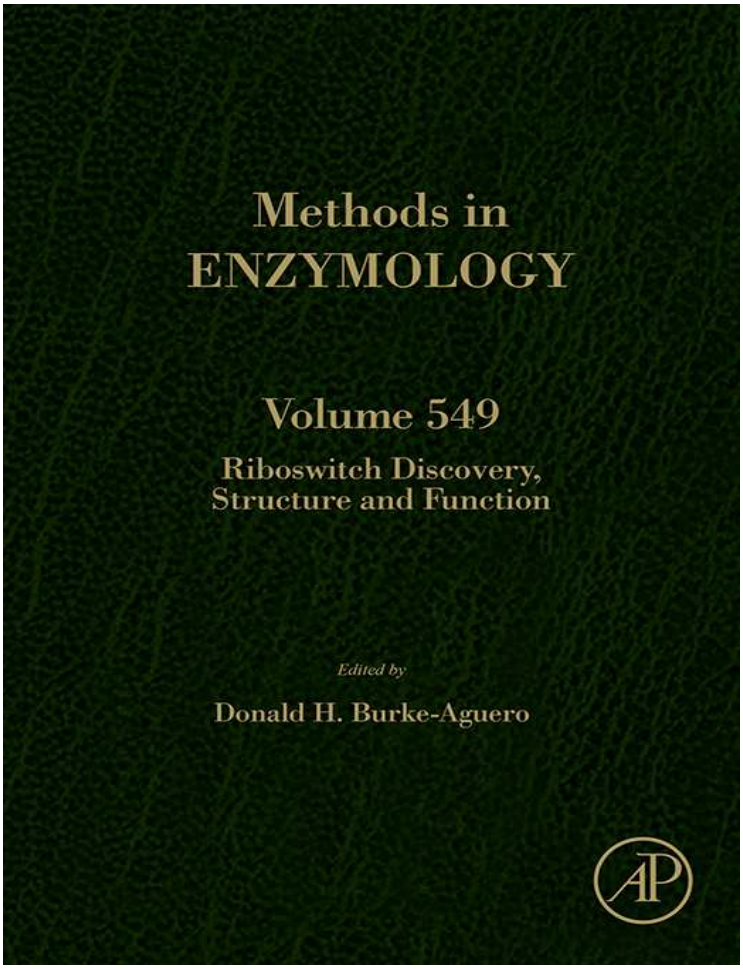

From Bao Ho, Julio Polanco, Randi Jimenez and Andrej Lupták, Discovering Human RNA Aptamers by Structure-Based Bioinformatics and Genome-Based *In Vitro* Selection. In: Donald H. Burke-Aguero, editor, *Methods in Enzymology, Vol. 549*, Burlington: Academic Press, 2014, pp. 29-46.
ISBN: 978-0-12-801122-5

Academic Press

CHAPTER TWO

Discovering Human RNA Aptamers by Structure-Based Bioinformatics and Genome-Based *In Vitro* Selection

Bao Ho[*,†,‡], Julio Polanco[*,†,‡], Randi Jimenez[*,†,‡], Andrej Lupták[*,†,‡,1]
[*]Department of Pharmaceutical Sciences, University of California, Irvine, California, USA
[†]Department of Molecular Biology and Biochemistry, University of California, Irvine, California, USA
[‡]Department of Chemistry, University of California, Irvine, California, USA
[1]Corresponding author: e-mail address: aluptak@uci.edu

Contents

Abstract

In vitro selection and structure-based searches have emerged as useful techniques for the discoveries of structurally complex RNAs with high affinity and specificity toward metabolites. Here, we focus on the design of a human genomic library that serves as the DNA template for *in vitro* selection of RNA aptamers. In addition, the structural solutions obtained from the *in vitro* selection can be used for structure-based searches for discovery of analogous aptamers in various genomic databases.

1. INTRODUCTION

Over the past two decades, *in vitro* selection (also known as SELEX) has served as a powerful tool for the discovery of novel DNA and RNA

Methods in Enzymology, Volume 549
ISSN 0076-6879
http://dx.doi.org/10.1016/B978-0-12-801122-5.00002-7

aptamers. Since then, extensive selection and structural studies have highlighted the structural diversity within a given random pool of DNA or RNA sequences (Stoltenburg, Reinemann, & Strehlitz, 2007). Initial *in vitro* selection studies utilized pools of synthetic random DNAs flanked by fixed, primer-binding sequences that were transcribed into amplifiable RNAs of similar diversity to determine the frequency of aptamers capable of binding a target molecule (Ellington & Szostak, 1990; Tuerk & Gold, 1990). More recently, modifications to the *in vitro* selection procedure have aimed to identify naturally occurring aptamers and other functional nucleic acids by using genome-derived DNA pools as templates for selections. In the case of adenosine, but not GTP, both synthetic and genomic DNA selections revealed a number of structurally conserved aptamer sequences (Burke & Gold, 1997; Curtis & Liu, 2013; Davis & Szostak, 2002; Sassanfar & Szostak, 1993; Vu et al., 2012). These adenosine-binding motifs are sequence-independent and represent a rare example of convergent molecular evolution spanning both genomic and synthetic sequence space.

Genomic SELEX was introduced by Singer and Gold using a genomic library for *in vitro* selection studies using a set of primers consisting of a fixed 5′ end sequence and a randomized 3′ tail. This allowed for amplification of fragmented human, yeast, and *Escherichia coli* genomic DNA *in vitro*, followed by size selection and primer extension to allow for transcription (Singer, Shtatland, Brown, & Gold, 1997). On the other hand, Salehi-Ashtiani et al. designed a genomic pool by partial digestion of human genomic DNA using DNase I. After digestion, hairpin sequences of known composition were ligated onto the genomic DNA, subjected to single-stranded digestion and then amplified by primer extension (Salehi-Ashtiani, Lupták, Litovchick, & Szostak, 2006).

Among the best-characterized aptamer structures is the adenosine-binding motif. Both synthetic and genomic selections reveal a conserved binding pocket consisting of an 11-nucleotide loop and a bulged G formed by two flanking helical motifs (Fig. 2.1). Nuclear magnetic resonance and mutation studies have shown that these conserved nucleotides and flanking helices are required for the formation of a binding pocket to allow base stacking and hydrogen-bonding interactions with the ligand (Dieckmann, Butcher, Sassanfar, Szostak, & Feigon, 1997; Dieckmann, Suzuki, Nakamura, & Feigon, 1996; Jiang, Kumar, Jones, & Patel, 1996; Vu et al., 2012). Although the sequence compositions of the flanking helical motifs vary, the adenosine-binding loop is largely sequence conserved, and both of these properties are exploited with structure-based search algorithms.

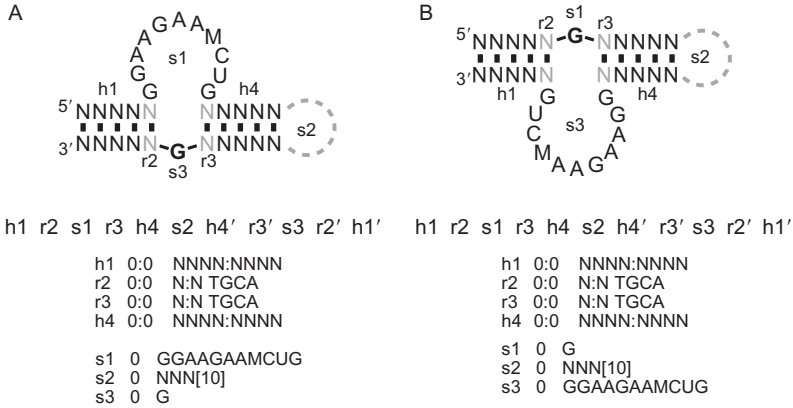

Figure 2.1 Secondary structure descriptors for an adenosine aptamer. The aptamer has been found to exist in both the 5′ (A) and 3′ (B) strands (Burgstaller & Famulok, 1994; Sassanfar & Szostak, 1993; Vu et al., 2012). The bulged guanosine nucleotide necessary for ligand binding is in bold.

Structure-based search algorithms are powerful tools in the discovery of functional RNAs. Their success comes from the ability to find sequences in unrelated, unprocessed sequence data that match complex motifs (Gautheret, Major, & Cedergren, 1990). Their appeal lies in the user accessibility: the ease of use, the flexibility in descriptor design, and the efficiency and speed of searches. Structure-based search programs are used to identify sequences capable of fitting into a given secondary structure. These programs match the patterns of base-paired and single-stranded regions as defined by the user in a descriptor file. Furthermore, the descriptor allows the user to specify regions of strict Watson-Crick base pairing, wobble pairs, mismatches, and single-nucleotide insertions in helices. Two user-friendly programs with similar syntax are RNABOB (ftp://selab.janelia.org/pub/software/rnabob/) and RNArobo (Jimenez, Rampasek, Brejova, Vinar, & Lupták, 2012); the implementation of neither of these programs requires extensive programming skills. The implementation of RNABOB is as previously described (Riccitelli & Lupták, 2010) and will be outlined briefly below.

Our approach here focuses on the design of a genomic DNA pool for use in *in vitro* selection. In principle, the pool can sample the entire genome of the target organism at single-nucleotide resolution (in both directions, with respect to the engineered RNA polymerase promoter), independent of expression of individual genes, but lacks sequences corresponding to spliced and otherwise processed transcripts. Structural characterization of the

resulting aptamers can be used to generate structure descriptors for mapping sequences against a genome database.

2. PRECAUTIONS

As ribonucleic acids are highly sensitive to degradation, it is advisable that the following procedures are conducted in an RNase-free laboratory environment. RNase-free reagents, consumables, and equipment are required. Proper handling of RNA includes, but is not limited to, frequent disinfection of gloves, benches, and instruments with ethanol or weak oxidizers (diluted bleach or hydrogen peroxide), regular change of gloves, and maintenance of RNA over ice while preparations of experiments take place. In addition, laboratory personnel should be properly equipped and trained for utilizing ^{32}P. Material safety data sheets of reagents included within these procedures are provided by their distributors and should be followed to minimize hazardous occurrences.

3. GENERATING A HUMAN GENOMIC DNA POOL

3.1. Materials

3.1.1 High molecular weight human genomic DNA

High molecular weight human genomic DNA isolated from whole blood cells is commercially available and can be purchased from suppliers such as Clontech or Promega. For the procedure outlined below, it is critical that the source DNA be of high molecular weight since the DNA will be subjected to sonication for fragmentation.

3.1.2 Adapter oligonucleotide sequences

The sequences of known composition consist of two pairs of synthetic oligonucleotides. The forward adapter contains a T7 promoter to allow for transcription, a 3′ dT overhang, and a 5′ phosphate modification on the complementary strand (Fig. 2.2). The reverse adapter contains a 5′ phosphate modification and a 3′ dT overhang on the complementary strand (Fig. 2.2). With regards to adapter design, the following requirements must be considered: oligonucleotide sequences should not form any interactions that may interfere with directional ligation of DNA, the melting temperature of each adapter set is within range of 55–70 °C for polymerase chain reaction (PCR) amplification and primers should not anneal to form primer dimers or nonspecific amplification byproducts. The sequences are chosen

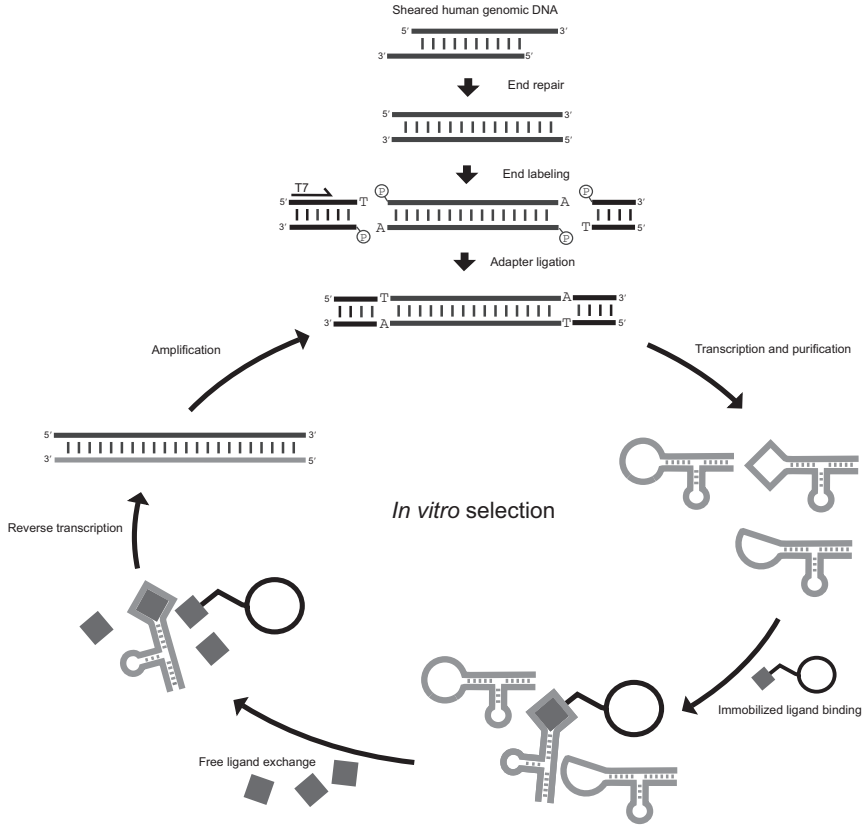

Figure 2.2 Overview of the design of a human genomic pool and *in vitro* selection. The library construction begins with sheared human genomic DNA, modified to allow ligation of sequences of known compositions. The library serves as a template for subsequent selection of RNA sequences with affinity and specificity toward a target molecule.

to promote efficient transcription and amplification, and if one of the goals of the *in vitro* selection is high-throughput sequencing (HTS), then the sequence is also matched to primer sequences used by the HTS platform of choice.

3.1.3 Enzymes

The following enzymes are used to prepare the genomic DNA for adapter ligation and PCR amplification: T4 DNA polymerase, T4 polynucleotide kinase (PNK), *E. coli* Klenow fragment (exo-), T4 DNA ligase, *Pfu* DNA Polymerase. All enzymes are commercially available from New England Biolabs and other suppliers.

3.1.4 Buffers

A solution 10 m*M* Tris·HCl, pH 8.0 can be used for the dilution, elution, or resuspension of the DNA throughout each step. For gel electrophoresis a 44.5 m*M* Tris·borate, pH 8.3, 1 m*M* EDTA solution is used as a running buffer. All gels are poststained in an ethidium bromide-water bath prior to imaging.

3.1.4.1 Tris/borate/EDTA buffer (10 ×)

Tris/borate/EDTA (TBE) buffer is widely utilized as the buffer for electrophoresis. The running concentration is typically 1 × or 0.5 ×. The following buffer is at 10 × and can be diluted with water to the appropriate concentration. Stock and running buffers can be stored at room temperature.

890 m*M* Tris·HCl, 890 m*M* boric acid, 20 m*M* EDTA

3.1.5 Instruments and miscellaneous

Covaris S2 focused acoustic shearer
Agarose and polyacrylamide gel electrophoresis (PAGE) loading equipment
PCR thermocycler (Eppendorf, BioRad)
Spectrophotometer (Nanodrop from Thermo scientific)
DNA purification kit (Qiagen)
UV light, short wavelength, with a camera mount
Image density analysis software (ImageJ)

3.2. Procedures

3.2.1 Preparation of genomic DNA

In a 0.5-mL microcentrifuge tube, aliquot 3 μg of high molecular weight human genomic DNA for a final volume of 130 μL in 10 m*M* Tris·HCl pH 8.0. Using the Covaris S2 acoustic shearer, tune the instrument to shear the genomic DNA to the size of interest. Note that shorter high frequency bursts correspond to larger fragment sizes. Verify the efficiency of fragmentation by running a small amount of the sheared DNA product on a 2% agarose gel along with a molecular weight standards. Sheared genomic DNA usually runs as a smeared band with the highest intensity corresponding to the median target size.

3.2.2 Repairing genomic DNA ends

Due to nonspecific physical shearing, the single-stranded overhangs generated by sonication are processed to generate double-stranded blunt ends.

To generate blunt-ended DNA products, a nucleotide polymerization reaction is performed using 500 μ*M* deoxynucleotide triphosphates, 50 m*M* NaCl, 10 m*M* Tris·HCl pH 8.0, 10 m*M* $MgCl_2$, 1 m*M* dithiotreitol (DTT), 3 units of T4 DNA polymerase, or other DNA polymerase with 3′ exonuclease activity, and 3 μg of sheared genomic DNA. The reaction is incubated at 12 °C for 15–30 min and subsequently purified using a DNA purification kit.

3.2.3 Addition of 5′ phosphate group onto genomic DNA

To prepare the genomic DNA for ligation, a 5′ phosphate is added enzymatically using T4 PNK. The following reagents are combined in a PCR tube and incubated at 37 °C overnight: 200 μ*M* adenosine triphosphate (ATP), 70 m*M* Tris·HCl pH 7.6, 10 m*M* $MgCl_2$, 5 m*M* DTT, 5% polyethylene glycol 8000, 2 m*M* spermidine, 10 units of T4 PNK, 3 μg of end-repaired genomic DNA product, and deionized water. Overnight incubation is significantly longer than is typically used for phosphorylation of oligonucleotides and is used here to ensure phosphorylation of long, blunt-ended DNAs. Once completed, the phosphorylated DNA product is purified using a silica membrane (Qiagen spin column) and quantified by spectrophotometry.

3.2.4 Addition of 3′ dA overhangs

To generate complementary cohesive ends to aid in adapter ligation, deoxyadenosine overhangs are introduced to the 3′ ends of the genomic DNA (Fig. 2.1). This reaction uses 100 μ*M* dATP, 10 m*M* Tris·HCl pH 8.0, 50 m*M* NaCl, 10 m*M* $MgCl_2$, 1 m*M* DTT, 5 units of Klenow fragment (3′-5′ exo-), 3 μg of phosphorylated genomic DNA product, and deionized water. The reaction is incubated at 37 °C for 30 min. The DNA product is then purified using a Qiagen DNA purification spin column and quantified by spectrophotometry.

3.2.5 Adapter ligation

To prepare for adapter ligation, a master stock is generated by dilution of each adapter oligonucleotide to a final concentration of 25 μ*M*. The adapter ligation reaction is set up in a PCR tube with 1 m*M* ATP, 10 m*M* DTT, 50 m*M* Tris·HCl pH 7.5, 10 m*M* $MgCl_2$, 30 units of T4 DNA ligase, 3 μL of 25 μ*M* stock adapter solution, ~2 μg of genomic DNA from Section 3.2.4, and deionized water. The reaction is then incubated at 16 °C for 30 min and prepared for PCR using a DNA purification kit.

3.2.6 PCR amplification

To ensure successful ligation of the adapters, the final DNA product is amplified by PCR using a set of two adapter sequences corresponding to the forward and reverse primers. The PCR reaction is set up with the entire adapter-ligated DNA product in 200 μ*M* dNTPs, 10 m*M* Tris·HCl, pH 8.3, 50 m*M* KCl, 1.5 m*M* $MgCl_2$, and 2 μ*M* of both forward and reverse primers. During PCR, aliquots are taken out every four cycles for a total of 24 PCR cycles. These PCR fractions are purified using a 2% agarose gel and density analysis to determine sequence diversity.

4. *IN VITRO* SELECTION OF RNA APTAMERS

4.1. Materials

4.1.1 Selection buffers

These buffers should provide physiological-like conditions to promote RNA structures similar to cellular ones. Binding buffer should facilitate binding of aptamers onto immobilized ligands. Elution buffer, in addition to being similar to its binding counterpart, contains elevated concentration of magnesium chloride to accommodate chelating by negatively-charged ligands (e.g., adenosine triphosphate; Vu et al., 2012), to maintain sufficient amount of Mg^{2+} to allow RNA to fold into a stable tertiary structure. In addition, the concentration of the free ligand should mimic its physiological concentration within the cell or exceed the approximate concentration of the ligand on the beads. Harsh elution buffer should contain denaturing condition in order to completely remove bound RNAs from the selection matrix. Storage of these buffers should adhere to appropriate conditions of their components. For example, where the ligand is a nucleotide triphosphate, the elution buffer should be frozen in multiple small aliquots to prevent hydrolysis.

4.1.1a Binding buffer: 140 m*M* KCl, 10 m*M* NaCl, 20 m*M* Tris·HCl, 5 m*M* $MgCl_2$.

4.1.1b Elution buffer: 140 m*M* KCl, 10 m*M* NaCl, 20 m*M* Tris·HCl, 5 m*M* $MgCl_2$, desired concentration of free ligand (in case of di- or triphosphorylated ligand, such as ATP, Mg^{2+} concentration needs to be increased by the ligand concentration to avoid changing the free Mg^{2+} concentration in the solution; e.g., we supplemented 5 m*M* ATP with an additional 5 m*M* $MgCl_2$; Vu et al., 2012).

4.1.1c Harsh elution buffer: 8 *M* urea and 5 m*M* EDTA.

4.1.2 Polyacrylamide gel electrophoresis

A stock of 15% can be prepared starting from the commercially available 40% acrylamide:bisacrylamide (19:1) solution. The 15% acrylamide stock should be preserved away from light and stored at 4 °C. Ammoniumpersulphate (APS) and tetramethylethylenediamine (TEMED) should be kept at 4 °C. The ^{32}P marker is for isolation of radiolabeled product once electrophoresis is completed.

4.1.2a 15% denaturing polyacrylamide stock: acrylamide:bisacrylamide (19:1) solution 40% w/v (CalBiochem), 1× TBE buffer, 7 *M* urea, water.

4.1.2b Gel polymerization agents: 10% APS in water and TEMED.

4.1.2c 2× RNA loading dye: 0.01% bromophenol blue, 0.005% xylene cyanol, and 8 *M* urea in 1× TBE buffer.

4.1.2d ^{32}P marker: [α-^{32}P] ATP, water, paper, scotch tape.

4.1.2e Ethanol precipitation: 300 m*M* KCl, GlycoBlue (Life Technologies), ethanol.

4.1.3 Agarose gel electrophoresis

4.1.3a 2% agarose gel: electrophoresis grade agarose, 0.5× TBE buffer, 1× ethidium bromide.

4.1.3b 2× DNA loading dye: 0.01% bromophenol blue, 0.005% xylene cyanol, and 8 *M* urea in 1× TBE buffer.

4.1.4 Transcription

The following reaction buffers should accommodate radiolabeled and nonlabeled transcripts, respectively. Albeit similar, the concentration of ATP included in the buffer for radioactive transcription is one order of magnitude less than that of the other nucleotides. This is designed to promote the incorporation of [α-^{32}P] ATP into the backbone of transcribed products, resulting in radiolabeled RNAs.

4.1.4a Buffer for radioactive transcription (2× stock): 4 m*M* Spermidine, 80 m*M* Tris·HCl pH 7.5, 20% DMSO, 20 m*M* DTT, 0.1% Triton X-100, 50 m*M* $MgCl_2$, 2 m*M* GTP, 2 m*M* CTP, 2 m*M* UTP, 0.2 m*M* ATP.

4.1.4b Buffer for nonradioactive transcription (2× stock): 4 m*M* Spermidine, 80 m*M* Tris·HCl pH 7.5, 20% DMSO, 20 m*M* DTT, 0.1% Triton X-100, 50 m*M* $MgCl_2$, 2 m*M* GTP, 2 m*M* CTP, 2 m*M* UTP, 2 m*M* ATP.

4.1.4c DNA template: approximately 0.1 μ*M* of genomic library or previous round's PCR product (~0.1 μ*M*) final concentration.

4.1.5 Reverse transcription

50 m*M* Tris·HCl (pH 8.3 at 25 °C), 75 m*M* KCl, 3 m*M* $MgCl_2$, 10 m*M* DTT, 0.5 m*M* dGTP, 0.5 m*M* dCTP, 0.5 m*M* dTTP, 0.5 m*M* dATP, 0.5 μ*M* reverse primer, RNA, 1 unit of reverse transcriptase.

4.1.6 Polymerase chain reaction

1 × *Taq* buffer (New England BioLabs), 0.5 m*M* dGTP, 0.5 m*M* dCTP, 0.5 m*M* TTP, 0.5 m*M* dATP, 2 μ*M* reverse primer, 2 μ*M* forward primer, cDNA, 1 unit of thermophillic DNA polymerase.

4.1.7 Enzymes

4.1.7a Transcription: T7 RNA Polymerase (New England Biolabs).

4.1.7b Reverse Transcription: ThermoScript Reverse Transcriptase (Life Technologies).

4.1.7c PCR: *Taq* DNA Polymerase (New England Biolabs).

4.1.8 Affinity column for in vitro selection

The selection matrix should have the target ligand immobilized onto it. Most experiments utilize commercially available sepharose matrix, e.g., ATP agarose (gamma phosphate-linked, Innova BioSciences; C8-linked; Sigma–Aldrich; Vu et al., 2012), or synthesized them via different linker with activated matrix, such as thiol sepharose (Davis & Szostak, 2002) or *N*-hydroxysuccinimide (NHS) sepharose (GE Healthcare Life Sciences). The concentration of immobilized ligand is often available form the manufacturer or should be measured by spectrophotometry (for example, using a UV–vis spectrophotometer with an integrating sphere or by fluorescence intensity if the ligand is fluorescent), and matched with the concentration of the free ligand within the elution buffer.

4.2. Procedure

4.2.1 Transcription

The protocols for radiolabeled and nonlabeled transcription are near identical, save for the addition of [α-^{32}P] ATP. Upon completion, the following reactions should be kept at −20 °C and should be used in the next experiment within the next few days to prevent RNA degradation and excessive decay of ^{32}P incorporated into the backbone. The transcription utilizes T7 RNA polymerase because the construct incorporates the T7 promoter.

The polymerase can be matched to a promoter of choice, depending on the design of the genomic pool.

4.2.1a Radioactive labeling of RNA via *in vitro* transcription: Mix 10 μL buffer for radioactive transcription, 2 μL DNA template, 0.5 μL T7 RNA polymerase, 0.5 μL [α-^{32}P] ATP (PerkinElmer), and deionized water to a total reaction volume of 20 μL in a small vial. Incubate at 37 °C for 3 h.

4.2.1b *In vitro* transcription: Mix 10 μL buffer for nonradioactive transcription, 2 μL DNA template, 0.5 μL T7 RNA polymerase, and deionized water to a total reaction volume of 20 μL in a small vial. Incubate at 37 °C for 3 h.

4.2.2 Purification of transcribed product

Transcribed RNAs can be purified by PAGE, as the reaction might contain side products, aborted constructs, and unreacted starting materials. In both radioactive and nonradioactive transcriptions, loading dye provides an approximate tracking of running samples, as well as increases the density of running samples to ensure that the RNAs stay within the wells. Visualization by UV shadowing (see step 4.2.2d) can identify nonradiolabeled RNAs on the polyacrylamide matrix, whereas the location of radiolabeled RNAs must be determined using ^{32}P markers and phosphor screen or photographic film. The markers help aligning the print-out onto the actual gel.

4.2.2a Polyacrylamide gel electrophoresis

1. Ensure that glass plates are clean and sterilized.
2. Prepare a 7.5% polyacrylamide solution by diluting polyacrylamide stock (15%) with equal volume of 8 *M* urea in 0.5 × TBE buffer.
3. For a gel cast of 16.5 cm by 22 cm dimension and 0.75 m*M* spacer, pour a plug at the end by mixing 5 mL 15% polyacrylamide, 50 μL 10% APS, and 5 μL TEMED. Wait until solidified.
4. Mix the rest of 15% polyacrylamide (20 mL) with 200 μL 10% APS and 20 μL TEMED. Pour into cast and apply comb.
5. Once gel is solidified, install onto electrophoresis apparatus. Overflow running wells with 0.5 × TBE.
6. Run gel at constant power of 20 W for 30 min.
7. Prepare running RNA solution by mixing *in vitro* transcription reaction and RNA loading dye (1:1 volume ratio).
8. Load solution onto wells. Run sample at 20 W. Duration is dependent upon the size of the transcription product.

4.2.2b Preparation of ^{32}P marker

1. Mix 5 μL 2× RNA loading dye, 5 μL water, and 0.5 μL [α-^{32}P] ATP
2. Pipette solution onto multiple spots of a small rectangular piece of paper
3. Once the liquid has evaporated, cover both side of the marker with tapes to prevent contamination. Now the dyed spots on the markers are radioactive.

4.2.2c Visualization of radioactively-labeled RNA

1. Remove gel plates from box once run is completed. Detach plates and cover gel with plastic wrap.
2. Secure two ^{32}P markers onto plastic-wrapped polyacrylamide gel with scotch tapes. Put a phosphor screen (GE Healthcare Life Sciences) on top of gel. Expose for at least 15 min.
3. Remove phosphor screen. Scan via Typhoon scanner under radioactive mode.
4. Obtain print-out of scanned image without scaling. Align image to gel based on location of ^{32}P markers.

4.2.2d Visualization of nonradioactive RNA

1. Remove gel plates from box once run is completed. Detach plates and cover gel with plastic wrap.
2. Visualize nonradioactive RNA with 254 nm UV light by placing the gel on a fluorescent thin layer chromatography plate. RNA absorbs the UV light, casting a shadow on the plate. Caution: prolonged exposure of RNA to UV light causes cross-linking.

4.2.2e Elution and precipitation of RNA from acrylamide matrix

1. Remove gel pieces containing visualized transcribed products.
2. Suspend gel pieces in 300 m*M* KCl. Elute at room temperature with shaking for at least 3 h. Add 0.5 μL GlycoBlue into the solution, mix well.
3. Transfer solution of eluted RNA onto a vial. Take care not to pipet out small bits of polyacrylamide. Add 95% ethanol at twice the volume of eluted solution. Let RNA precipitate in ethanol solution at −20 °C for at least 1 h.
4. Centrifuge at 20,000 × *g* for 30 min. Discard supernatant. Repeat at lower duration if necessary.
5. Let pellet dry at room temperature for 15 min. Resuspend evaporated RNA pellet in 20 μL physiological buffer.

4.2.3 In vitro *selection of RNA aptamers*

The selection strategy for aptamers is designed to enrich the pool for molecules that display high affinity and specificity toward the target ligand. The starting RNA library should exhibit a sufficiently high diversity to cover the whole genome at single-nucleotide resolution in both directions. The desired RNAs bind to the immobilized ligands on the selection agarose matrix, whereas the unbound RNAs are eliminated from the column by washing. Bound RNAs are then eluted through exchange with free ligand. Once the elutions are combined, free ligands and ions within the fraction can be removed using a desalting step (e.g., using Sephadex G25 column). It is critical that the presence of the free ligand be reduced prior to ethanol precipitation, as high concentration of free ligand can coprecipitate with selected RNA and inhibit subsequent reverse transcription. The progress of an *in vitro* selection is monitored by comparing the percentage of eluted RNAs across the rounds. In theory, the binding of the pool should increase with each round of selection, although it is often not measurable in the first few rounds of the selection. A counter-selection step can be included to increase stringency of the experiment, by incubating the RNA pool with beads containing a compound related to, but distinct from, the target molecule. A less stringent counter-selection step involves just the beads and the linker, and these are used to subtract molecules that have affinity for these components of the target system. A competitive elution with a ligand analog before the elution with the target ligand can be utilized to further increase selectivity of the aptamers (e.g., for ATP selection, a pre-elution step with dATP or ADP may be desirable to enrich for ATP-specific aptamers).

4.2.3a *In vitro* selection through affinity binding to targeted ligand

1. Withdraw desired volume of selection matrix onto a spin-filter or a disposable column. In principle, any amount of beads that results in correct effective concentration of the target molecule can be used, if the beads are agitated to ensure mixing of the RNA pool with the targets. The beads should ideally cover the filter surface, so that the solution is exposed to them during each step. Wash the beads with the same volume of binding buffer. Repeat at least three times. Incubate resuspended radioactive RNA solution at 70 °C for 3 min to unfold.
2. Transfer unfolded RNA onto beads and let incubate with shaking for 30 min at room temperature. Centrifuge column at $3000 \times g$ for 1 min. Collect flow-through.

3. Wash the beads with the same volume of selection buffer. Centrifuge at $3000 \times g$ for 1 min. Collect wash. Repeat three times.
4. Incubate the beads with the same volume of elution buffer. Incubate at room temperature with shaking for 30 min. The free ligand within the elution buffer and increased duration help promote ligand exchange. Centrifuge column at $3000 \times g$ for 1 min. Collect elution. Repeat three times.
5. Transfer the same reaction volume of harsh (see step 4.1.1c) elution buffer onto the beads. Agitate the reaction at room temperature for 15 min by rocking or shaking. Centrifuge at $3000 \times g$ for 1 min. Collect harsh elution. Add physiological buffer to dried beads.
6. Measure radioactivity of each fraction by Cherenkov counting on a liquid scintillation counter. This value will help quantify the percentage of eluted RNAs within the purified transcription.

4.2.3b Desalting of selected RNA

1. Combine eluted fractions and transfer to a Microcon centrifugal filter of appropriate size (YM-10).
2. Centrifuge at $14{,}000 \times g$ for 12 min. Invert the filter onto another collection vial, centrifuge at $1000 \times g$ for 3 min. Wash the filter with small volume (20–30 μL) of 300 m*M* KCl.
3. Collect filtered RNA, add 1 μL of GlycoBlue (or glycogen) and precipitate with ethanol.

4.2.4 Reverse transcription of selected RNAs

1. Prepare a 20 μL solution of reverse transcription buffer.
2. Dissolve RNA pellet in transcription buffer.
3. Place the reaction vial in a thermocycler.
4. Set the thermocyler at 45 °C for 30 min.
5. Store cDNA at −20 °C or continue on to PCR.

4.2.5 Polymerase chain reaction

The concentration of selected sequences after each round of *in vitro* selection is often not sufficient to move onto the next round. As such, these sequences need to be reverse transcribed and amplified to an appropriate concentration for the next transcription. When analyzing aliquots of the PCR, the aliquot with the fewest PCR cycles yielding a full-length band should be used for the subsequent selection.

4.2.5a Amplification of cDNA

1. Prepare a reaction mixture by mixing 5 μL 10× *Taq* buffer, 1.25 μL of 20 μ*M* forward primer, 1.25 μL of 20 μ*M* reverse primer, 2 μL of each dNTP at 5 m*M*, 1.5 unit of *Taq* enzyme (or other PCR-competent DNA polymerase), 10 μL of cDNA from the reserve transcription, and 39 μL of distilled deionized water to a total of 50 μL reaction volume.
2. Run the following steps on a thermocyler
   a. 95 °C for 3 min
   b. 95 °C for 30 s
   c. 55 °C for 30 s
   d. 72 °C for 1 min
   e. Repeat steps *b* to *d* for 32 cycles, withdrawing aliquots of 9 μL every 4 cycles, starting at cycle 8.

4.2.5b Agarose gel casting

1. Ensure that the plastic cast for agarose gel is clean and sterilized
2. Assemble the agarose casting apparatus. Install plastic comb with desired number of wells.
3. In a flask, combined 50 mL 0.5× TBE buffer with 1 g agarose.
4. Heat up the mixture in a microwave until all agarose is dissolved. Caution: overboiling of agarose might create spilling and burning hazard.
5. Add 0.5 μL ethidium bromide into 2% agarose in 0.5× TBE solution.
6. Pour the above solution onto agarose casting apparatus.
7. Let the agarose gel solidify at room temperature.

4.2.5c Agarose Gel Electrophoresis

1. Prepare running samples by mixing 2 μL of each aliquot with 2 μL of 2× DNA loading dye.
2. Remove the plastic comb and place the solidified 2% agarose gel into the electrophoresis box.
3. Fill the box with 0.5× TBE buffer until the buffer level is high enough to cover the agarose gel.
4. Load 2 μL of running samples into each well, and a DNA ladder for reference.
5. Collect the electric cables with a direct current power supply so that the samples run from cathode to anode.
6. Run electrophoresis at a constant potential of 200 V for 15 min.

4.2.5d Visualization of DNA on agarose gel

1. Remove agarose gel from the electrophoresis box.
2. Visualize DNA via UV light. Caution: prolonged exposure of DNA to UV light causes cross-linking.

5. STRUCTURE-BASED SEARCHES FOR NATURALLY OCCURRING APTAMERS

5.1. Materials

5.1.1 Unix compliant operating system

Any computer with a Unix platform can implement the RNABOB program. Our searches were run on iMac with Mac OS X 2.8 GHz Intel Core 2 Duo Processor and 2 GB 667 MHz DDR2 SDRAM.

5.1.2 RNABOB

The program can be downloaded from: ftp://selab.janelia.org/pub/software/rnabob/. Information regarding implementation can be found in the accompanying files (rnabob.man, rnabob.ps).

5.1.3 RNArobo

The program can be downloaded from http://compbio.fmph.uniba.sk/rnarobo/. Information regarding implementation is found on the same webpage.

5.2. Procedures

5.2.1 Descriptor

The descriptors (Fig. 2.2) used to find adenosine aptamers in the human genome are based on *in vitro* selected RNAs that bind adenosine containing molecules (Burgstaller & Famulok, 1994; Burke & Gold, 1997; Sassanfar & Szostak, 1993).

5.2.1a The motif of interest should be defined as a pattern using the following code: h for helical elements (allows G–U wobble pairs), r for relational elements or user-defined stringency of Watson-Crick base pairing for each residue, and s for single-stranded elements. The first line of the descriptor is the order of these elements beginning at the 5′ end of the motif. Each strand of helical and relational elements must be represented. For example, h1 and h1′ represent each side of helix 1.

5.2.1b The next lines of the descriptor define the nucleotide content of each individual element and each line contains three or four fields:

1. The name of the element from the motif topology given in the first line of the descriptor.
2. The number of nucleotide mismatches allowed. For each side of helical and relational elements, mismatches are specified separately. For example, "0:1" means the 5′ side of the helix may not contain any sequence mismatches (mutations), but the 3′ side can contain up to one base-pairing mismatch (mis-pair).
3. Sequence specificity written in IUPAC nucleotide code.
4. The list of strict base pairs required (T, C, G, or A). Only relational elements contain this field.

Note that the length of helices is intended to limit the size of output from structure-based searches and therefore also limit false positives returned. In general, more stringent descriptors yield a group of sequences with higher confidence to display *in vitro* activity. For adenosine aptamers, as long as the sequence and correct folding of the binding pocket are maintained, the length of helices and the structure of the peripheral domain (Fig. 2.2, s2) are arbitrary. Loosening the length requirement of helices can be written as "h1 0:0 NNNN[x]:NNNN[x]" which defines helix 1 as being at least four base pairs in length but it can be up to $4 + x$ base pairs. In addition, loosening s2 can be accomplished by increasing the number in the square brackets.

5.2.2 Sequence data

This file must contain DNA or RNA sequences in common database formats. Any sequence information can be searched, including whole genomes, metagenomic data, genes, ESTs, etc. Searches can focus on 5′ UTRs, intergenic regions, introns, 3′ UTRs, or noncoding regions.

REFERENCES

Burgstaller, P., & Famulok, M. (1994). Isolation of RNA aptamers for biological cofactors by in-vitro selection. *Angewandte Chemie International Edition in English*, *33*(10), 1084–1087.

Burke, D. H., & Gold, L. (1997). RNA aptamers to the adenosine moiety of S-adenosyl methionine: Structural inferences from variations on a theme and the reproducibility of SELEX. *Nucleic Acids Research*, *25*(10), 2020–2024. http://dx.doi.org/10.1093/nar/25.10.2020.

Curtis, E. A., & Liu, D. R. (2013). Discovery of widespread GTP-binding motifs in genomic DNA and RNA. *Chemistry & Biology*, *20*(4), 521–532. http://dx.doi.org/10.1016/j.chembiol.2013.02.015.

Davis, J. H., & Szostak, J. W. (2002). Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(18), 11616–11621.

Dieckmann, T., Butcher, S. E., Sassanfar, M., Szostak, J. W., & Feigon, J. (1997). Mutant ATP-binding RNA aptamers reveal the structural basis for ligand binding. *Journal of Molecular Biology*, *273*(2), 467–478.

Dieckmann, T., Suzuki, E., Nakamura, G. K., & Feigon, J. (1996). Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA*, *2*(7), 628–640.

Ellington, A. D., & Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, *346*(6287), 818–822.

Gautheret, D., Major, F., & Cedergren, R. (1990). Pattern searching/alignment with RNA primary and secondary structures: An effective descriptor for tRNA. *Computer Applications in the Biosciences*, *6*(4), 325–331.

Jiang, F., Kumar, R. A., Jones, R. A., & Patel, D. J. (1996). Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature*, *382*(6587), 183–186.

Jimenez, R. M., Rampasek, L., Brejova, B., Vinar, T., & Lupták, A. (2012). Discovery of RNA motifs using a computational pipeline that allows insertions in paired regions and filtering of candidate sequences. *Methods in Molecular Biology*, *848*, 145–158. http://dx.doi.org/10.1007/978-1-61779-545-9_10.

Riccitelli, N. J., & Lupták, A. (2010). Computational discovery of folded RNA domains in genomes and in vitro selected libraries. *Methods*, *52*(2), 133–140. http://dx.doi.org/10.1016/j.ymeth.2010.06.005.

Salehi-Ashtiani, K., Lupták, A., Litovchick, A., & Szostak, J. W. (2006). A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*, *313*(5794), 1788–1792.

Sassanfar, M., & Szostak, J. W. (1993). An RNA motif that binds ATP. *Nature*, *364*(6437), 550–553.

Singer, B. S., Shtatland, T., Brown, D., & Gold, L. (1997). Libraries for genomic SELEX. *Nucleic Acids Research*, *25*(4), 781–786.

Stoltenburg, R., Reinemann, C., & Strehlitz, B. (2007). SELEX—A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering*, *24*(4), 381–403. http://dx.doi.org/10.1016/j.bioeng.2007.06.001.

Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, *249*(4968), 505–510.

Vu, M. M., Jameson, N. E., Masuda, S. J., Lin, D., Larralde-Ridaura, R., & Luptak, A. (2012). Convergent evolution of adenosine aptamers spanning bacterial, human, and random sequences revealed by structure-based bioinformatics and genomic SELEX. *Chemistry & Biology*, *19*(10), 1247–1254. http://dx.doi.org/10.1016/j.chembiol.2012.08.010.