Chapter 3

Selection of RNA-Binding Peptides
Using mRNA-Peptide Fusions

This work has been adapted from the following publication:

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.
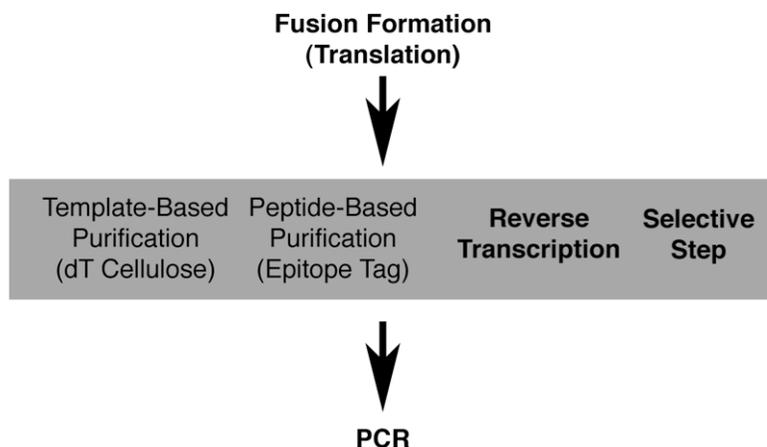
Fusion Formation
(Translation)

| Template-Based Purification (dT Cellulose) | Peptide-Based Purification (Epitope Tag) | **Reverse Transcription** | **Selective Step** |

PCR

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).
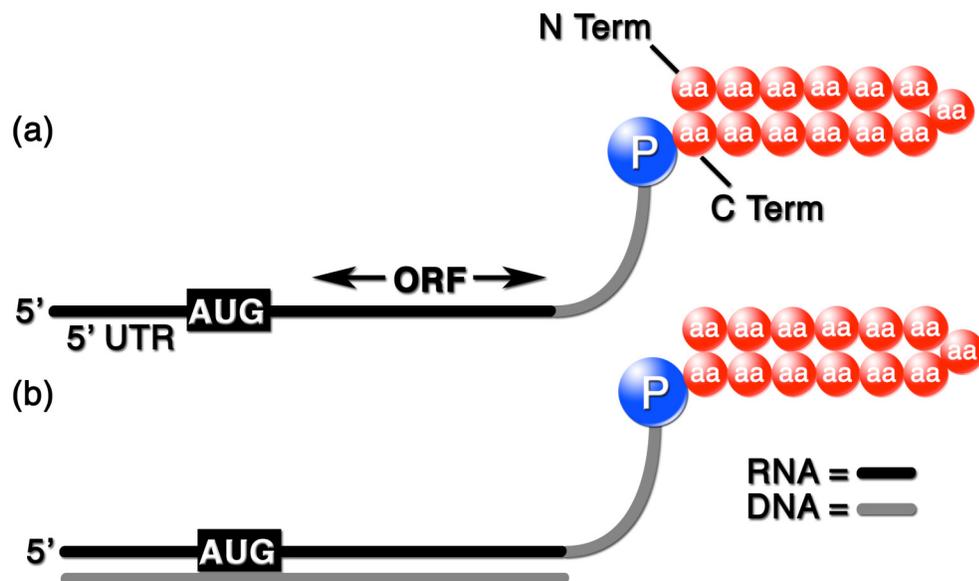

Figure 3.2. Two forms of mRNA-peptide fusions. **(a)** Schematic indicating the structure and connectivity of a mRNA-peptide fusion after synthesis on the ribosome. Linkage occurs between the C terminus of the peptide and the 3' end of the template through puromycin (P). **(b)** cDNA/mRNA-peptide fusion 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).
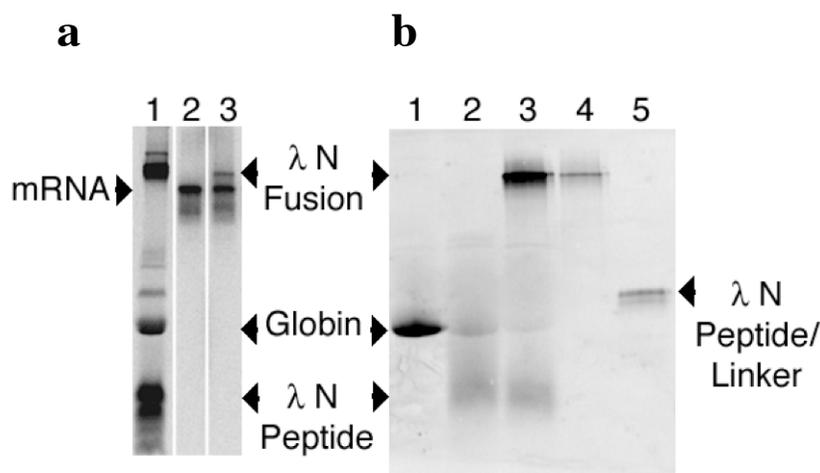

Figure 3.3. Efficiency of fusion synthesis. **(a)** The fraction of template converted to fusion. Translation of 400 nM ^{32}P-labeled template (lane 2) produces a fusion product with lower mobility (lane 3) as assayed by SDS-Tricine PAGE. Lane 1 shows ^{35}S-labeled fusion as a size standard. **(b)** 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
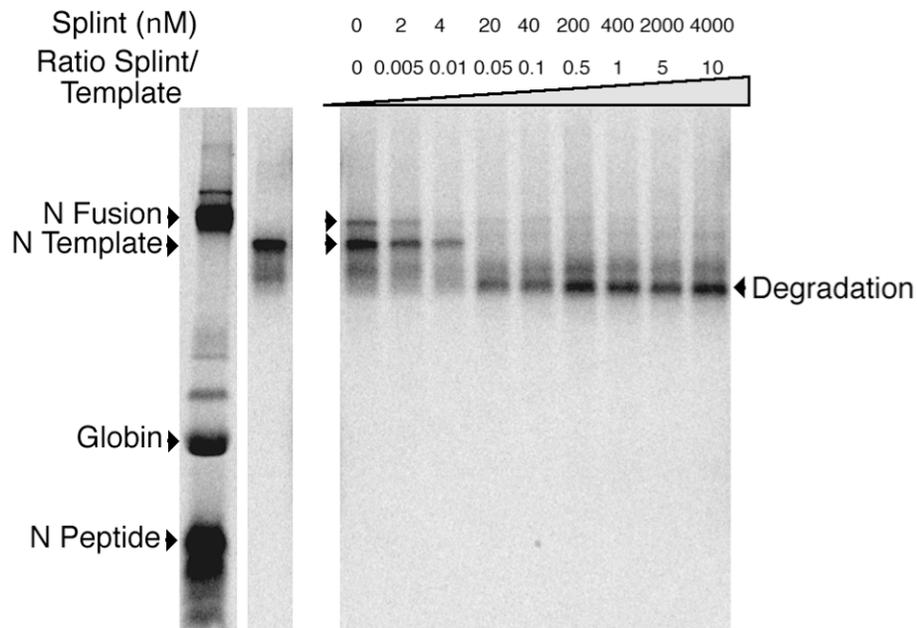

Figure 3.4. Splint-mediated degradation of template during translation. Varying amounts of splint oligonucleotide (see Materials and Methods) were added to translation reactions containing a ^{32}P-labeled template. The presence of sub-stoichiometric amounts of splint (0.05/1, splint to template) causes 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'-GGGACAATTACTATTTACAATTACAATGGA CGCCCAGACCCGCCGGCGCGAGCGCAGGGCCGAGAAGCAGGCCCAGTGGAA GGCCGCCAACGACTACAAGGACGACGATGACAAG-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'-GGGACAATTACTATTTACAATTACAATGGACGCCCAGACCNNCBNGCGCGAG CGCAGGGCCGAGAAGCAGGCCCAGTGGAAGGCCGCCAACGACTACAAGGAC GACGATGACAAG-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 (α-^{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$_2$O 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.