of selected peptides, and a QLRNSCA constant region (13).

λ N$_{(1-22)}$	MDAQTRRRRERRAEKQAQWKAAN
P22 N$_{(14-30)}$	NAKTRRHERRRKLAIER
ϕ21 N$_{(12-39)}$	TAKTRYKARRAELIAERR
HIV-Rev$_{(34-50)}$	RQARRNRRRRWRERQR
HIV-TAT$_{(49-57)}$	RKKRRQRRR
BMV Gag$_{(7-25)}$	KMTRAQRRAAARRNRWTAR
CCMV Gag$_{(7-25)}$	KLTRAQRRAAARKNKRNTR
Yeast PRP6$_{(129-144)}$	TRRNKRNRIQEQLNRK
Human U2AF$_{(142-153)}$	SQMTRQARRLYV
L16$_{(51-63)}$	RRAMSRKFRRNSK
S7$_{(91-99)}$	KTKLERRNK
HTLV-II Rex$_{(4-16)}$	TRRQRTRRARRNR
FHV coat$_{(35-49)}$	RRRRNRTRRNRRRVR
Consensus	RRXRR

Figure 6.3. Alignment of several arginine rich peptides. Adapted from (35). Arginine rich regions are often of the sequence RRXRR.

D-CR7 Selection

Utilizing the RRXRR library, we synthesized a pool of mRNA protein fusions containing 100 trillion (10^{14}) independent sequences and sieved the pool against immobilized D-CR7. We performed a total of fifteen rounds of selection, alternating matrices every few rounds to avoid selecting matrix-binding sequences (Figure 6.4). The round 15 pool bound reasonably well, but also contained modest affinity to the matrix

with no target (data not shown). Attempts to increase binding affinity by increasing binding stringency or decrease nonspecific binding by using different immobilization matrices and by preclearing the library were unsuccessful. The round 15 library was sequenced (Figure 6.5) yielding one sequence that dominated the pool. The winning sequence contains a two-base deletion in the 3' NgoMIV restiction site causing a shift into the +3 reading frame.
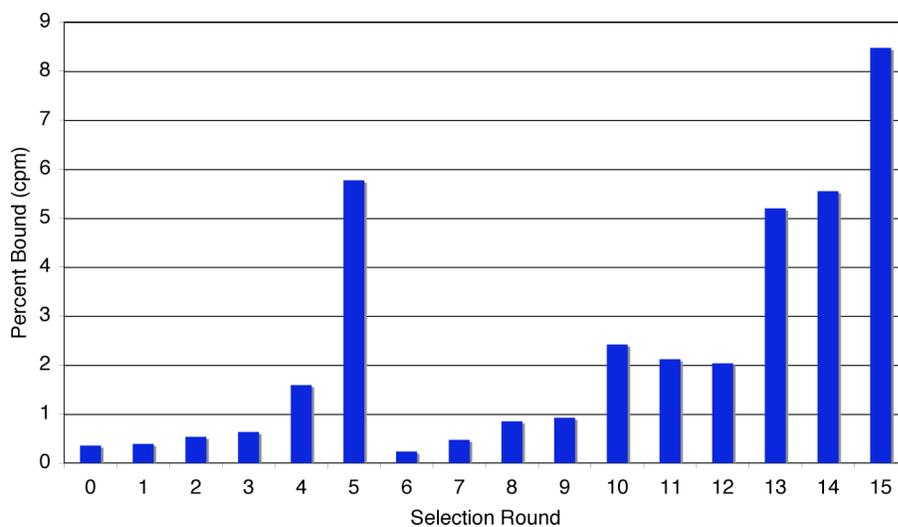
Figure 6.4. Binding of the round 0-15 pools to immobilized D-CR7. Binding percentages are corrected for matrix binding.

Clone	Peptide Sequence

```
                     1        10        20        30
                     •        •         •         •
        Pool      MXXXXXrrXrrXXXXXXXXXXXAGYQLRNSCA

CR7_15J1   MFLIYFDRVRRRMKIDFIPSGSLPAQKQLR
CR7_15J2   MFLIYFDHVRRRMKIDFIPSGPLPAQKQLR
CR7_15J3   MFLIYFDRVRRRMKIDFIPSGPLPAQKQLR
CR7_15J4   MFVIYFDRVRRRMKIDFIPSGSLPAQKQLR
CR7_15J5   MFLIYFDHVRRRMKIDFIPSGSLPAQKQLR
CR7_15J6   MFLIYFDRVRRRMKIDFIPSGSLPAQKQLR
CR7_15J7   MFLIYFDHVRRRMKIDFIPSGSLPAQKQLR
CR7_15J8   MFVIYFDRVRRRMKIDFIPSGSLPAQKQLR
CR7_15J9   MFVIYFDRVRRRMKIDFIPTGSLPAQKQLR
CR7_15J10  MFVIYFDRVRQRMKIDFIPSGSLPAQKQLR
```

Figure 6.5. Peptide sequences from round 15J. Mutations are highlighted in bold text. The RRXRR pool sequence is shown as a reference.

A nonmutated winning sequence (15J3) and two sequences representing the common mutations R8H (15J2) and L3V (15J8) were selected for further study. Each clone was amplified from plasmid DNA, synthesized as a ^{35}S-labeled fusion, and used in an *in vitro* binding assay against immobilized CR7. The results (Figure 6.6) show that 15J2, 15J3, and 15J8 bind CR7 with similar affinity, suggesting that the R8H and L3V mutations do not affect the binding affinity of the peptides. However, the 15J3 sequence shows some affinity toward the matrix, explaining why attempts to reduce nonspecific binding failed. The 15J3 sequence also shows low net binding to two other immobilized RNAs (D-P6.1 and D-HCVIIId RNA).
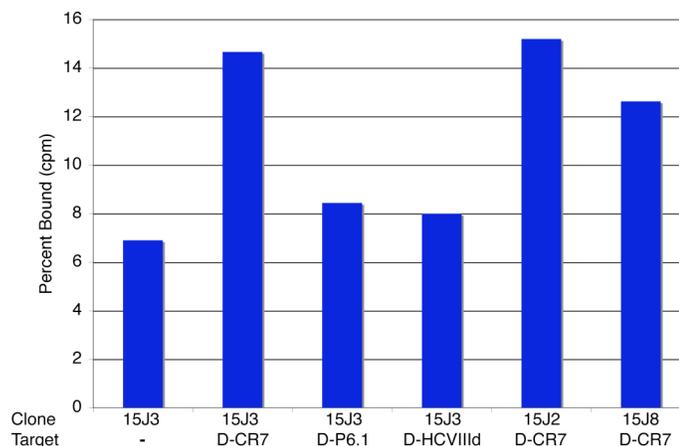
Figure 6.6. Binding of different round 15J clones to immobilized RNA targets. Individual clones corresponding to 15J2, 15J3, and 15J8 were synthesized as ^{35}S-labeled fusions and tested for binding to immobilized D-CR7, DHCVIIId, and D-P6.1. A no target control is shown for the 15J3 sequence (-).

DP6.1 Selection

We performed a similar selection targeting immobilized D-P6.1 using the RRXRR library. Using a 100 trillion-member library (10^{14} sequences), we performed a total of eight rounds of selection against immobilized D-P6.1 (Figure 6.7). The matrix was switched to streptavidin magnetic beads at round five in order to reduce matrix-binding sequences. Increasing binding stringency at round eight resulted in no further increase in binding, and the round eight pool was sequenced (Figure 6.8). Once again, a single sequence (termed 8H1) dominated the pool. 8H1 and a truncated variant changing the C-terminal constant region (QLRNSCA) to GGGG were amplified from plasmid DNA, synthesized as ^{35}S-labeled fusions, and tested for binding against immobilized D-P6.1. Figure 6.9 shows that both 8H1 and truncated 8H1 bind to D-P6.1, however truncated 8H1 possesses higher affinity than full-length 8H1 demonstrating that the QLRNSCA constant region is not required for, and may even hinder binding.
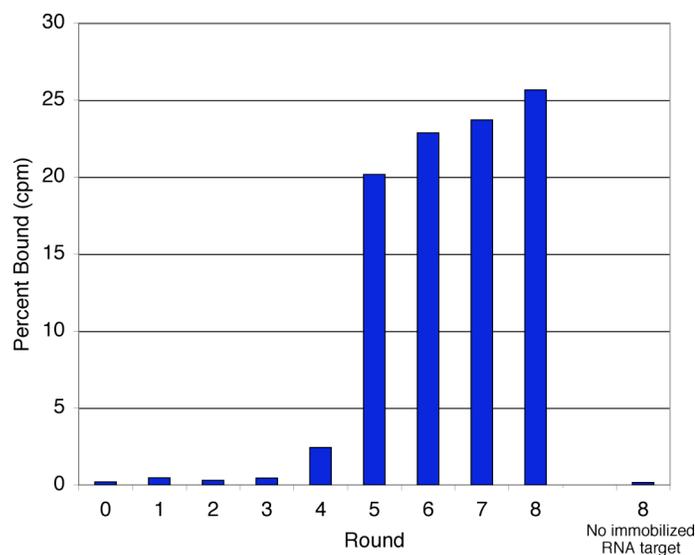
Figure 6.7. Binding of rounds 0-8 to immobilized D-P6.1. A no target control from round 8 is shown.

Clone	Peptide Sequence			
	1	10	20	30
	•	•	•	•
Pool	MXXXXXrrXrrXXXXXXXXXXXAGYQLRNSCA			
RRXRR8H.1	MNDARRNRKYLRVKRLRIQKM--YQLRNSCA			
RRXRR8H.2	MNDARRNRKYLRVKRLRIQKM--YQLRNSCA			
RRXRR8H.3	MNDARRNRKYLRVKRLRIQKM--.QLRNSCA			
RRXRR8H.4	TNDARRNRKYLRVKRLRIQKT--YQLRNSCA			
RRXRR8H.6	MNDARRNRKYLRVKRLRIQKM--YQLRNSCA			
RRXRR8H.7	MNDARRNRKYLRVKRLRIQKM--YQLRNSCA			
RRXRR8H.8	MNDARRNRKYLRVKRLRIQKM--YQLRNSCA			

Figure 6.8. Sequences from round 8H targeting D-P6.1. The starting library is shown for reference. Lowercase (r) positions indicate a 50% probability of Arg. Dashes (-) indicate a deletion. The red period indicates a stop codon. Peptide mutations are highlighted in bold text.
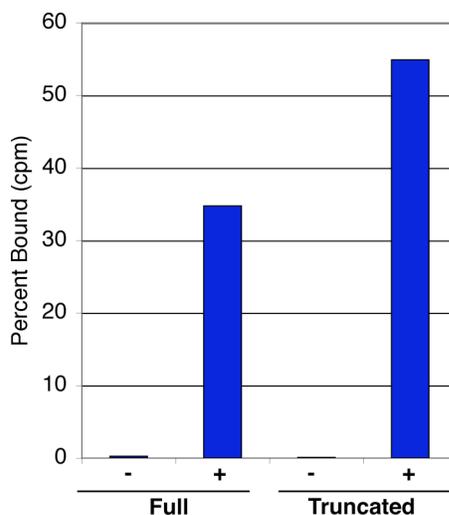
Figure 6.9. Binding of full-length and truncated 8H1. Truncated 8H1 replaces the QLRNSCA constant region with GGGG. Binding reactions containing no target (-) are shown as controls.

We chemically synthesized a peptide based on truncated 8H1 for further biophysical characterization. Attempts to determine a binding constant by a gel shift assay or by titration against 2-aminopurine labeled P6.1 were unsuccessful (data not shown). Preliminary experiments using surface plasmon resonance yield a k_{on} of 2.2 x 10^7 M^{-1}s^{-1}, a k_{off} of 0.048 s^{-1}, and a K_d of 2.2 nM (Figure 6.10). However, binding under these experimental conditions is mass-transport limited and exacerbated by the fact that only ~25% of immobilized D-P6.1 is functional, indicating that P6.1 may be conformationally heterogeneous.

8H1 is highly homologous to the winning sequence (18B) selected from a 10^{13}-member library targeting D-P6.1 using the X27S library (C. Ueda, unpublished observations). Alignment of 8H1 and 18B (Figure 6.11) shows that ten amino acids are conserved, suggesting these positions are involved in binding. Both sequences are

predicted to form α-helices (36) and the conserved residues are spaced with an i, i+3 or i,

i+4 pattern suggesting they are located on the same face of an α-helix.
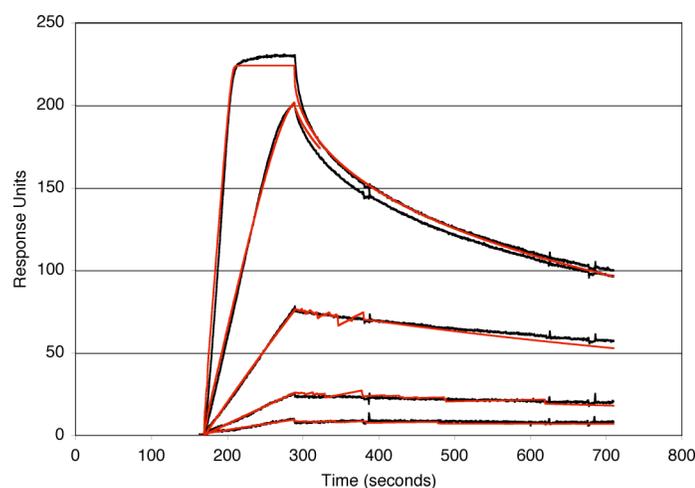
Figure 6.10. The binding of 8H1 studied by surface plasmon resonance. 8H1 was injected at 1.22, 3.67, 11, 33, and 99 nM and bound to D-P6.1 immobilized on a streptavidin sensor chip. Sensorgrams (black) are corrected for drift using a flow cell containing no RNA target. Kinetic fits are shown in red.

```
   8H1        MNDARRNRKYLRVKRLRIQKMYQLRNSCA
   18B       MTTSARSLRKYYRVRLLRAFKTRPARHVTSGGLRASAI
Consensus       AR   RKY RV+ LR   K
```

Figure 6.11. Alignment of 18B and 8H1 peptides. 18B was selected from the X27S library while 8H1 was selected from the RRXRR library. Constant regions are underlined. Positions common to both peptides are highlighted in bold text. (+) designates the position requires a positive charge (Lys or Arg).

The 8H1 sequence also contains a six-base deletion that destroys the NgoM IV

restriction site. The deletion of the NgoM IV restriction site has also been observed in

the D-CR7 selection (above) and other selections using the RRXRR library (data not

shown). The deletion of the site is unlikely to be due to poor construction of the library

as the majority of the sequences in the starting pool contain the NgoM IV site (Figure

6.S1). It is possible that the amino acid sequence (AG) is a site for protease cleavage, however fusion formation of the RRXRR library is comparable to other libraries used for selection. A likely explanation is that the nucleotide sequence of the restriction site (5'-GCCGGC-3') resulted in poor amplification of the template sequence. The NgoM IV restriction site has been used previously (13), however in that experiment, it was not subjected to multiple cycles of selection, as in these experiments. Regardless of the reason for bias, the restriction site should be avoided for future mRNA display libraries.

L-P6.1 Selection

It was unclear structurally how and if an L-peptide would bind to an L-RNA. We thus chose the totally random X27S library to target L-P6.1 since it would not bias the
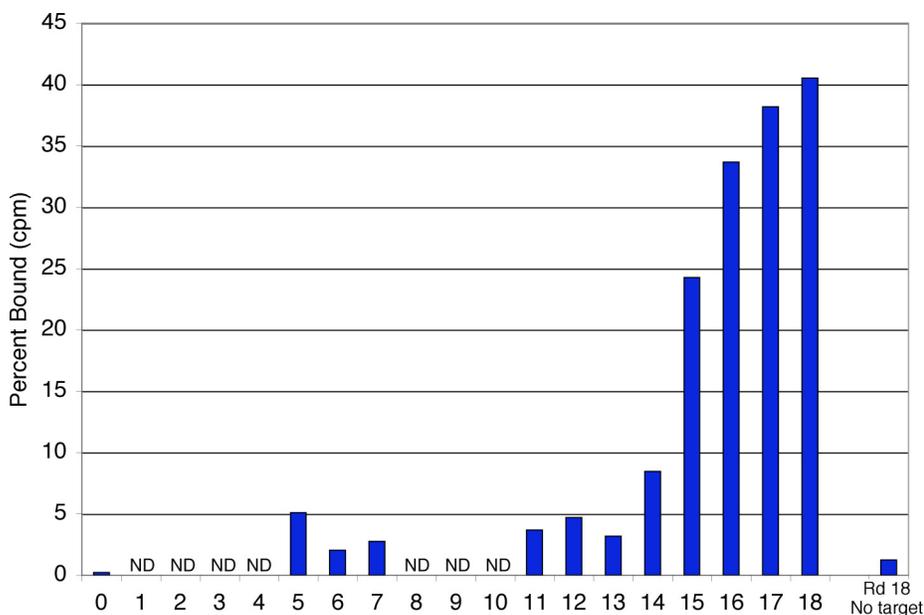
Figure 6.12. Binding of rounds 0-18 to immobilized L-P6.1. (ND) indicates that no binding assay was performed for that particular round. Binding of the round 18 pool to matrix containing no L-P6.1 is shown as a control.

selection. We constructed an mRNA display library totaling ten trillion (10^{13}) independent sequences and sieved the library against immobilized L-P6.1. After 16 rounds of selection, the library bound L-P6.1 very well (Figure 6.12) and we tested the ability of the pool to bind other immobilized RNA targets. Incubation of ^{35}S-labeled fusions with L- versions of *boxB* and CR7 as well as D-P6.1 resulted in little binding to these noncognate targets (Figure 6.13). However, the pool does show binding to an L-HCV IIId target (data not shown, discussed in Chapter 7).
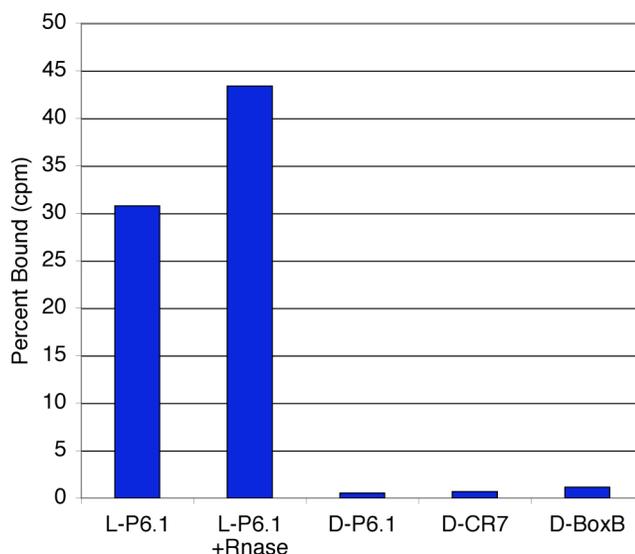
Figure 6.13. The specificity of the round 16 pool of the L-P6.1 selection. ^{35}S-labeled fusions were incubated with different immobilized RNA targets and binding percentages were determined by scintillation counting. A sample was treated with RNase (+RNase) to show that binding is not due to the mRNA component of the fusion.

It was also necessary to test if the mRNA contributed any affinity to the binding interaction. Addition of RNase to the binding reaction shows that the fusions retain binding to the L-P6.1 target ruling out any binding contribution of the mRNA (Figure 6.20). In fact, the RNase-treated fusions show higher binding affinity than the mRNA-

protein fusions, as observed previously (13). These data also demonstrate that the L-enantiomer of RNA is resistant to degradation by RNase.

After two additional rounds of selection, little increase in binding was seen and the pool was sequenced. Figure 6.14 shows that the selected sequences belong to three sequence families with no consensus to each other. One sequence (18C1) composed roughly 60% of all clones (10/17) while another sequence (18C5) composed roughly a third of all sequences (6/17). The third sequence (18C17) was represented only once.

Clone	Peptide Sequence			
	1	10	20	30
	•	•	•	•
Pool	MXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXTSGGLRASAI			
18C1, 2, 9, 12, 19	MMDWKRAKLNRLSVRKLRKYADYF----TSGGLRASAI			
18C7	MMDWKRAKL**S**RLSVRKLRKYADYF----**A**SGGLRASAI			
18C13	-MDWKRAKL**S**RLSVRKLRKYADYF----TSGGLRASAI			
18C11	MT**DY**KRAKLN**L**LSVRKLRKYADYF----TSGGLRASAI			
18C14	MT**AC**KRAKLNRLSVRKLRKYADYF----TSGGLRASAI			
18C18	MMDW**R**RAKLNRLSA**R**KLRKYADYF----T**R**GGLRASAI			
18C5, 10	MGYLTPKGRALKLRMLDRNRRRKAKSGVTSGGLRASAI			
18C20	MGYLTPKGRALKLRMLDRNRRRK**V**KSGVTSGGLRASAI			
18C4	MGYLTPKGRALKLRML**S**RNRRR**E**AKSGVTSGGLRASAI			
18C6	MGYLTPKGRALKLRMLDR**I**RRR**E**AKSGVTSGGLRASAI			
18C16	MGYLTPKGRALKLRMLDR**I**RRR**EV**KSGVTSGGLRASAI			
18C17	MKHSNSSRGRKTLWRALTLWLLMQSLKRTSGGLRASAI			

Figure 6.14. Sequences of the round 18 pool targeting L-P6.1. The starting pool is shown as a reference. Peptide sequences were grouped into three families. Peptides belonging to the 18C1 family contain a 12 base deletion resulting loss of four amino acids, indicated by (-). Mutations are highlighted in bold text.

Representative sequences from each family (18C1, 18C5, and 18C17) were synthesized as ^{35}S-labeled fusions, and tested for binding to immobilized L-P6.1 (Figure 6.15). The results show that all three families bind immobilized L-P6.1, however the sequences possess different binding affinities. 18C1 showed the highest affinity,

explaining why it came to dominate the round 18 pool. Surprisingly, 18C5 showed much less affinity to L-P6.1 than 18C17 and 18C1; the frequency it appeared in the round 18 pool implied that it should have higher affinity. These data suggest that each family binds L-P6.1 utilizing different contacts to the RNA. Additionally, there is an inverse correlation between the net charge (+6, +9, +7 for 18C1, 18C5, and 18C17, respectively) and affinity for LP6.1, implying that contacts other than ionic interactions are important.
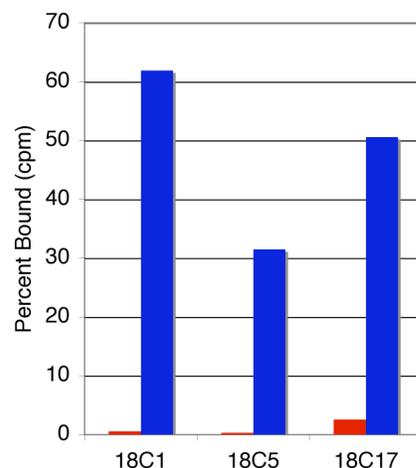
Figure 6.15. Binding of individual clones from round 18 of the L-P6.1 selection. Sequences representative of the three peptide families were synthesized as ^{35}S-labeled fusions and tested against immobilized L-P6.1. Fusions were also bound to matrix lacking target (red bars) to show binding was specific for L-P6.1.

We constructed a number of deletion constructs in order to define the minimal binding domain of each peptide. The peptides are predicted to adopt α-helical conformations (36) and these predicted helical regions were used to design deletion constructs. The deletion constructs were synthesized as ^{35}S-labeled fusions and several of the constructs showed little binding affinity toward immobilized L-P6.1 (Figure 6.16).

For peptide 18C1, nearly the entire randomized region is necessary for binding, while 18C17 shows reduced affinity if N- or C-terminal regions are deleted.
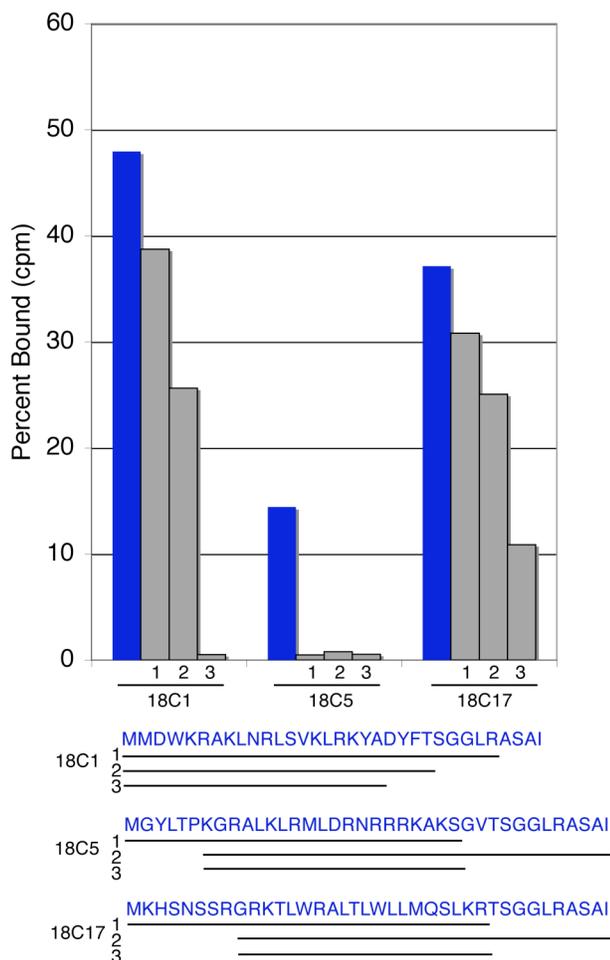
Figure 6.16. Deletion mapping of round 18C clones. Deletion constructs were constructed by PCR and tested for binding to immobilized L-P6.1. Black lines represent amino acids included in each respective construct. Methionine was added to constructs containing N-terminal deletions in order to initiate translation.

Conclusions

The peptides we have selected are highly basic, arginine rich, and are all predicted to form α-helicies. Our data support the notion that arginine is an important amino acid

for RNA recognition, both for its charge complementarity to RNA and its ability to form hydrogen bonds. While helical prediction programs are not always accurate, the fact that all sequences are predicted to be helical may stem from the favorable helical propensities of Arg and Lys. Other arginine rich sequences adopt α-helicies (Rev (37)) or β-hairpins (Tat (38)), however sequences that recognize RNA hairpins are helical (39-43). It is therefore possible that α-helical structures are a general method to recognize RNA hairpins.

It is surprising that only a few solutions to a particular target were found. Only one in 10^{13} peptides possessed the ability to specifically bind the immobilized RNA targets with high affinity. Two selections using different random libraries resulted in the same consensus sequence, suggesting that very few peptide solutions exist for recognition of D-P6.1. The large number of selection cycles required for our experiments most likely reflects the relative paucity of solutions in the initial library. While it is probable that there are many more sequences in the selection pools, it is likely that few of these possess affinities comparable to the peptides we have selected since they were selected against. It is important to note that the peptides we have discovered would not have been found by phage display since a maximum of 10^9 sequences can be sieved by phage display.

It might seem, to a first approximation, that many RNA binding peptides should exist in the initial library since any sequence with basic residues should bind with high affinity (44-46). However, many of these sequences lack *specificity* and were removed by addition of tRNA competitor. Additionally, an inverse correlation is seen between binding affinity and total net charge of the selected peptides suggesting that other, non-electrostatic interactions are also important for RNA binding due to contributions to

overall affinity and specificity. Indeed, hydrophobic contacts that contribute to specificity are often found in other RNA-peptide or RNA-protein complexes (47,48).

In contrast with the selections described here, previous experiments targeting the *boxB* RNA resulted in selection for a diverse collection of RNA binding peptides (13). The difference in the number of selected solutions is probably due to the use of a constant region in the *boxB* selection that possessed micromolar affinity. The constant region probably increased the number of solutions in the starting pool, resulting in more sequences at the end of the selection. Indeed, the selection for *boxB* converged in fewer rounds than most of the selections described here. Additionally, mutagenesis and reselection of winning sequences described here against L-P6.1 and the L-HCV IIId targets resulted in many more independent solutions (Chapter 7).

Our results show that it is possible to excise a domain from a biologically important RNA and isolate high affinity peptides that bind this domain. While we have targeted RNA hairpins, natural peptides bind a variety of RNA structures (47, 48) and it is therefore likely that high affinity ligands to many other folded RNAs can be isolated from random peptide libraries. Since RNA constructs often fold into similar structures when removed from, or in the context of larger RNAs, our peptides probably bind to CR7 and P6.1 in the context of full-length hTR. Lastly, based on the *in vitro* binding assays, our peptides probably possess nanomolar affinitites. Future experiments will address the ability of our peptides to inhibit telomerase activity.

Materials and Methods:

Construction of RRXRR Library

Single stranded DNA (5'-GGCGCAGCTGTTTCTGAGCTGGTAGCCGGCSNNS NNSNNSNNSNNSNNSNNSNNSNNS21S21SNNS21S21SNNSNNSNNSNNSNN CATTGTAATTGTAAATAGTAATTGTCCC where N= an equimolar mixture of A,C,G,T; S= an equimolar mixture of G,C; 1=35:10:5:50 of T:C:A:G; 2=10:70:10:10 of T:C:A:G) was synthesized at the Keck Oligo Facility at Yale (http://keck.med.yale.edu/oligos/). The composition of each mixed base was adjusted to correct for the different coupling rates of each base (N= 3:3:2:2 for A:C:G:T; S= 3:2 for C:G; 1=23:10:5:33 for T:C:A:G; 2= 66:100:100:466 for T:C:A:G). The synthetic DNA was purified by 8% urea-PAGE and electroeluted. PCR conditions were optimized and were: 10 mM Tris-HCl, pH 9.0 at room temperature, 50 mM KCl, 2.5 mM MgCl$_2$, 0.1% Triton X-100 (Sigma), 4 pmol of template, 3 µM of the 42.108 5' primer (5'-TAATACGACTCACTATAGGGACAATTACTATTTACAATTACA-3'), 3 µM of the 3' primer 3RRXRR (5'-GGCGCAGCTGTTTCTGAGCTGGTA-3'), and 0.2 mM each of dATP, dTTP, dCTP, and dGTP. PCR reactions were performed in 0.65 mL thick-walled tubes (final volume 200 µL, no heated lid on the thermocycler), heated to 95 ºC before Taq was added, and cycled at 95 ºC for two minutes followed by six cycles of two minutes at 95 ºC, two minutes at 53 ºC, and four minutes at 72 ºC. A total of 24 mL of PCR was performed yielding 510-1200 µg of DNA product as determined by quantification against a mass ladder on an agarose gel. The PCR reactions were

combined, phenol extracted, ethanol precipitated, and resuspended in 1.2 mL of 1X TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0).

The 5' primer (42.108) was ^{32}P-labeled with polynucleotide kinase (PNK) and γ-^{32}P ATP. One cycle of PCR was performed under the conditions described above, except the 3' primer was omitted and template: 5' primer molar ratios were varied from 2:1, 5:1, 10:1, and 20:1. Roughly 20% of the input template was extendable (data not shown).

Eight hundred picomoles of PCR product were used in eight, one mL transcription reactions for more than 2 hours at 37 °C. Transcription reactions were 40 mM HEPES-KOH, pH 7.5, 2 mM spermidine, 40 mM DTT, 25 mM MgCl$_2$, 1X RNASecure (Ambion) and 4 mM each of UTP, GTP, ATP, and CTP (Sigma) at pH 8.0. The reactions were stopped by addition of 100 μL of 0.5 M EDTA, pH 8.0, and were phenol extracted and ethanol precipitated. The RNA was purified by 8% urea-PAGE, electroeluted, and ethanol precipitated.

The RNA was ligated to pF30P (5' pdA21-[C9]3-dAdCdC-puromycin-3', where p is a 5' phosphate and C9 is spacer phosphoramidite 9 (Glen research)) using the splint RRXRR splint (5'- TTTTTTTTTTTNGGCGCAGCTGT-3', where N is 70% T and 10% of A, C, and G) and T4 DNA Ligase (NEB) in T4 DNA ligase buffer (50 mM Tris-HCl, pH 7.5, 10 mM MgCl$_2$, 1 mM ATP, 10 mM DTT, and 25 μg/mL Bsa). The ligation reaction consisted of RNA, splint, and pF30P at a molar ratio of 1:1.5:1, respectively. After two hours at room temperature, the ligation reaction was purified by denaturing PAGE, electroeluted, and ethanol precipitated. The RNA and ligated template were quantified by absorbance at 260 nm and using the biopolymer calculator (http://paris.chem.yale .edu/extinct.html).

Rabbit reticulocyte lysate was prepared according to the method of Jackson and Hunt (49) and translation and fusion formation efficiencies were determined to be comparable to those obtained in a commercial rabbit reticulocyte lysate (Novagen). Fusion formation efficiencies were determined to be ~20% by translation of a ^{32}P-labeled ligated template and visualization on a tricine gel or urea-PAGE. A total of 20 mL of the homemade lysate and 5 mL of the commercial lysate (Novagen) were used to translate 10,000 pmol of ligated template (0.4 μM final concentration). Both lysates contained 20 mM HEPES-KOH, pH 7.6, 100 mM KCl, 0.5 mM MgCl$_2$, 2 mM DTT, 8 mM creatine phosphate, 25 μM of each amino acid except for Leu and Cys, which were each 12.5 μM. A small amount of ^{32}P-labeled ligated template was added so that purification yields could be determined. Translation was conducted for one hour at 30 °C, after which KCl and MgCl$_2$ were added to final concentrations of 500 mM and 50 mM, respectively.

In order to simplify sample purification, the lysate was processed in 2.5 mL batches. For each 2.5 mL of translation, dT purification was performed with 100 mg of dT cellulose (NEB) prewashed with Isolation buffer (IB, 100 mM Tris-HCl, pH 8.0, 1 M NaCl, 0.2% Triton X-100) and 25 mL of IB. After one hour at 4 °C, the entire sample was added to a plastic disposable column (Biorad) containing a frit blocked with glass wool. The beads were washed with 20 mL (~20 column volumes) of IB and eluted with five, 0.5 mL fractions of ddH$_2$O. The sample was precipitated in the presence of linear acrylamide (Ambion) and reverse transcribed in a total volume of 400 μL. The reactions were then incubated with 800 pmol of annealed target immobilized on neutravidin agarose (Pierce), and the buffer adjusted with 36 μL of 2.5 M KCl, 2.4 μL of 5% Nonidet P-40 (NP-40), 3 μL of 20 mg/mL yeast tRNA (Roche), and 680 μL of ddH$_2$O. The beads

were transferred to a Spin-X column, washed three times with 700 μL of 1X N binding

buffer (NBB, 10 mM HEPES-KOH pH 7.5, 0.5 mM EDTA, pH 8.0, 100 mM KCl, 1 mM

$MgCl_2$, 1 mM DTT, 0.01% (v/v) NP-40), + 50 μg/mL tRNA, and transferred to a new

eppendorf tube with 1 mL of 1X NBB + tRNA, and the supernatant removed.

One-tenth of the beads was used to determine the number of cycles required for

amplification of the bound DNA. The PCR reactions were performed under conditions

described above except that the cycle times were shortened to 30 seconds at 94 °C, 30

seconds at 55 °C, and 45 seconds at 72 °C.

The yields of dT and ethanol precipitation ranged from 57-74% of input mRNA

template recovered. The RT efficiencies were determined to be >90%. This resulted in

roughly 3×10^{14} molecules that entered the selective step for the D-P6.1 selection (~3.1

copies of each individual sequence) and $\sim 3 \times 10^{14}$ molecules that entered the selective

step for the D-CR7 selection (~2.9 copies of each individual sequence).

RRXRR Selection Rounds

Subsequent rounds of selection were similar to the first round except reaction volumes

were reduced. Each round consisted of one mL of PCR, which was phenol extracted and

ethanol precipitated, and resuspended in 200 μL of 1X TE buffer. Half of the DNA stock

was used in a one mL *in vitro* transcription for >2 hours at 37 °C. The mRNA was

desalted with a NAP-25 column and used in a 250-500 μL ligation. Ligated mRNA was

translated in a 100 μL reaction after which salts were added and the reaction incubated at

room temperature for one hour or overnight at -20 °C. The fusions were purified by dT-

cellulose with one mL of IB and 100 μL of 25% (v/v) dT cellulose pre-equilibrated in IB.

After one hour at 4 ℃, the slurry was transferred to a Spin-X column, washed five times with 700 μL fractions of IB, and eluted with two fractions of 250 μL of 65 ℃ H_2O. The eluent was ethanol precipitated in the presence of linear acrylamide, and reverse transcribed in a total volume of 40 μL.

The binding buffer was switched at round 2 for the D-P6.1 selection and the D-CR7 selection to 10 mM Tris-HCl pH 7.5, 0.5 mM EDTA, pH 8.0, 100 mM NaCl, 1 mM $MgCl_2$, 1 mM DTT, 0.01% (v/v) Tween-20, and 50 μg/mL tRNA. At round 6 of the D-P6.1 selection, the immobilization matrix was changed from neutravidin agarose to magnetic streptavidin beads (Genovision). Competitor tRNA concentrations were also increased to 500 μg /mL and 5,000 μg /mL as binding affinities of the pools increased. At round 6 of the D-CR7 selection, the buffer was changed to 1X SB and the protocols for selection were changed to those of the L-P6.1 selection (See section below).

Construction of X27S Library:

The X27S library DNA was amplified from 1,200 pmol of ssDNA 142.1 template (5'-TTAAATAGCGGATGCACGCAGACCGCCACTAGTSNNSNNSNNSNNSNNS NNSNNSNNSNNSNNSNNSNNSNNSNNSNNSNNSNNSNNSNNSNNSNNSNNSNN SNNSNNSNNSNNCATTGTAATTGTAAATAGTAATTGTCCC-3') using 1 μM of the 5' primer, 47T7FP (5'- GGATTCTAATACGACTCACTATAGGGACAATTACTATT TACAATTAC-3') and 1 μM of the 3' primer, 3X27S (5'-AATAGCGGATGCAC GCAGACC-3'). Samples were heated to 94 ℃, Taq was added and allowed to equilibrate for two minutes at 94 ℃ and followed by six cycles of 94 ℃ for one minute, 55 ℃ for 1 minute, and 72 ℃ for 2 minutes.

The DNA was phenol extracted, ethanol precipitated, and resuspended in 1 mL of 1X TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, pH 8.0) yielding a solution of 370 ng of DNA/μL. DNA concentration was determined by comparing a dilution series to a mass ladder (100 bp ladder, NEB).

Roughly 670 pmol of DNA (~73 μg of DNA) were *in vitro* transcribed in a 6 mL volume for 3 hours at 37 ºC. RNA was gel purified by 6% denaturing PAGE and electroeluted. RNA was ligated to pF30P using the splint X27Splint (5'-TTTTTTTTTTTNAATAGCGGATG-3', where N is 70% T and 10% of A, C, and G) and T4 DNA Ligase (NEB). The ligation reaction consisted of RNA, splint, and pF30P at a molar ratio of 1:1.5:1, respectively. After two hours at room temperature, the ligation reaction was purified by denaturing PAGE, and electroeluted, yielding 3,600 pmol of ligated product.

Ligated mRNA (3,000 pmol) was translated in a total volume of 7.5 mL (0.4 μM template) as described previously (13), salt was added to facilitate fusion formation, and the reaction incubated overnight at -20 ºC. A small amount of ^{32}P-labeled ligated template was translated so that product yields could be determined.

The translated material was purified with 350 mg of dT-cellulose (NEB) that had been prewashed with Isolation buffer (IB, 100 mM Tris-HCl, pH 8.0, 1 M NaCl, 0.2% Triton X-100). The dT purification was performed at 4 ºC for one hour, after which the beads were washed with 20 mL of IB on three disposable columns (Biorad) containing frits blocked with glass wool. Bound fusions on each column were eluted with eight, 0.5 mL washes of 65 ºC ddH$_2$O. All eluates were pooled, then ethanol precipitated in the presence of linear acrylamide, as described previously (14). The pellets were

resuspended in 1X RT buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM MgCl$_2$, 10 mM DTT, 0.25 mM each dNTP, 1 µM of the 3' X27S primer) filtered with a 0.45 µm Spin-X column (Costar, Corning), and the top of the column washed with one 150 µL and two 50 µL fractions of ddH$_2$O. This step was necessary to remove dT cellulose that had been carried over from the dT purification step. All fractions were combined and the buffer adjusted to 1X RT in a final volume of 1,212 µL. The reactions were heated to 65 °C for 5 minutes, cooled on ice, and Superscript II (Invitrogen) was added. The samples were incubated at 42 °C for 1 hour then heated to 65 °C for 5 minutes to deactivate the Superscript II. The overall yield of material carried through dT purification and ethanol precipitation was 43-54%. Fusion efficiency was measured to be 15% of input template by translation of a ^{32}P-labeled ligated template (data not shown) while RT efficiency was measured to be >90%. Half the sample was used for selection against L-P6.1 while the other half was bound to D-P6.1 (C. Ueda, unpublished observations). Based on these yields, 5-6 x 10^{13} molecules entered the selective step for each RNA target, of which 1-2 x 10^{13} were independent sequences.

L-P6.1 (600 pmol) was annealed in 1X selection buffer (SB, 10 mM Tris-HCl, pH 7.5, 150 mM KCl, 0.5 mM EDTA, pH 8.0, 0.01% (v/v) Tween-20, 1 mM DTT) by heating the sample to 90 °C for one minute, then cooling rapidly on ice. Neutravidin agarose (112.5 µL of beads) was pre-equilibrated in 1X SB. Annealed L-P6.1 was added to the beads and the mixture incubated at room temperature for 30 minutes after which the supernatant was removed.

The buffer of the reverse transcription reaction was adjusted to selection conditions by the addition of 54 µL of 2.5 M KCl, 1.8 µL of 10% Tween-20, 9 µL of 10

mg/mL tRNA, 2 μL of EDTA, pH 8.0, and 1,000 μL of H2O. The entire reaction was added to the immobilized L-P6.1 on neutravidin beads and incubated at 4 ℃ for 1.5 hours. Twenty microliters of 2 mM biotin in 1X PBS (4.3 mM Na_2HPO_4-$7H_2O$, 1.4 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl) was added and the reaction incubated at 4 ℃ for an additional 30 minutes. The beads were transferred to a Spin-X column, washed three times with 700 μL of 1X SB + 50 μg/mL tRNA, and transferred to a new eppendorf tube with 1 mL of 1X SB + tRNA, and the supernatant removed.

One-tenth of the beads was used to determine the number of cycles required for amplification of the bound DNA. The PCR reactions were performed under conditions described above except that the cycle times were shortened to 30 seconds at 94 ℃, 30 seconds at 55 ℃, and 45 seconds at 72 ℃. Six, nine, and eleven cycles of PCR were performed and the products run on a sodium borate-agarose gel (50). The remaining bound DNA was then amplified for 11 cycles in a total volume of 1 mL.

L-P6.1 Selection Rounds

The selection rounds for L-P6.1 using the X27S library are similar to those described for the RRXRR library, except as detailed below. One microliter of 100 μM L-P6.1 was added to 50 μL of binding buffer -tRNA, and annealed by heating to 90 ℃ and rapidly cooling on ice. The annealed RNA was added to 400 μL of binding buffer and 50 μL of 50:50 (v/v) neutravidin agarose slurry prewashed with binding buffer, and bound for one hour at room temperature. The supernatant was removed, and the reverse transcribed fusions, one mL of binding buffer with tRNA, and 10 μL of 2 mM D-biotin in 1X PBS was added. The fusions were incubated at 4 ℃ for one hour after which the

slurry was transferred to a Spin-X column and washed five times with 700 μL fractions of binding buffer with tRNA. The beads were resuspended with 700 μL of binding buffer, transferred to a 1.7 mL eppendorf tube, and the supernatant removed. 75 μL of ddH$_2$O was added and 10 μL of the slurry used to determine the number of cycles required for PCR.

For rounds where the matrix changed, Ultralink streptavidin (Pierce) was used instead of neutravidin agarose. Rounds where the magnetic streptavidin matrix (Genovision) was used were similar to those described above, except that after the fusions were bound, the magnetic beads were washed four times with 1 mL of binding buffer with tRNA using a robotic liquid handling system (Kingfisher, ThermoLab Systems). The last wash was conducted using a hand-held magnet and the beads were resuspended in 100 μL of ddH$_2$O before PCR. For rounds where tRNA concentration was increased, high tRNA concentrations (500 μg/mL or 5,000 μg/mL) were used in both the binding reaction and washes, except for the last few washes where tRNA concentration was decreased by ten-fold until it reached 50 μg/mL.

The matrix used for each round of selection is denoted in Figure 6.S.2.

In Vitro Binding Assay

dT-purified ^{35}S-labeled fusions (~100,000 cpm) were incubated with 0.1-0.2 μM immobilized, annealed RNA target (50 μL of a 50/50 (v/v) slurry) in binding buffer with 50 μg/mL tRNA for one hour at 4 °C. The samples were transferred to Spin-X columns

(Costar) and washed three times with 700 μL of binding buffer. The supernatant, washes, and beads counted by scintillation.

Synthesis of RNA Constructs

For NMR and native gel shift experiments, RNA constructs were synthesized by *in vitro* transcription of partially ssDNA templates using T7 RNA Polymerase (34). The D-CR7 template (5'-GTCCCACAGCTCAGGGACTATAGTGAGTCGTATTACGAATT-3') and D-P6.1 template (5'-GGAGAGCCCAACTCTCCTATAGTGAGTCGTATTACGAATT-3') were used with T7 promoter strand (5'- AATTCGTAATACGACTCACTATA-3') for the transcription of D-CR7 and D-P6.1 RNAs, respectively. The RNA was gel purified by 20% denaturing PAGE, electroeluted, and ethanol precipitated. The electroeluted RNA was used directly for native gel experiments.

RNA used for NMR was dissolved in 2 M NaCl and dialyzed extensively against ddH$_2$O to remove any Tris present in the sample. RNA was then freeze-dried, and resuspended in NMR buffer (10 mM phosphate, pH 6.0, 50 mM NaCl) in H$_2$O/D$_2$O (90:10 (v/v)).

D-P6.1 (5'-photocleavable biotin (Glen research)-AAAGGAGAGUUGGGCUC UCC-3') and D-CR7 (5'-photocleavable biotin-AAAGUCCCUGAGCUGUGGGAC-3') were synthesized at the Caltech Oligo Facility. The 2'-silyl protecting groups were removed by addition of 360 μL of triethylamine trihydrofluoride (Aldrich), incubation overnight at room temperature, and butanol precipitation. L-P6.1 (5'-GGAGAGUUGGGCUC UCCAAA-Biotin-3') was synthesized and supplied deprotected by Chemgenes. RNAs were purified by 20% denaturing PAGE and electroeluted.

Peptide Synthesis

The 8H1 peptide (NH$_2$-MNDARRNRKYLRVKRLRIQKMY-COOH) was synthesized using fluorenylmethyloxycarbonyl (Fmoc) chemistry on a 432A Synergy Peptide Synthesizer (Applied Biosystems). The peptide was deprotected in a trifluoroacetic acid (TFA)/1,2-ethanediol/thioanisole mixture (90:5:5, (v/v)) for 2 hours then washed with methyltertbutyl ether and resuspended in ddH$_2$O. The crude peptide mixture was purified by reverse-phase HPLC using a C18 semiprep column (250x10 mm, Vydac). Buffer A was a 95:5:0.1 (v/v) mixture of water/acetonitrile/TFA while buffer B was a 10:90:0.1 (v/v) mixture of water/acetonitrile/TFA. A gradient from 0%-40% B was run in 120 minutes and fractions containing the peptide were pooled. Final purity (>95%) was assessed by analytical HPLC on an analytical C18 column (250x4.6 mm, Vydac) using buffer A and buffer B (gradient of 0-100% buffer B in 50 minutes). The peptide mass was confirmed by MALDI-TOF mass spectrometry. The peptide concentration was determined by absorbance of tyrosine using the biopolymer calculator (http://paris.chem.yale.edu/extinct.html).

Native Gel Electrophoresis

RNAs were annealed in running buffer (0.5X TBE (44.5 mM Tris base, 44.5 mM boric acid, pH 8.3, 1 mM EDTA) containing 10% glycerol) by heating to 90 °C and rapidly cooling on ice. RNAs were run on a 10% acrylamide gel in a thermostated gel box (Hoeffer) in 0.5X TBE at 4 °C. Roughly 5,000 picomoles of RNA were run on each lane in order for the RNA to be visualized by UV-shadowing.

NMR Spectroscopy

NMR spectra were collected at 25 ºC as previously described (13, 15). RNAs were annealed before the NMR experiments by heating the sample to 90 ºC and cooling rapidly on ice. The concentration of P6.1 was 1.5 mM while CR7 was ~0.5 mM.

Sequencing

Pool templates were PCR amplified, gel purified using the QiaQuick gel extraction kit (Qiagen), and cloned using a TOPO-TA cloning kit (Invitrogen). Single colonies were grown in LB broth containing 50-100 µg/mL Ampicillin, and purified using the QiaPrep Spin Miniprep kit (Qiagen). Sequences were analyzed by Sequencher.

Biacore Analysis

Biacore analysis was performed on a Biacore 2000 instrument at 25 ºC. A streptavidin chip (Sensor Chip SA, Biacore) was preconditioned by three 50 µL injections of 40 mM NaOH, 2 M NaCl. Annealed P6.1 (5'-GGAGAGUUGGGCU CUCC-biotin) in HBS-EP buffer (10 mM HEPES-KOH pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% (v/v) Tween-20) was injected at a 1 µM concentration yielding ~1,000 RU of immobilized P6.1. 8H1 was diluted into running buffer (HBS-EP, 1 mM DTT, 50 µg/mL tRNA, 0.5 mg/mL Bsa) and injected at 99 nM, 33 nM, 11 nM, 3.67 nM, and 1.22 nM. The sensor chip was regenerated with injections of 2 M NaCl. Data were processed and analyzed using the BIAEvaluation software (Biacore).

References

1.   Cech, T.R. Life at the end of the chromosome: telomeres and telomerase. (2000) *Angew Chem Int Ed Engl* **39**, 34-43.

2.   Cong, Y.S., Wright, W.E. and Shay, J.W. Human telomerase and its regulation. (2002) *Microbiol Mol Biol Rev* **66**, 407-425, table of contents.

3.   Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W.   Specific association of human telomerase activity with immortal cells and cancer. (1994) *Science* **266**, 2011-2015.

4.   Hahn, W.C., Counter, C.M., Lundberg, A.S., Beijersbergen, R.L., Brooks, M.W. and Weinberg, R.A. Creation of human tumour cells with defined genetic elements. (1999) *Nature* **400**, 464-468.

5.   Counter, C.M., Hirte, H.W., Bacchetti, S. and Harley, C.B. Telomerase activity in human ovarian carcinoma. (1994) *Proc Natl Acad Sci U S A* **91**, 2900-2904.

6.   Harrison, R.J., Gowan, S.M., Kelland, L.R. and Neidle, S. Human telomerase inhibition by substituted acridine derivatives. (1999) *Bioorg Med Chem Lett* **9**, 2463-2468.

7.   Sun, D., Thompson, B., Cathers, B.E., Salazar, M., Kerwin, S.M., Trent, J.O., Jenkins, T.C., Neidle, S. and Hurley, L.H. Inhibition of human telomerase by a G-quadruplex-interactive compound. (1997) *J Med Chem* **40**, 2113-2116.

8. Strahl, C. and Blackburn, E.H. Effects of reverse transcriptase inhibitors on telomere length and telomerase activity in two immortalized human cell lines. (1996) *Mol Cell Biol* **16**, 53-65.

9. Asai, A., Oshima, Y., Yamamoto, Y., Uochi, T.A., Kusaka, H., Akinaga, S., Yamashita, Y., Pongracz, K., Pruzan, R., Wunder, E., Piatyszek, M., Li, S., Chin, A.C., Harley, C.B. and Gryaznov, S. A novel telomerase template antagonist (GRN163) as a potential anticancer agent. (2003) *Cancer Res* **63**, 3931-3939.

10. Norton, J.C., Piatyszek, M.A., Wright, W.E., Shay, J.W. and Corey, D.R. Inhibition of human telomerase activity by peptide nucleic acids. (1996) *Nat Biotechnol* **14**, 615-619.

11. Pitts, A.E. and Corey, D.R. Inhibition of human telomerase by 2'-O-methyl-RNA. (1998) *Proc Natl Acad Sci U S A* **95**, 11549-11554.

12. Dominick, P.K., Keppler, B.R., Legassie, J.D., Moon, I.K. and Jarstfer, M.B. Nucleic acid-binding ligands identify new mechanisms to inhibit telomerase. (2004) *Bioorg Med Chem Lett* **14**, 3467-3471.

13. Barrick, J.E., Takahashi, T.T., Ren, J., Xia, T. and Roberts, R.W. Large libraries reveal diverse solutions to an RNA recognition problem. (2001) *Proc Natl Acad Sci U S A* **98**, 12374-12378.

14. Barrick, J.E., Takahashi, T.T., Balakin, A. and Roberts, R.W. Selection of RNA-binding peptides using mRNA-peptide fusions. (2001) *Methods* **23**, 287-293.

15. Xia, T., Frankel, A., Takahashi, T.T., Ren, J. and Roberts, R.W. Context and conformation dictate function of a transcription antitermination switch. (2003) *Nat Struct Biol* **10**, 812-819.

16. Takahashi, T.T., Austin, R.J. and Roberts, R.W. mRNA display: ligand discovery, interaction analysis and beyond. (2003) *Trends Biochem Sci* **28**, 159-165.

17. Burd, C.G. and Dreyfuss, G. Conserved structures and diversity of functions of RNA-binding proteins. (1994) *Science* **265**, 615-621.

18. Futaki, S., Suzuki, T., Ohashi, W., Yagami, T., Tanaka, S., Ueda, K. and Sugiura, Y. Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. (2001) *J Biol Chem* **276**, 5836-5840.

19. Wender, P.A., Mitchell, D.J., Pattabiraman, K., Pelkey, E.T., Steinman, L. and Rothbard, J.B. The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. (2000) *Proc Natl Acad Sci U S A* **97**, 13003-13008.

20. Williams, K.P., Liu, X.H., Schumacher, T.N., Lin, H.Y., Ausiello, D.A., Kim, P.S. and Bartel, D.P. Bioactive and nuclease-resistant L-DNA ligand of vasopressin. (1997) *Proc Natl Acad Sci U S A* **94**, 11285-11290.

21. Schumacher, T.N., Mayr, L.M., Minor, D.L., Jr., Milhollen, M.A., Burgess, M.W. and Kim, P.S. Identification of D-peptide ligands through mirror-image phage display. (1996) *Science* **271**, 1854-1857.

22. Eckert, D.M., Malashkevich, V.N., Hong, L.H., Carr, P.A. and Kim, P.S. Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. (1999) *Cell* **99**, 103-115.

23. Antal, M., Boros, E., Solymosy, F. and Kiss, T. Analysis of the structure of human telomerase RNA *in vivo*. (2002) *Nucleic Acids Res* **30**, 912-920.

24. Chen, J.L., Blasco, M.A. and Greider, C.W. Secondary structure of vertebrate telomerase RNA. (2000) *Cell* **100**, 503-514.

25. Chen, J.L., Opperman, K.K. and Greider, C.W. A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. (2002) *Nucleic Acids Res* **30**, 592-597.

26. White, L.K., Wright, W.E. and Shay, J.W. Telomerase inhibitors. (2001) *Trends Biotechnol* **19**, 114-120.

27. Lukowiak, A.A., Narayanan, A., Li, Z.H., Terns, R.M. and Terns, M.P. The snoRNA domain of vertebrate telomerase RNA functions to localize the RNA within the nucleus. (2001) *Rna* **7**, 1833-1844.

28. Allain, F.H. and Varani, G. Structure of the P1 helix from group I self-splicing introns. (1995) *J Mol Biol* **250**, 333-353.

29. Leeper, T., Leulliot, N. and Varani, G. The solution structure of an essential stem-loop of human telomerase RNA. (2003) *Nucleic Acids Res* **31**, 2614-2621.

30. Ueda, C.T. and Roberts, R.W. Analysis of a long-range interaction between conserved domains of human telomerase RNA. (2004) *Rna* **10**, 139-147.

31. Mitchell, J.R. and Collins, K. Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. (2000) *Mol Cell* **6**, 361-371.

32. Rist, M. and Marino, J. Association of an RNA kissing complex analyzed using 2-aminopurine fluorescence. (2001) *Nucleic Acids Res* **29**, 2401-2408.

33. Beattie, T.L., Zhou, W., Robinson, M.O. and Harrington, L. Functional multimerization of the human telomerase reverse transcriptase. (2001) *Mol Cell Biol* **21**, 6151-6160.

34. Milligan, J.F. and Uhlenbeck, O.C. Synthesis of small RNAs using T7 RNA polymerase. (1989) *Methods Enzymol* **180**, 51-62.

35. Tan, R. and Frankel, A.D. Structural variety of arginine-rich RNA-binding peptides. (1995) *Proc Natl Acad Sci U S A* **92**, 5282-5286.

36. Rost, B. and Sander, C. Prediction of protein secondary structure at better than 70% accuracy. (1993) *J Mol Biol* **232**, 584-599.

37. Battiste, J.L., Mao, H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D. and Williamson, J.R. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. (1996) *Science* **273**, 1547-1551.

38. Puglisi, J.D., Chen, L., Blanchard, S. and Frankel, A.D. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. (1995) *Science* **270**, 1200-1203.

39. Scharpf, M., Sticht, H., Schweimer, K., Boehm, M., Hoffmann, S. and Rosch, P. Antitermination in bacteriophage lambda. The structure of the N36 peptide-*boxB* RNA complex. (2000) *Eur J Biochem* **267**, 2397-2408.

40. Legault, P., Li, J., Mogridge, J., Kay, L.E. and Greenblatt, J. NMR structure of the bacteriophage lambda N peptide/*boxB* RNA complex: recognition of a GNRA fold by an arginine-rich motif. (1998) *Cell* **93**, 289-299.

41. Cilley, C.D. and Williamson, J.R. Structural mimicry in the phage phi21 N peptide-*boxB* RNA complex. (2003) *Rna* **9**, 663-676.

42. Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A. and Patel, D.J. Solution structure of P22 transcriptional antitermination N peptide-*boxB* RNA complex. (1998) *Nat Struct Biol* **5**, 203-212.

43. Faber, C., Scharpf, M., Becker, T., Sticht, H. and Rosch, P. The structure of the coliphage HK022 Nun protein-lambda-phage *boxB* RNA complex. Implications for the mechanism of transcription termination. (2001) *J Biol Chem* **276**, 32064-32070.

44. Mascotti, D.P. and Lohman, T.M. Thermodynamics of oligoarginines binding to RNA and DNA. (1997) *Biochemistry* **36**, 7272-7279.

45. Mascotti, D.P. and Lohman, T.M. Thermodynamics of single-stranded RNA binding to oligolysines containing tryptophan. (1992) *Biochemistry* **31**, 8932-8946.

46. Mascotti, D.P. and Lohman, T.M. Thermodynamics of single-stranded RNA and DNA interactions with oligolysines containing tryptophan. Effects of base composition. (1993) *Biochemistry* **32**, 10568-10579.

47. Draper, D.E. Protein-RNA recognition. (1995) *Annu Rev Biochem* **64**, 593-620.

48. Draper, D.E. Themes in RNA-protein recognition. (1999) *J Mol Biol* **293**, 255-270.

49. Jackson, R.J. and Hunt, T. Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA. (1983) *Methods Enzymol* **96**, 50-74.

50. Brody, J.R. and Kern, S.E. Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. (2004) *Biotechniques* **36**, 214-216.

Chapter 7

In Vitro Evolution of RNA Binding Peptides for Increased Specificity

Abstract

The Hepatitis C Virus internal ribosme entry site (HCV IRES) is an attractive target for inhibition of HCV translation. Using peptides that we selected to bind P6.1, a domain of human telomerase RNA, we perform mutagenic PCR and additional cycles of selection to isolate sequences that bind to P6.1 and to domain IIId of the HCV IRES. The P6.1-binding sequences are highly specific for their target but the HCV IIId-binding sequences bind to both P6.1 and domain IIId. Our data show that although only a few positions are necessary for RNA binding specificity, these mutations are relatively rare.

Introduction

Hepatitis C Virus (HCV) infects nearly 170 million people worldwide, of which 75-85% are chronically infected (1). HCV can cause liver cirrhosis and hepatocellular carcinoma, however most individuals are asymptomatic. Pegylated interferon and ribavirin are currently used to treat HCV, but are only 40-80% effective in all patients (2). Like the human immunodeficiency virus (HIV), HCV mutates rapidly and becomes resistant to drugs targeting viral enzymes (2).

The 5' end of the HCV genome, however, is relatively well conserved among isolated strains of HCV (3). The 5' untranslated region (UTR) folds into an internal ribosome entry site (IRES) that allows cap-independent translation of the HCV genome. The HCV IRES binds to the 40S ribosomal subunit (K_d= 2 nM) as well as the initiation factor eIF3 (K_d= 35 nM) (4). The high conservation seen in the IRES region is most likely due to the fact that it needs to interact with these cellular components.

Because of its high conservation, the HCV IRES, represents a good target for inhibition of HCV. The secondary structure of HCV has been determined (Figure 7.1) and contains four conserved domains. Domains II, IIIc-f, and IV interact with the 40S subunit (5-7) while domains IIIa and IIIb interact with eIF3 (8, 9). These domains have been targeted using ribozymes, RNA aptamers, and antisense oligos (10) and most inhibit IRES-mediated translation *in vitro*, validating the HCV IRES as a potential drug target.

Using mRNA display, we have previously selected RNA binding peptides that bind with high affinity and specificity (11-13). mRNA display enables peptide selection
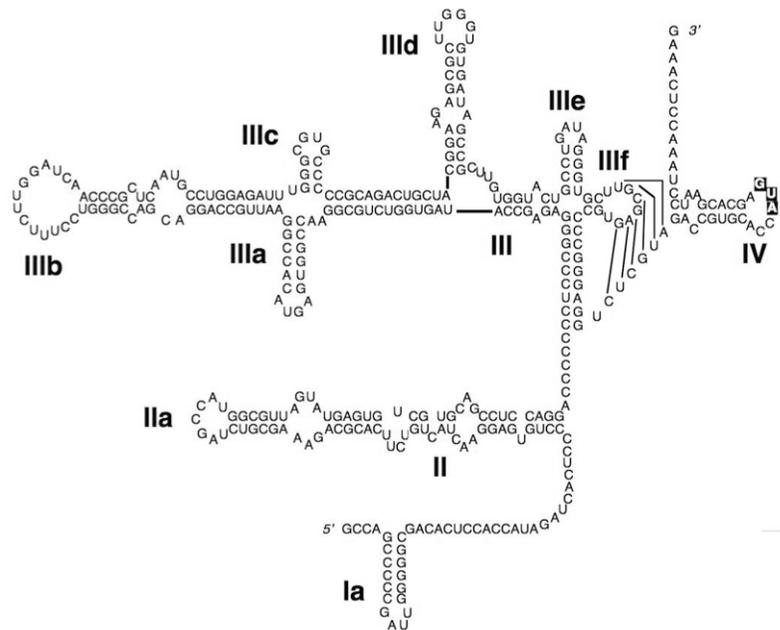
Figure 7.1. Secondary structure of the Hepatitis C Virus internal ribosome entry site (HCV IRES). Domains I-IV are indicated. Translation starts at the AUG codon (boxed) in domain IV. eIF3 binds to domains IIIa and IIIb while the 40S ribosomal subunit binds to domains II, IIIc-f, and IV.
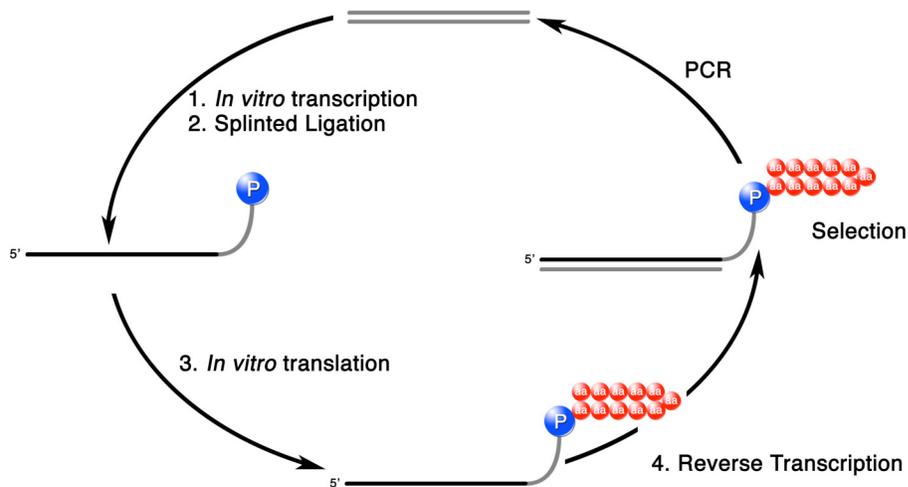
Figure 7.2. An mRNA display selection cycle. A double stranded DNA library is transcribed using T7 RNA polymerase [1] , and the resulting mRNA ligated to a synthetic oligonucleotide containing puromycin [2]. *In vitro* translation of the ligated product results in attachment of a peptide to its encoding mRNA [3]. Reverse transcription generates a cDNA/mRNA hybrid [4], which is used in affinity selection. PCR generates an enriched pool which is used in further cycles of selection.

by covalently linking a peptide to its encoding mRNA (Figure 7.2) (14). We have recently isolated peptides that bind to a catalytically important hairpin from human telomerase RNA, L-P6.1. These peptides were isolated using a reflection selection strategy whereby L-peptides that bind to an L-RNA are first isolated, enabling the synthesis of D-peptides that recognize natural RNA (15, 16). These D-peptides should be more stable against degradation by proteases (17). Thus, we believed that we would be able to use a similar strategy to target domains of the HCV IRES and obtain D-peptides that would inhibit IRES-mediated translation.

We noticed that the HCV IIId hairpin is very similar in sequence to the P6.1 hairpin from telomerase (Figure 7.3). The IIId hairpin contains the same loop sequence, 5'-UUGGG-3', except that it contains an extra 3' U nucleotide (U18, Figure 7.3) that is flipped out of the loop (18, 19). Other RNAs have been observed to extrude loop nucleotides when folding. For example, the *boxB* hairpin contains a pentaloop where the
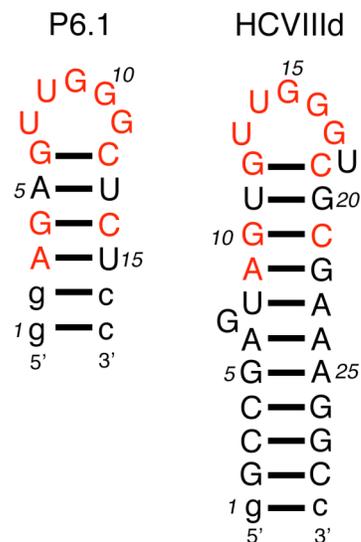
Figure 7.3. RNA models of P6.1 (left) and HCV IIId (right). Nucleotides that are common to both hairpins are colored red while nucleotides added for *in vitro* transcription are shown in lower case.

fourth nucleotide is extruded, resulting in a stable GNRA tetraloop fold (20, 21). Both P6.1 and IIId contain a GU wobble pair at the base of the loop, a GC closing pair, and several common nucleotides in the stem.

The P6.1 hairpin adopts a fold similar to that of UNCG tetraloops, with G10 (Figure 7.3) extruded (22). The HCV IIId hairpin does not exhibit the same conformation as the P6.1 hairpin, however the IIId loop is disordered (18), possibly suggesting structural homogeneity. It is therefore possible that the IIId loop may adopt conformations similar to L-P6.1 (23); upon peptide binding, a different structure would be enforced (24-26).

We were therefore interested to test if the peptides we isolated from the L-P6.1 selection would bind to an L-HCV IIId RNA. If the peptides possessed moderate affinity for L-HCV IIId, we felt that we would be able to then isolate peptides which were specific for either L-HCV IIId or L-P6.1 based on the fact that previous work has shown that RNA binding specificity is determined by a relatively few (2-4) number of mutations (27).

Results and Discussion

We first tested the ability of the pool targeting L-P6.1 to bind to L-HCV IIId. ^{35}S-labeled fusions from round 16 were tested in an *in vitro* binding assay using biotinylated L-HCV IIId as a target. Figure 7.4 shows that the pool binds to both L-P6.1 and to L-HCV IIId, although it prefers the L-P6.1 by a factor of ~1.5-fold.
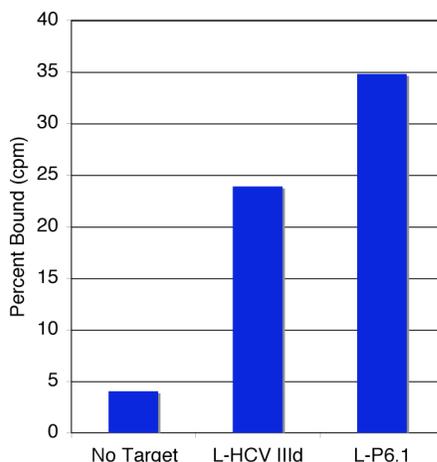
Figure 7.4. Binding of the round 16 pool to L-P6.1 and L-HCV IIId. A sample where target was omitted (no target) is shown as a control.

Previous data (Chapter 6) showed that the pool was composed of three sequence families, which possessed different affinities toward L-P6.1. This suggested that the individual peptide families utilized different binding contacts and it was therefore possible that individual clones would discriminate between L-P6.1 and L-HCV IIId differently. Figure 7.5 shows that representative clones from the L-P6.1 do indeed possess different specificities toward the L-P6.1 and L-IIId targets. Clone 18C1 is highly specific for the L-P6.1 target while 18C17 shows virtually no specificity between the two hairpins. 18C5 shows specificity representative of the pool as a whole; it prefers the L-P6.1 target by approximately twofold.

Improving RNA Binding Specificity

We have previously shown that there are "hot spots" responsible for specificity in RNA binding peptides; as few as four mutations were required to change the binding specificity between the λ N and P22 N peptides (24, 27). We therefore felt that it might
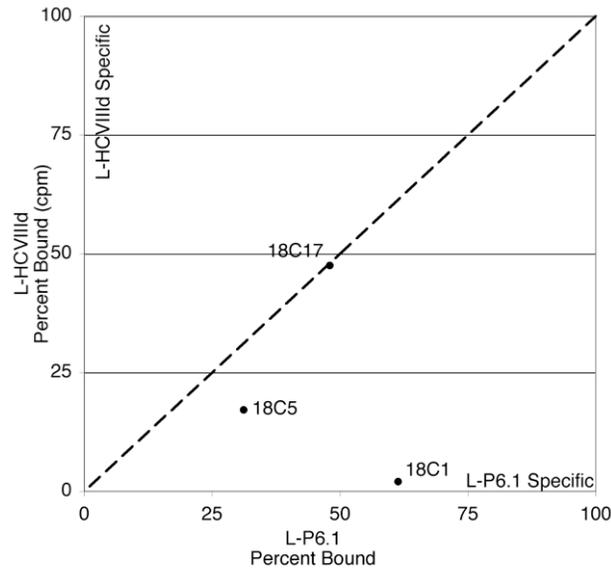
Figure 7.5. Specificities of the individual sequences from the round 18 pool targeting L-P6.1. Individual clones were synthesized as ^{35}S-labeled fusions and tested for binding to immobilized L-HCV IIId and L-P6.1. The x-axis represents L-P6.1 binding while the y-axis represents L-HCV IIId binding. Therefore, peptides in right, lower corner of the graph possess high specificity for the L-P6.1 target. The dotted line shows 1:1 binding.

 be possible to isolate specific peptides for each RNA target by introducing mutations in the round 18 pool, then performing additional round of selection on each RNA target.

Mutagenic PCR was performed on the round 18 pool and the resulting molecules sieved against immobilized L-P6.1 or L-HCV IIId. We chose to mutagenize the pool, rather than individual sequences (e.g., mutagenizing 18C1 for the L-P6.1 selection and 18C17 for the L-HCV IIId selection), since it was possible that other functional sequences could have been present at low copy numbers (28). This was more important for the HCV IIId selection since the selective pressure on the pool was changed; weak, low copy L-P6.1 binding sequences could have higher affinity for L-HCV IIId.

A total of three rounds of mutagenic PCR and selection were performed, followed by three more rounds of selection with no mutagenic PCR. Pool binding to L-P6.1 and

L-HCV IIId was monitored by measuring the binding of ^{35}S-labeled fusions against each immobilized RNA target. Figure 7.6 shows the binding of the L-P6.1 selection while Figure 7.7 shows the binding of the L-HCV IIId selection. Both pools showed an initial decrease in binding during rounds that contained mutagenic PCR, then an increase in binding in the subsequent rounds.

As the selection progressed, each pool showed increased specificity for its cognate target, although the L-HCV IIId pool never increased its specificity much more than 1:1 (Figure 7.8). The selection stringency for specificity was increased by addition of a negative selection step where the pool was precleared with immobilized noncognate target (e.g., the L-P6.1 selection was bound to immobilized L-HCV IIId before positive
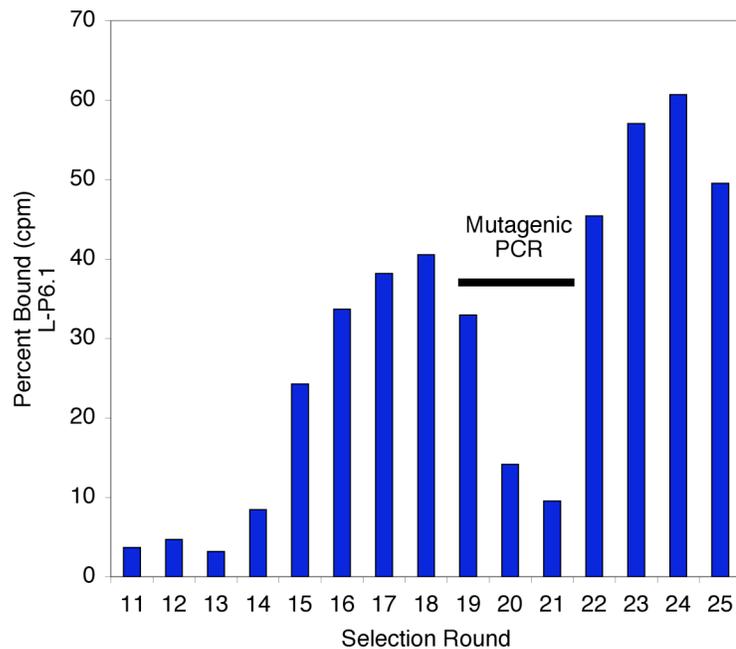
Figure 7.6. Binding of rounds 11-25 of the L-P6.1 selection. Mutagenic PCR was introduced in rounds 19, 20, and 21 resulting in a decrease in pool binding.
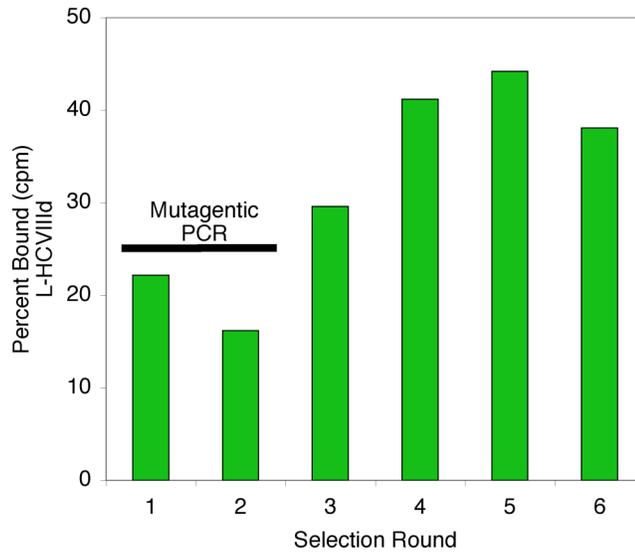
Figure 7.7. Binding of rounds 1-6 of the L-HCV IIId selection. Round 19 from the L-P6.1 selection was subjected to mutagenic PCR, then selected against immobilized L-HCV IIId. Mutagenic PCR was also added to rounds 1 and 2 of the L-HCV selection.
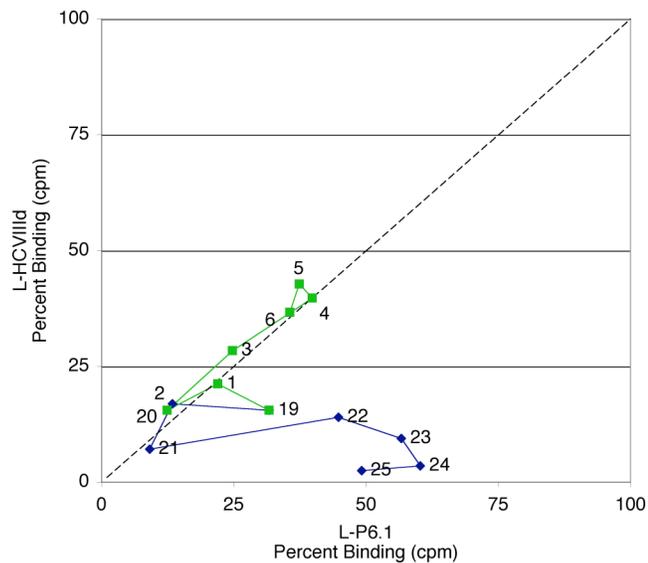
Figure 7.8. The specificity of rounds 0-6 of the L-HCV IIId selection (green) and rounds 19-25 of the L-P6.1 selection (blue). The dotted line represents equal affinities to L-P6.1 and L-HCV IIId. The x-axis represents L-P6.1 binding while the y-axis represents L-HCV IIId binding.

selection against L-P6.1). Additional rounds of selection resulted in no increase in

binding and round 24 from the L-P6.1 selection and round five from the L-HCV IIId

selection were sequenced.

P6.1 Binding Peptides

Figure 7.9 shows the sequences from the round 24 pool that were selected against

L-P6.1. A single family of sequences descended from the 18C1 family dominated the

pool; no sequences from the 18C5 family or the 18C17 family were seen. The 18C1

family showed the highest affinity and highest specificity for L-P6.1, explaining why it

overtook the round 24 pool.

```
Clone              Peptide Sequence

                   1         10        20        30
                   •         •         •         •
Parental Sequence  MMDWKRAKLNRLSVRKLRKYADYFTSGGLRASAI

            24C1   MMVRKRTKLSRLSVRKMRKYADYFSGSGLRASAI
            24C4   MSDYKRAKLNRLSVRKLRKYADYFTRVGLRASAI
            24C5   -MDWKRSKLNRLSARKLRKYADYFTSSGLRASAI
            24C6   MMDWKRTKLSRLSVRKLRKYADYFTGVGLRASAI
            24C7   MKDWKRTKFYRLSDRKLRKYSDYFTSVGLRASAI
            24C9   MMDWKRSKLNRLSVRKLRKYADYFARGGLRASAI
           24C10   MSDSKRAKLNRISARKTRKYSDYFSRVGLRASAI

       Consensus   M-D-KR-KL-RLS-RK-RKY-DYF---GLRASAI
```

Figure 7.9. Sequences from the round 24 pool targeting L-P6.1. The parental 18C1 is
shown above as a reference. Mutations are highlighted in bold text. The 3' primer codes
for the amino acid sequence GLRASAI, thereby preventing any mutations from occuring
in those positions. A consensus sequence is derived from the peptide sequences.

Assuming that the mutagenic PCR introduced mutations at every position in the

template, alignment of the round 24 clones could be used to show positions important for

RNA binding. A consensus sequence of DxKRxKLxRLSxRKxRKYxDYF was generated from the alignment. This sequence explains why attempts to truncate the peptide at position 21 resulted in nonbinding peptides (Chapter 6). No mutations to D22, Y23, or F24 were seen, implying that these positions are important for peptide binding. Additionally, all basic residues (Arg and Lys) are conserved, consistent with the fact that basic charge is important for RNA recognition by peptides.

We then tested the specificity of individual clones toward L-P6.1 and L-HCV IIId using the *in vitro* binding assay (Figure 7.10). Although it was difficult to obtain accurate binding affinities for most of the peptides because of low binding affinities to L-HCV IIId (typically, less than 3%), most of the 24C sequences exhibited similar preferences for L-P6.1 as the parental 18C1 sequence (Figure 7.10B). However, two clones, 24C7 and 24C10, showed an increase in specificity for L-P6.1 of 3-5 fold, as compared to the parental sequence (Figure 7.10B). Both 24C7 and 24C10 contain the A21S mutation, while all other sequences do not, suggesting that this position is a source for the increased specificity exhibited by these peptides. In order to test this hypothesis, we constructed an 18C1 mutant containing the A21S mutation and tested its specificity toward the two hairpins. The single A21S mutation increases the specificity for L-P6.1 from 30-fold to 60-fold (data not shown), showing that it contributes, but is not the sole determinant, of specificity.
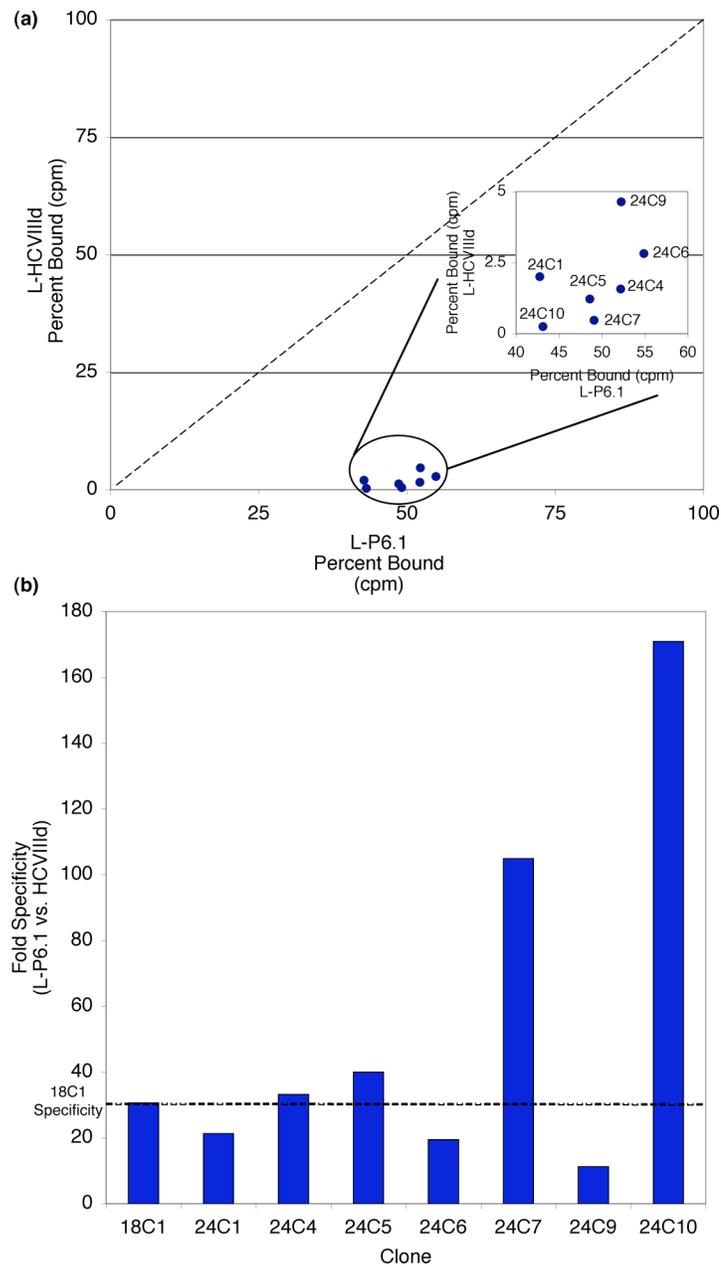
Figure 7.10. Specificity of individual round 24 clones. **(a)** The binding of the clones is magnified for clarity (inset). **(b)** Fold specificity of L-P6.1 binding versus L-HCV IIId binding. Higher bars represent more specific L-P6.1 sequences. The specificity of the 18C1 parental sequence is shown by the dotted line.

HCV IIId Binding Peptides

The round 5 pool targeting L-HCV IIId was sequenced and yielded three families of sequences (Figure 7.11). As expected, descendents from the 18C17 sequence, which had shown little selectivity for L-P6.1 or L-HCV IIId, were present, and the 18C1 and 18C5 sequences, which preferred L-P6.1, were absent. Two new families not seen in the original 18C pool were also isolated, supporting our decision to randomize the entire pool rather than the 18C17 sequence. Little sequence similarity is seen between the three families, preventing any consensus sequence from being deduced.

```
Clone              Peptide Sequence

                   1         10        20        30
                   •         •         •         •
Parental Sequence  MKHSNSSRGRKTLWRALTLWLLMQSLKRTSGGLRASAI

        HCV5-2     MKLPNSSRGRKTLWRAMTLWLLMQSLQRTSGGLRASAI
        HCV5-9     MKYSNSSRGRKTLWRALTLWLLMQSLKRTSDGLRASAI
        HCV5-10    MKYSNSSRGRKTLWRGWTLWILIQSLKRTRDGLRASAI


        HCV5-1     MNTLKELVLLYLSERRGLSVSKADFLKRTSDGLRASAI
        HCV5-5     MNSLKELVLLYLRERRGLPVSKPDFAKWVKDGLRASAI
        HCV5-8     MNTLKELVLLYLSEQRGLPISKSDFLRWTRGGLRASAI


        HCV5-7     MMKRFSAKVSTLSRERRRMLRTLIQRRLTGGGLRASAI
```

Figure 7.11. Sequences from the round 5 pool targeting L-HCV IIId. The parental sequence (18C17) for HCV5-2, 9, and 10 is shown above as a reference. Peptides are grouped into three sequence families. Muations are highlighted in bold text.

Additionally, the sequences contain fewer mutations per template when compared to the round 24 sequences from the L-P6.1 selection. An average of ~5.5 mutations per template were seen round 24 of the L-P6.1 selection while only ~3.5

mutations were template occurred in the L-HCV IIId selection. Although a difference in error rates could be a possible explanation, the fact that both mutagenesis experiments were performed in parallel under the same experimental conditions argues against this hypothesis. Instead, a more plausible explanation is that more positions are required for binding for the sequence families from the L-HCV IIId selection than from the L-P6.1 selection.

Figure 7.12 shows the binding specificity of the round five clones as determined by the *in vitro* binding assay. Two clones, HCV5-1 and HCV5-7 bind the L-HCV IIId
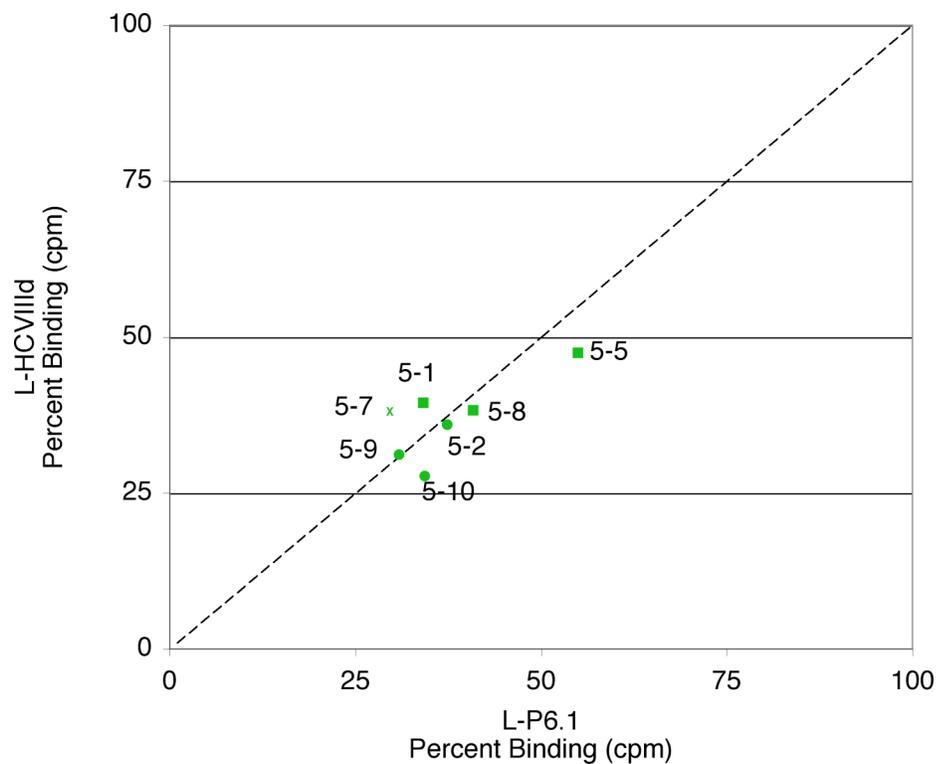
Figure 7.12. Specificity of individual round 5 clones targeting L-HCV IIId. Peptides belong to three sequence families (Figure 7.11) and families 1, 2, and 3 are denoted by circles, squares, and crosses, respectively.

hairpin with higher affinity than the L-P6.1 hairpin. Even these clones, however, lack the ability to effectively discriminate between the two hairpins; the entire pool exhibits little specificity for either hairpin.

It is worth noting that the two sequences that bind L-HCV IIId better than L-P6.1, HCV5-1 and HCV5-7, belong to peptide families that were not seen in round 18. It therefore must be much easier to enrich for a novel sequence family than it is to endow a degenerate sequence with more specificity; there are more unique sequences in the pool than there are multiple mutations conferring specificity.

Conclusions

Starting from a pool of sequences that bound L-P6.1, we have isolated sequences that bind to L-P6.1 and L-HCV IIId. We attempted to increase the selectivity of the peptides for each target by including mutagenic PCR in the selection cycle, however we were only successful in isolating L-P6.1-specific sequences.

Although only a few (2-4 mutations) mutations can result in specificity changes, these mutations must be relatively rare. For example, in order to search all possible combinations of twenty amino acids at two positions in a sequence of n length, there will be $200 \times (n^2 - n)$ sequences. Therefore, a total of ~1.4x10^5 two-position mutations are possible for the X27S library. Additionally, there are other factors that make it more difficult to search sequence space for highly specific mutants. Mutagenic PCR is heavily biased towards certain mutations, and the degeneracy of the genetic code prevents many mutations from occurring by a one-nucleotide change.

These data suggest that there are two possible alternatives to isolate more specific L-HCV binding peptides. First, returning to an earlier round (e.g., round 12) and reselecting against the L-HCV target may yield more unique peptides with higher affinities toward L-HCV IIId. Such a strategy also has the advantage of not requiring negative selection steps to remove L-P6.1 binding sequences. Secondly, doping several of the round 5 HCV IIId-binding sequences and reselecting against immobilized L-HV IIId would be a more effective strategy than utilizing mutagenic PCR in the selection cycle. Our data argue that mutations increasing specificity are relatively rare, resulting in the need for higher mutagenesis rates and more cycles of selection. Doping the sequences will also result in less mutational bias and access to all twenty amino acids at a single position, rather than mutations resulting from a one-base change. Future experiments will show if it is possible to isolate peptides with increased specificity.

Materials and Methods

In Vitro Selection

The selection was performed as described in Chapter 6, for the L-P6.1 selection rounds (pg. 149).

Mutagenic PCR

PCR conditions were the same as the selection rounds, except the dNTP concentrations were 1 mM of dCTP and dTTP and 0.2 mM of dATP and dGTP, the $MgCl_2$ concentration was increased to 7 mM, and $MnCl_2$ was added to a final concentration of 0.5 mM. Samples were cycled for 1 minute at 94 $^{\circ}$C, 1 minute at 55 $^{\circ}$C, and 2 minutes at 72 $^{\circ}$C. After every four cycles, 10 µL of PCR reaction was transferred to 90 µL of fresh buffer, dNTPs, primers and Taq heated to 72 $^{\circ}$C. This process was repeated for up to 15 transfers. Samples from different cycles were pooled and used for the next round of selection. The mutagenic rates for the first, second, and third mutagenic PCR reactions were an average of 0.031 mutations/codon (1 amino acid change per template), 0.055 mutations/codon (1.8 aa changes/template), and 0.057 mutations/codon (1.9 aa changes/template).

In Vitro Binding Assay

dT-purified ^{35}S-labeled fusions (~100,000 cpm) were incubated with 0.1-0.2 µM immobilized, annealed RNA target (50 µL of a 50/50 (v/v) slurry) in binding buffer with 50 µg/mL tRNA for one hour at 4 $^{\circ}$C. The samples were transferred to Spin-X columns

(Costar) and washed three times with 700 μL of binding buffer. The supernatant, washes, and beads counted by scintillation.

RNA Synthesis

Biotinylated L-HCV IIId RNA (5-GGCCGAGUAGUGUUGGGUCGCGAAAGGC CAAA-biotin-3') was synthesized and deprotected by Chemgenes. RNA was gel purified by 20% urea-PAGE, electroeluted, and ethanol precipitated.

Sequencing

Pool templates were PCR amplified, gel purified using the QiaQuick gel extraction kit (Qiagen), and cloned using a TOPO-TA cloning kit (Invitrogen). Single colonies were grown in LB broth containing 50-100 μg/mL Ampicillin, and purified using the QiaPrep Spin Miniprep kit (Qiagen). Sequences were analyzed by Sequencher.

References

1.   Cohen, J. The scientific challenge of hepatitis C. (1999) *Science* **285**, 26-30.

2.   Tan, S.L., He, Y., Huang, Y. and Gale, M., Jr. Strategies for hepatitis C therapeutic intervention: now and next. (2004) *Curr Opin Pharmacol* **4**, 465-470.

3.   Honda, M., Beard, M.R., Ping, L.H. and Lemon, S.M. A phylogenetically conserved stem-loop structure at the 5' border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation. (1999) *J Virol* **73**, 1165-1174.

4.   Kieft, J.S., Zhou, K., Jubin, R. and Doudna, J.A. Mechanism of ribosome recruitment by hepatitis C IRES RNA. (2001) *Rna* **7**, 194-206.

5.   Gallego, J. and Varani, G. The hepatitis C virus internal ribosome-entry site: a new target for antiviral research. (2002) *Biochem Soc Trans* **30**, 140-145.

6.   Lytle, J.R., Wu, L. and Robertson, H.D. Domains on the Hepatitis C virus internal ribosome entry site for 40s subunit binding. (2002) *Rna* **8**, 1045-1055.

7.   Spahn, C.M., Kieft, J.S., Grassucci, R.A., Penczek, P.A., Zhou, K., Doudna, J.A. and Frank, J. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. (2001) *Science* **291**, 1959-1962.

8.   Collier, A.J., Gallego, J., Klinck, R., Cole, P.T., Harris, S.J., Harrison, G.P., Aboul-Ela, F., Varani, G. and Walker, S. A conserved RNA structure within the HCV IRES eIF3-binding site. (2002) *Nat Struct Biol* **9**, 375-380.

9. Kieft, J.S., Zhou, K., Grech, A., Jubin, R. and Doudna, J.A. Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation. (2002) *Nat Struct Biol* **9**, 370-374.

10. Toulme, J.J., Darfeuille, F., Kolb, G., Chabas, S. and Staedel, C. Modulating viral gene expression by aptamers to RNA structures. (2003) *Biol Cell* **95**, 229-238.

11. Barrick, J.E., Takahashi, T.T., Ren, J., Xia, T. and Roberts, R.W. Large libraries reveal diverse solutions to an RNA recognition problem. (2001) *Proc Natl Acad Sci U S A* **98**, 12374-12378.

12. Barrick, J.E., Takahashi, T.T., Balakin, A. and Roberts, R.W. Selection of RNA-binding peptides using mRNA-peptide fusions. (2001) *Methods* **23**, 287-293.

13. Xia, T., Frankel, A., Takahashi, T.T., Ren, J. and Roberts, R.W. Context and conformation dictate function of a transcription antitermination switch. (2003) *Nat Struct Biol* **10**, 812-819.

14. Takahashi, T.T., Austin, R.J. and Roberts, R.W. mRNA display: ligand discovery, interaction analysis and beyond. (2003) *Trends Biochem Sci* **28**, 159-165.

15. Schumacher, T.N., Mayr, L.M., Minor, D.L., Jr., Milhollen, M.A., Burgess, M.W. and Kim, P.S. Identification of D-peptide ligands through mirror-image phage display. (1996) *Science* **271**, 1854-1857.

16. Williams, K.P., Liu, X.H., Schumacher, T.N., Lin, H.Y., Ausiello, D.A., Kim, P.S. and Bartel, D.P. Bioactive and nuclease-resistant L-DNA ligand of vasopressin. (1997) *Proc Natl Acad Sci U S A* **94**, 11285-11290.

17. Eckert, D.M., Malashkevich, V.N., Hong, L.H., Carr, P.A. and Kim, P.S. Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. (1999) *Cell* **99**, 103-115.

18. Lukavsky, P.J., Otto, G.A., Lancaster, A.M., Sarnow, P. and Puglisi, J.D. Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function. (2000) *Nat Struct Biol* **7**, 1105-1110.

19. Klinck, R., Westhof, E., Walker, S., Afshar, M., Collier, A. and Aboul-Ela, F. A potential RNA drug target in the hepatitis C virus internal ribosomal entry site. (2000) *Rna* **6**, 1423-1431.

20. Legault, P., Li, J., Mogridge, J., Kay, L.E. and Greenblatt, J. NMR structure of the bacteriophage lambda N peptide/*boxB* RNA complex: recognition of a GNRA fold by an arginine-rich motif. (1998) *Cell* **93**, 289-299.

21. Scharpf, M., Sticht, H., Schweimer, K., Boehm, M., Hoffmann, S. and Rosch, P. Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex. (2000) *Eur J Biochem* **267**, 2397-2408.

22. Leeper, T., Leulliot, N. and Varani, G. The solution structure of an essential stem-loop of human telomerase RNA. (2003) *Nucleic Acids Res* **31**, 2614-2621.

23. Leulliot, N. and Varani, G. Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. (2001) *Biochemistry* **40**, 7947-7956.

24. Austin, R.J., Xia, T., Ren, J., Takahashi, T.T. and Roberts, R.W. Differential modes of recognition in N peptide-*boxB* complexes. (2003) *Biochemistry* **42**, 14957-14967.

25. Menger, M., Eckstein, F. and Porschke, D. Dynamics of the RNA hairpin GNRA tetraloop. (2000) *Biochemistry* **39**, 4500-4507.

26. Barrick, J.E. and Roberts, R.W. Achieving specificity in selected and wild-type N peptide-RNA complexes: the importance of discrimination against noncognate RNA targets. (2003) *Biochemistry* **42**, 12998-13007.

27. Austin, R.J., Xia, T., Ren, J., Takahashi, T.T. and Roberts, R.W. Designed arginine-rich RNA-binding peptides with picomolar affinity. (2002) *J Am Chem Soc* **124**, 10966-10967.

28. Bartel, D.P. and Szostak, J.W. Isolation of new ribozymes from a large pool of random sequences. (1993) *Science* **261**, 1411-1418.